HARVARD UNIVERSITY
Graduate School of Arts and Sciences

DISSERTATION ACCEPTANCE CERTIFICATE

The undersigned, appointed by the

Department of Public Policy

have examined a dissertation entitled

"Essays on Shock Propagation in Economic Production Networks: Applications to U.S. Oil Price Episodes and Green Jobs"

presented by Stuart Iler

candidate for the degree of Doctor of Philosophy and hereby certify that it is worthy of acceptance.

Signature ______________________

Typed name: Professor Joseph Aldy, Chair

Signature ______________________

Typed name: Professor Robert Stavins

Signature ______________________

Typed name: Professor Soroush Saghafian

Date: May 2, 2022

Essays on Shock Propagation in Economic Production Networks:

Applications to U.S. Oil Price Episodes and Green Jobs

A dissertation presented

by

Stuart Iler

to

The Department of Public Policy

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Public Policy

Harvard University

Cambridge, Massachusetts

May 2022

Dissertation Advisor:
Professor Joseph Aldy

Author:
Stuart Iler

Essays on Shock Propagation in Economic Production Networks: Applications to U.S. Oil Price Episodes and Green Jobs

Abstract

This dissertation is composed of three essays that explore how dynamics in economic production networks can lead shocks to have differential impacts for producing entities and for workers. In Chapter 1, I propose a set of indicators that aim to capture, and predict, such differential impacts for producing entities based on both the configuration and the evolution of network connections. In terms of the latter, I focus specifically on how producing entities use their inputs as either substitutes or complements at the pairwise level. I confirm the indicators' predictive power by leveraging synthetic data produced by a computational model. To bridge this conceptual work and empirical applications, I also propose and test an approach that categorizes production processes' input pairs as substitutes or complements based on percentage changes in the usage of those inputs. In Chapter 2, I apply these ideas to the context of historical oil price episodes to construct a set of indicators for U.S. manufacturing industries over the period 1968-2018. Leveraging a regression framework, I find substantial heterogeneity in industry outcomes during oil price episodes that were tied both to industries' places in the network and to the changing input usage of the industries around them. The results also suggest that many oil price increases were at least partially demand-driven and that the supply-side pass-through of prices was due, at times, to industries' inability to substitute away from higher-priced petroleum products. In Chapter 3, I explore the relationship between green jobs and occupational transitions. Using data for the United States for 2001-2012, I train machine learning models that relate economic changes at the state-industry level to historical occupation-to-occupation flows. I leverage these models to explore how several scenarios of economic change could impact a set of 84 focus occupations, which were identified in the previous literature as either green or brown based on their titles, the tasks they perform, and/or the industries in which they are employed. I find that the models' predictions have both similarities and differences with these earlier categorizations, suggesting that the distinction between green and brown jobs may be blurred when viewed from the perspective of the broader networked economy.

Table of Contents

List of Tables

List of Figures

Acknowledgments

I am very grateful to everyone who has supported me throughout my Ph.D.

I would like to thank my advisors Joseph Aldy, Soroush Saghafian, and Robert Stavins for their valuable feedback on this dissertation research. Their insightful comments and suggestions helped to improve this work and to make it both more relevant and more clearly communicated. I would also like to thank our doctoral programs director, Nicole Tateosian, for her consistent support, responsiveness, and willingness to help.

I want to thank several other Harvard faculty members, including Pol Antràs, Richard Zeckhauser, Finale Doshi-Velez, and David Parkes, for taking the time to share their thoughts on this work. Their perspectives provided another lens through which to look at, and continue to develop, the research that eventually became this dissertation. I also want to acknowledge attendees of the Harvard Environmental Economics lunch whose feedback on early iterations of these ideas helped to clarify the direction in which this research might evolve.

My parents, Don and Karyn, and my sister and brother-in-law, Terri and Kristian, were instrumental in supporting me over these years. Their kind words, love, and encouragement helped me to keep perspective and to continue even when things were difficult. I am also fortunate to have an incredible group of friends whose support was a constant throughout this long process. Many thanks to you all—both family and friends—for everything you have given me.

Finally, and most importantly, I want to thank my wife, Emily, for being by my side during this journey from beginning to end. More than anyone, you have been a part of this process and everything it entailed. I cannot thank you enough for your love, for your encouragement, and for always believing in me. I have reached this milestone in large part because I have had you as my partner both in this process and in life. I love you.

Introduction

Production networks provide a conceptualization of the economy that explicitly accounts for the direct and indirect connections among producing entities. In contrast to models that take these entities as independent, production networks suggest that shocks to one part of the economy may have substantial impacts both for other parts of the economy and for the economy as a whole.

In the following three chapters, I consider the relationships among economic shocks, economic production networks, and outcomes for industries and workers. Chapter 1 is primarily conceptual and methodological, which sets the stage for Chapter 2 and Chapter 3 where I apply the ideas to U.S. oil price episodes and green jobs, respectively.

The main purpose of Chapter 1 is to explore the propagation of shocks in production networks with an emphasis on considering, and distinguishing, two components of shock effects. The first is the aspect of the effect due to the connectivity between a shocked entity and the other producing entities within the network. The second is the part of the effect due to how connections evolve over time, which critically depends on the ability or inability of producing entities to substitute among their inputs in response to varying prices.

To investigate these dynamics, I build a computational model of production networks that can simulate a range of artificial economies with a specific focus on heterogeneity in producing entities' usage of inputs. I use the synthetic data produced by the model for two purposes. This first is to examine a set of indicators that intend to capture both components of shock effects described above. I find that these indicators, which are related to—but distinct from—the Leontief inverse, are able to distinguish between the heterogeneous connectivity and evolutionary impacts experienced by producing entities in the presence of a shock. The results simultaneously build on the existing literature by providing further evidence that shock propagation is meaningfully affected by the idiosyncratic differences in how producing entities use their inputs as substitutes or complements.

In the last part of the chapter, I use the synthetic data to test an approach that aims to extract the pairwise substitutability or complementarity of producing entities' inputs using only changes in the quantities employed of those inputs. My goal in proposing the method is to create a bridge between conceptual work—such as the indicators developed in the chapter—and empirical applications. The method differs from previous approaches in several ways, including its focus on pairs of inputs rather than on groups of inputs. I find that the approach generally performs well in categorizing input pairs vis-à-vis the ground truth of entities' production processes.

In Chapter 2, I use a production networks view of the economy to explore the relationship between industries' economic performance and oil price episodes in the United States over the half-century between 1968 and 2018. Leveraging the results from Chapter 1, I construct a set of indicators that relate U.S. manufacturing industries to petroleum refining and products, oil and gas extraction, and automobile manufacturing. These indicators intend to capture industry-specific and economically relevant dynamics in the production network across three interrelated areas: supply shocks, demand shocks, and the substitutability and complementarity of industries' inputs. I evaluate the differentiating power of these indicators by examining their statistical significance for industries' year-to-year output and value added within a regression framework. The results reveal substantial heterogeneity in industry outcomes during oil price episodes and suggest that an industry's relative place in the production network—as well as the changing input usage of the industries around it—are important predictors of its economic growth.

The results also support the notion that oil episodes themselves have been associated with heterogeneous causes and consequences. Specifically, I find a key role for demand-driven oil price increases throughout most of the past half-century. This corroborates some of the more recent work in the oil episodes literature that has attributed price increases to demand factors in addition to supply factors, which stands in contrast to the earlier view that many episodes were primarily supply-driven.

At the same time, the results suggest a strong role for the supply-side transmission of shocks during some periods, including the finding that industries were differentially impacted by the relative inability of their upstream suppliers to shift away from petroleum products. This provides an explanation for why the supply-side channel may have been important during certain episodes even though the cost share of oil in production is relatively small. In addition, to the extent such supply-side transmission is indicative not just of a mechanism by which industries were affected, but also a sign of the underlying causes of the associated oil price changes, the results further suggest a decrease in the relative importance of supply-driven—versus demand-driven—oil price fluctuations in the latter half of the study period.

In Chapter 3, I take an empirical approach to exploring the relationship between occupational transitions and green jobs. To do so, I take the evolution of U.S. industries during the period 2001-2012 as given and train a series of machine learning models that relate this structure to historical occupation-to-occupation flows. I then leverage these models to explore the potential implications of five scenarios of economic change for a set of 84 focus occupations. I choose these particular occupations because they have been categorized in the previous literature as either "green" or "brown," depending on a combination of the tasks they perform, the industries in which they are employed, and/or their general prognosis for growth/decline in a transition to a green economy.

I take predicted change in net inflow—the difference between total flows of workers into and out of each occupation—as the central metric of interest. In this way, I define greenness not by characteristics of the occupations themselves or by features of the industries in which they work, but by their predicted growth or decline under the various scenarios. In such a classification scheme, jobs that would otherwise be considered brown—or be thought of as outside of the green-brown spectrum—can be placed onto either the green end or the brown end of the spectrum based on their potential growth trajectory in the overall movement towards a green economy.

Leveraging this metric and the predictions from the models, I rank the 84 focus occupations from most green to least green in each scenario. In comparing these rankings to previous categorizations of green and brown jobs, I find both similarities and differences, which suggests that the movement towards a green economy may affect occupations heterogeneously due to factors beyond the "greenness" of the tasks, tools, and/or industries with which they are associated.

In addition, although the rankings I produce are generally correlated across the scenarios, inspection at the level of individual occupations reveals some patterns of diversity. I explore these patterns using a hierarchical clustering algorithm, which splits the 84 occupations into groups based on differences in their rankings. The results suggest that about half of the occupations in the study experience moderate to large changes in their rankings when moving from some scenarios to others. For these occupations, the groupings indicate that predicted change in net inflow may be related both to the particular industries affected and to the manner in which resulting price changes are propagated through the U.S. production network.

Lastly, I examine patterns of worker movements between the 84 focus occupations and the "exchange occupations" with which they trade workers. I find that certain of these exchange occupations are predicted to play a central role, in the sense that there is a statistically significant relationship between their individual transition contributions and the overall rankings that result. This is most true in the case of worker exchange with unemployment, which plays this type of role across all of the scenarios.

Overall, the results suggest that the evolution towards a green U.S. economy may not be accompanied by a homogeneous transfer of workers from some occupations to others. Instead, a given set of industrial changes may have heterogeneous effects based on the back-and-forth trading of workers between focus and exchange occupations, where such transitions may benefit some workers while detrimentally affecting others. In turn, a key avenue for future research is to better understand these dynamics from an equity and welfare standpoint.

Chapter 1

A Computational Approach to Shock Propagation in Production Networks

1.1 Introduction

Production networks provide a conceptualization of the economy that explicitly accounts for the direct and indirect connections among producing entities. In contrast to models that take these entities as independent, production networks suggest that shocks to one part of the economy may have substantial impacts both for other parts of the economy and for the economy as a whole.

One key finding in the production networks literature is that producing entities' ability to substitute among their inputs is an important factor in the transmission of shocks. For instance, in an empirical context, Barrot and Sauvagnat (2016) analyze a dataset of U.S. firms and show (1) that suppliers hit by natural disasters impose substantial sales losses on their customers and (2) that the effect is larger if the supplier produces a specific input (i.e., if the input cannot easily be substituted for other inputs). The authors also find that the shock can be propagated horizontally from one supplier to another—via the impacted customers—if the supplier hit by the natural disaster produces a specific input. Taken together, these results suggest a role both for supplier-customer connections and for a particular aspect of those connections—input substitutability—in the transmission of shocks.

Similar findings have been echoed in theoretical work. As an example, Carvalho and Tahbaz-Salehi (2019), drawing on the model in Baqaee and Farhi (2018), derive a first-order approximation for the impact of a shock to one industry j on another industry i. In their model economy, each industry produces its output by combining labor with a bundle of intermediate inputs produced by other industries. The authors describe two channels through which industry i may be affected by the shock to j: (1) a downstream propagation effect whereby an increase in j's output price affects the output prices of its customers, and those customers' customers, and so on; and (2) an effect whereby j's customers (as well as those customers' customers, and so on) potentially readjust demand for their intermediate inputs, which propagates back upstream. Analogous to above, this suggests that it is both the connections among producing entities and the changes induced in those connections that are relevant for outcomes in the presence of shocks.

In this way, although such effects are the result of multiple forces occurring simultaneously, they can be conceptualized as having two parts. The first is the portion of the effect due to how producing entities are connected, directly or indirectly, to one another. The second is the portion of the effect due to how the network changes over time, and more precisely, how the magnitude of each connection evolves as entities adjust their inputs in response to price variation. For the purposes of this chapter, I refer to the aspects of the network that give rise to each of these components as the "connectivity" and the "evolutionary" parts of the network, respectively.

Mathematically, a network can be represented by an adjacency matrix: a two-dimensional grid of numbers that contains the magnitude of the edge between every pair of nodes. In an economic context, the nodes represent the firms or the industries, and the edges reflect the supplier-customer relationships—typically measured in dollars—among them. An input-output table, which records the financial flows between industries in the economy, is essentially a weighted, directed, adjacency matrix (McNerney, 2009). For this reason, empirical analyses of the industry-level production network often use input-output tables as a main data source. (Given the conceptual similarities between adjacency matrices and input-output tables, for the remainder of the chapter I slightly abuse terminology and refer to both adjacency matrices and input-output tables as simply input-output tables. I do so with the acknowledgment that, in an empirical context, a network representation of the economy can be constructed from various data sources, including but not limited to the traditional input-output table as published within the national economic accounts.)

If we call the input-output table A, then the Leontief inverse $(I - A)^{-1}$ (where I is the identity matrix) measures the direct and indirect connections among all producing entities in the economy. In input-output analysis, developed by Wassily Leontief in the late 1930s, the Leontief inverse is used to calculate the production required by each entity to satisfy final demand taking network dependencies into account.

A key aspect of the standard input-output model is that the ratios of each industry's inputs remain fixed. In turn, every pair of inputs in an industry's production process are perfect complements—in the sense that output cannot be increased by increasing one input without increasing the other—and the overall production process is captured mathematically by the classic Leontief production function. In a parallel manner, analyses of production networks that rely on one input-output table, Leontief inverse, or other single network snapshot will, without additional information, only directly capture the connectivity part of the dynamics described above, as they ignore—or at least do not differentiate—the portion of the shock effect due to how producing entities adjust their input usage.

As I describe below in more detail, some empirical studies have incorporated changing input use by estimating the elasticities of substitution in nested production function models (e.g., Atalay, 2017) or by proxying for complementarity using measures such as R&D expenditures (e.g., Barrot and Sauvagnat, 2016). Other studies have instead leveraged covariates based only on a single snapshot of the production or supply chain network, sometimes drawing directly on the Leontief inverse and sometimes using other measures (e.g., Acemoglu, Akcigit, and Kerr, 2016; Carvalho, 2014; Carvalho et al., 2021).

In this chapter, I take a computational approach to exploring the propagation of shocks in production networks with a particular emphasis on considering, and distinguishing, the connectivity and the evolutionary components of shock effects. Via the model I construct, the indicators I propose, and the input-categorization method I develop, I attempt to bridge the static Leontief inverse on one hand with the role of input substitution and complementarity on the other. Although the chapter is primarily conceptual and methodological, my intention is for the results to have direct applicability to empirical contexts.

In considering these dynamics, I make three main contributions to the existing literature. The first is to build a computational model of production networks that can simulate a range of artificial economies with a specific focus on heterogeneity in producing entities' usage of inputs. I take a primarily computational, rather than an analytical, approach to model-building because it allows me to examine shock transmission across a wide variety of network structures and production processes while also precisely defining the particular heterogeneity within each artificial economy. Through the synthetic datasets it produces, the model also facilitates the two other parts of the study, which I describe below.

A key feature of my computational approach is a generalization of the notion of nested production functions, which I reconceptualize as networks. In this formulation, a production process is a collection of production steps, and each production step is either a two-input Cobb-Douglas production function or a two-input Leontief production function. These steps are linked together to form a network that represents the overall production process for each entity. This framework has two related advantages: (1) it provides both the detail and the flexibility—at the input-by-input level—to capture a wide spectrum of individual production processes; and (2) it allows for the unambiguous classification of every pair of inputs as either substitutes or complements based only on the structure of the process network, where these classifications are symmetric for each input pair and independent of prices. I implement this idea of production-processes-as-networks computationally and then embed it within an input-output framework.

The second main contribution of the chapter is to propose and explore a set of indicators that aim to capture both the connectivity and the evolutionary parts of shock dynamics in economic networks. These indicators, which are distinct from—yet related to—the Leontief inverse, can be thought of as having two related uses: on one hand, they represent a way to predict or explain differential impacts to producing entities in the presence of various types of shocks, where such shocks might be known or postulated. From this view, such indicators could be useful in anticipating which entities would be most affected by the enactment of a proposed policy, such as the implementation of a carbon tax. At the same time, these indicators also provide a method of inferring the presence of different types of shocks given the observed (and varying) economic performance of entities. For example, in a retrospective study using historical data, such indicators could help to uncover the extent to which, and by what mechanisms—such as supply- versus demand-side, and connectivity versus evolutionary—entities were differentially affected by various shocks. In Chapter 2, I use the indicators for this latter purpose in the context of historical oil price episodes.

To explore the differentiating power of the indicators, I use the computational model to simulate a large number of random, artificial economies composed of networked industries, where the production process of each industry is a network of production steps. I leverage a regression analysis of the model's artificial data to show that the indicators contain relevant information about both the connectivity and the evolutionary parts of shock dynamics in the economic networks that I model. The results simultaneously provide evidence—building on the existing literature—that idiosyncratic, input-level substitution heterogeneity is a substantial factor in shock transmission, even when controlling for the information provided by the classic Leontief inverse.

Finally, the third main contribution of the chapter is to propose an approach that aims to extract the pairwise substitutability and complementarity of producing entities' inputs using only changes in the quantities employed of those inputs. I test the approach using synthetic data from the computational model and show that it generally performs well in recovering the substitutability and complementarity of input pairs vis-à-vis the ground truth of entities' production processes.

This method differs from previous approaches in: (1) its focus on pairs of inputs rather than on groups of inputs; (2) its binary categorization of inputs as substitutes or complements, without an estimate of elasticity; and (3) its use of machine learning rather than linear regression or other econometric techniques. The primary purpose in developing the method is to serve as a bridge between conceptual work—such as the indicators in this chapter—and empirical applications. For example, although the connectivity indicators I propose can be constructed from available data (that is, directly from an input-output table or other network snapshot), the evolutionary indicators are based on underlying information about producing entities' production processes, which must be inferred from available data. In this way, the input-categorization approach provides a critical step in translating the evolutionary indicators to an empirical setting.

Related Literature. This chapter is primarily related to previous work in the areas of economic production networks, nested production functions, and estimation of elasticities of substitution. It is also broadly related to research that has applied machine learning to synthetic datasets.

Economic Production Networks. This chapter builds on a growing theoretical and empirical literature that has investigated the dynamics and the implications of production networks.

Theoretical papers in the production networks literature include early work by Long and Plosser (1983) as well as more recent studies by Gabaix (2011), Acemoglu et al. (2012, 2017), Baqaee (2018), and Baqaee and Farhi (2019). A common theme across these papers is an examination of how industry or firm interconnections can lead micro-level shocks to influence aggregate outcomes. Related empirical work by Foerster, Sarte, and Watson (2011) and Atalay (2017) use multi-sector models to assess the relative importance of sectoral versus aggregate shocks in contributing to aggregate fluctuations. Another set of research—including papers by Acemoglu, Akcigit, and Kerr (2016), Barrrot and Sauvagnat (2016), Carvalho et al. (2021), and Baqaee and Farhi (2018)—has instead focused on how firm- or industry-level shocks propagate through production networks to affect other firms or industries.

The current study is broadly related to this literature in its consideration of production networks as a conduit for shocks. It is more specifically related to the papers in the last category that focus on the transmission of shocks from one producing entity to another. I discuss several of these latter papers briefly here and then highlight the similarities and differences with the current study.

As mentioned above, Barrot and Sauvagnat (2016) use a firm-level dataset to examine shock propagation—precipitated by natural disasters—based on covariates that capture direct supplier-customer relationships as well as supplier input specificity (as proxied by sale on an organized exchange, R&D expenses, and number of patents). These covariates are non-constant and vary for each quarter in the study time period (1978-2013). As previously described, the authors find that input specificity is a key determinant of shock transmission both from suppliers to customers and from suppliers to other suppliers.

In a similar vein, Carvalho et al. (2021) leverage a dataset of Japanese firms to examine shock propagation over input-output linkages in the wake of the Great East Japan Earthquake of 2011. Their covariates capture whether each firm was—prior to the shock in 2010—a supplier to or a customer of (both directly and at distances of two, three, and four in the network) a firm that would be affected by the shock. They find that, after the earthquake, the growth rates of firms at distances one, two, three, or four from a shocked firm were all lower than the growth rate of a control group. Their results also show that this growth rate effect decreases in magnitude the farther away each firm is from an earthquake-affected firm.

Acemoglu, Akcigit, and Kerr (2016) instead look at U.S. industries (rather than firms) and consider four shocks: imports from China; federal government spending; total factor productivity (TFP); and foreign-industry patents. For each of these shocks, the authors construct three industry-level covariates: (1) the direct magnitude of the shock; (2) the shock as transmitted via upstream connections; and (3) the shock as transmitted via downstream connections. These last two covariates are based on the Leontief inverse, which the authors construct from the U.S. Bureau of Economic Analysis' 1992 input-output tables. They find that the propagation of these shocks through the production network is statistically significant, and as their model suggests, the two demand-side shocks (Chinese imports and federal government spending) tend to propagate upstream while the two supply-side shocks (TFP and foreign patenting) tend to propagate downstream.

Lastly, Baqaee and Farhi (2018) take a theoretical approach to investigating the propagation of shocks from one producer to another. A key feature of their model is to allow for heterogeneity in consumption preferences and production functions, which they use to shed light on a number of applied questions, including sectoral comovement in business cycles. One of their general insights is that once the economy is disaggregated into the primitives in their model, "local" elements—including elasticities of substitution—play an important role in the transmission of shocks.

The current study is similar to these papers in several ways: like all four papers, it focuses on shock propagation from one producing entity to another; like Barrot and Sauvagnat (2016) and Baqaee and Farhi (2018), it explicitly takes account of input substitutability; like Acemoglu, Akcigit, and Kerr (2016), it includes covariates based on the Leontief inverse; and like both Carvalho et al. (2021) and Acemoglu, Akcigit, and Kerr (2016), it considers shock propagation over multiple levels in the production network.

It also differs from these studies in several ways: it considers the pairwise substitutability and complementarity among inputs by conceptualizing production processes as networks, while also providing a methodology to estimate such substitutability/complementarity from data; it suggests indicators beyond either distance-to-shock or the classic Leontief inverse, with a specific focus on both the connectivity and evolutionary portions of shock dynamics; and, more broadly, it takes a computational rather than an analytical approach to model-building.

Nested Production Functions. The current chapter also has a relationship to previous research that uses nested production functions, and especially to those papers that leverage and/or focus on nesting as a way to incorporate heterogeneity in the elasticity of substitution. Nested production functions have been used extensively in the production networks literature as well as in other literatures, including in studies that estimate production input elasticities and in research using computable general equilibrium (CGE) models.

For instance, the theoretical basis provided by nested constant elasticity of substitution (CES) production functions has been leveraged to empirically estimate the substitutability among capital, labor, and energy (Brockway et al., 2017; Dissou et al., 2015; Papageorgiou et al., 2017). These types of estimates have been widely applied within CGE models, which are often used for the assessment of public policies.

Nested production functions also appear in the production networks context and in closely related research based on multi-sector models. For example, Atalay (2017) and Carvalho et al. (2021) both use a nested model that combines a set of material inputs (with one substitution parameter) to create an intermediate output, which is then combined with labor or labor/capital (via a second substitution parameter) to produce a final output. As described above, Carvalho et al. (2021) empirically analyze a large dataset of Japanese firms, which they interpret through the lens of their model. Atalay (2017) instead leverages the model to estimate the two substitution parameters based on data for 30 U.S. industries (see below).

In a theoretical context, Baqaee and Farhi (2018) (also referenced above) present their main results using a model where producers create their outputs based on an arbitrary set of nested CES production functions. Unlike much of the other literature, however, the authors generalize these results to arbitrary neoclassical production functions by introducing what they call the "input-output substitution operator," which captures the substitutability between every pair of inputs for a producer via the combination of the Allen-Uzawa elasticity of substitution and other primitives in their model.

Altogether, my approach is similar to these previous studies in that it relies on nested production functions, but it differs in that it generalizes and then re-interprets the composition of functions as a network. Specifically, in the implementation I describe below, a production process of n inputs has $(n-1)$ nests, which I formalize as a network of production nodes. Adding a nest for every input allows for the representation of a broad set of production processes.

In addition, the framework I propose is similar to Baqaee and Farhi (2018) in that it allows for heterogeneity in the substitutability of every pair of inputs, and in this way, it differs from many other models that use a single parameter to represent substitutability across multiple inputs. At the same time, it also differs from Baqaee and Farhi (2018) in that my definitions of substitutability and complementarity are based directly on the structure of the production process network without regard to prices.[1]

Lastly, my approach differs from previous research in its focus on computation. As with my model more broadly, the production process framework is designed so that it can be implemented computationally to produce large synthetic datasets. This serves as both an alternative and a complement to the analytical approach taken in the existing literature.

Estimating Elasticities of Substitution. The input-categorization approach I propose in this chapter is broadly related to existing work that looks to estimate the elasticities of substitution among inputs. This previous work has appeared across a range of literatures, including those related to CGE modeling, trade, and multi-sector models.

There are many different forms for the elasticity of substitution, beginning with the concept as it was first introduced by Hicks in the early 1930s (Mundra and Russell, 2004; Stern, 2011). For example, in addition to the Hicks formulation, other variations include the Allen-Uzawa elasticity of substitution and the Morishima elasticity of substitution. Although many of these elasticities are equal in the two-input case, they are generally not equal when there are three or more inputs (Brockway et al., 2017; Stern, 2011).

Analogously, empirical investigation has revealed that different forms of the elasticity of substitution can yield varying classifications of inputs even when applied to the same dataset. For this reason, some researchers have argued that different elasticity concepts should be used in different contexts depending on the questions being asked (Mundra and Russell, 2004; Stern, 2011).

Approaches to estimating elasticities of substitution are as varied as the definitions themselves. One example, as mentioned above, is the estimation of the substitution parameters of capital-labor-energy nested CES production functions to serve as an input to CGE and other macroeconomic modeling. Studies within this literature have employed nonlinear least squares, solving systems of linear equations, and direct linear estimation, among other techniques (Brockway et al., 2017). Although there are various trade-offs involved with these particular methods, one general challenge with this approach is that estimates are dependent on the order of the nests. This has led some

[1] I note that although my definition is based only on the structure of a production process, it is loosely akin to the Allen-Uzawa taxonomy, which classifies input pairs as direct substitutes or complements if a price increase for one of the inputs increases or decreases, respectively, the quantity demanded of the other (Mundra and Russell, 2004).

researchers to estimate, or to advocate estimating, all possible nest configurations so that the results can be directly compared (Brockway et al., 2017; Dissou et al., 2015).

Another example, from the trade literature, is the work by Boehm, Flaaen, and Pandalai-Nayar (2019), who use a firm-level dataset—focused on the time around the 2011 Tōhoku earthquake—to estimate the elasticity between intermediate materials sourced from Japan and intermediate materials from all other sources.[2] To do so, they take firm-level production to be of a nested CES form, where Japanese materials are combined with all other materials to create an aggregate, which is then combined with a capital-labor aggregate to produce final output. They structurally estimate this production function for each firm using data from the pre-earthquake period, and then apply those results to the post-earthquake period to infer the elasticity. They find that the ability of firms to substitute between Japanese and other material inputs is relatively low, and especially so for Japanese multinational firms. In turn, they conclude that rigid production complementarities are a key factor in the cross-country transmission of shocks.

As a last example, Atalay (2017) (also discussed above) uses a two-stage least squares approach to estimate, for a set of U.S. industries, the substitution elasticities between intermediate material inputs and between an aggregate of those materials and capital/labor. Using instruments similar to Acemoglu, Akcigit, and Kerr (2016), the author finds that the first elasticity is near zero, suggesting that the intermediate goods produced by industries enter as complements in downstream industries' production processes.

The approach I propose to recover the pairwise substitutability or complementarity among inputs differs from these earlier methods in several key ways. The first is that the estimates produced by the approach are at the level of individual input pairs. This differs from previous methods that, as in the examples above, group multiple inputs (and especially intermediate material inputs) into an aggregate associated with a single substitution parameter.

A second difference is that the approach categorizes input pairs in a binary way (as simply either substitutes or complements) based only on changes in input and output quantities. In this way, it eschews the notion of the elasticity of substitution, including definitions that rely on prices. I describe later in the chapter how the degree of substitutability could potentially be estimated (and the estimates tested) by expanding both the computational model and the approach itself.

Finally, the approach differs from previous work in that it employs machine learning rather than econometric techniques. I choose machine learning in an attempt to allow for more flexibility in, and

[2] Another example in the trade literature is the work of Feenstra et al. (2018), who—leveraging both simulated data and U.S. data—use two-stage least squares and two-step generalized method of moments to estimate the substitution elasticity (1) between different foreign sources and (2) between foreign and domestic sources.

to make fewer assumptions about, the underlying form of entities' production processes than might be the case in some statistical frameworks. This said, I evaluate the accuracy of the approach using synthetic data created within a model of production-processes-as-networks, and in this way, these accuracy results can be translated to applied contexts only to the degree to which the modeling reflects the important features of real-world production processes.

Machine Learning and Synthetic Datasets. Lastly, the current study is also broadly related to previous work that applies machine learning to synthetic datasets.[3]

For instance, given the difficulty in some contexts of acquiring or developing datasets to use in training machine learning algorithms, an alternative is to create synthetic data. This approach has been used successfully in a variety of applications—many of which are related to computer vision—including motion recognition (Mayer et al., 2018), identification of individuals (Barbosa et al., 2018), and vehicle detection (Moate et al., 2018). In this chapter, I use synthetic data both to test the proposed indicators and to evaluate the input-categorization approach.[4]

The chapter proceeds as follows: in section 1.2, I develop an analytical variant of the model to provide background and intuition for the computational model and the indicators; in section 1.3, I describe the production-process-as-network framework; in section 1.4, I describe the details of the computational model, which relies on that framework; in section 1.5, I propose and evaluate the connectivity and evolutionary indicators; in section 1.6, I describe and test the input-categorization approach; and in section 1.7, I provide some concluding thoughts.

1.2 Analytical Model

I begin by developing an analytical variant of the model for two purposes. The first is that the analytical variant serves to illustrate the functioning of the computational model, especially in the sense that both versions view each instance of the input-output table as the result of particular industry input choices. The primary difference between the two models is that the analytical variant takes these choices as the result of a general cost-minimization process, whereas the computational

[3]In other contexts, the term "synthetic data" sometimes refers to datasets constructed from underlying samples that have been modified to prevent the disclosure or reconstruction of sensitive information (Berg et al., 2016). For example, the U.S. Census Bureau publishes the Synthetic Longitudinal Business Database, a longitudinal microdata product that contains altered data values at the establishment level. In this chapter, I use the term "synthetic data" to instead refer to entirely artificial datasets created for the purpose of further analysis.

[4]I also note that machine learning has been used to infer general rules and theories about large and complex systems across several domains. As three examples, this includes the work of Schmidt and Lipson (2009) for physical systems, Brummitt et al. (2020) for economic systems, and Dutting et al. (2019) for optimal auction design. In the current chapter, I use an analytical variant of my model to suggest a set of connectivity and evolutionary indicators, which I then test using synthetic data from the computational model. In future work, machine learning could be used for the reverse purpose: the computational model could be used to generate large, synthetic datasets that contain hundreds or thousands of features about the economic production network, and machine learning could be applied to extract potential indicators for use in empirical contexts.

model embeds the production-processes-as-networks framework within this structure, allowing for the exploration of a variety of specific functional forms.

The second purpose of the analytical variant is to provide intuition about the mechanisms by which supply and demand shocks might combine with the production network to result in differentiated outcomes for producing entities. In this respect, the model parallels some of the previous models in the literature, including its consideration of the propagation of supply shocks from suppliers to customers, the propagation of demand shocks from customers to suppliers, the horizontal propagation of shocks from suppliers to other suppliers (via affected customers), and more generally, the transmission of shocks via both connectivity and evolutionary conduits. It differs from previous models primarily in its focus on the explicit heterogeneity among producing entities' usage of their inputs, which (as mentioned in the previous paragraph) takes the form of embedding these input choices within an input-output framework. I draw on the insights from the analytical model to propose the indicators in section 1.5.

1.2.1 Model Derivation

At the broadest level, the analytical model provides a mapping from prices and final demand to values for intermediate demand and producing entities' revenue and profits. In this way, the model can be used to explore the comparative statics of an incremental change in a specific price or final demand component (a "shock"). As described in the introduction, in traditional input-output analysis, the ratio of inputs for an industry are taken as fixed, whereas in the model below, they are allowed to vary when prices change. Specifically, a given set of prices results in a particular instantiation of the economy's input-output matrix; input-output analysis can then be used to solve for the required vector of intermediate inputs.[5]

For simplicity and tractability, I consider a model composed of three industries, each of which produces a unique good. We might think of this model as representing a sub-economy within a larger economy, where inputs to and demand from the sub-economy are specified by a set of exogenous functions. Another interpretation is that one or more of these three industries are actually collections of industries, which have been collapsed into single network nodes to simplify the analysis. In either case, for ease of language, I refer to the model economy/sub-economy as simply "the economy" and to each industry/industry collection as one of "industry one," "industry two," or "industry three."

Let industries be indexed by $i, j \in \{1, 2, 3\}$. Each industry i produces its good by using a combination of the other industries' goods, as well as (potentially) using one or more inputs from

[5] Please see the appendix for a brief background discussion of input-output analysis.

outside of the three-industry economy. In cases where industries rely only on outside suppliers, marginal production costs are provided by exogenous functions and may (or may not) be independent of the prices and production of the other industries inside of the three-industry economy.

Demand for industry i's good is the sum of: (1) what is needed for the production of other industries' goods, and (2) final demand, which exists outside of the three-industry economy and is specified by an exogenous function. I refer to goods that are used only as inputs to other industries' production processes as "intermediate goods." If there is non-zero final demand for a good, I refer to the good as a "final good." (Note that a final good may also be used as an intermediate input.)

Each industry i's production technology is given by $g_i(x_1, x_2, x_3)$, where x_1, x_2, and x_3 are the input quantities (in physical terms) of the goods produced by industries one, two, and three, respectively. The output of $g_i(\cdot)$ is a scalar value representing the quantity (in physical terms) of industry i's good produced using the given input quantities. In this general formulation, an industry might rely on its own good in its production process. For simplicity in the model, I generally take these self-dependencies to be zero.

I abstract from the price-setting mechanism and take prices in dollars, $p' = [p_1 \quad p_2 \quad p_3]$, as given. I take final demand (in physical terms) as specified by some set of exogenous functions, $q_i(p)$, $i \in \{1, 2, 3\}$. Final demand in dollars is given by $d' = [d_1(p) \quad d_2(p) \quad d_3(p)]$ where $d_i(p) = q_i(p) \cdot p_i$. Note that below I use the shorthand q_1, q_2, and q_3, and d_1, d_2, and d_3, with the understanding that these values depend on the current vector of prices.

Each industry i—given prices and total demand (both intermediate and final)—chooses its input quantities to satisfy demand at the minimum total cost. I assume that the functions $g_i(\cdot)$ exhibit constant returns to scale, so I consider the inputs required to produce one unit of an industry's output. As a result of this process, I say that industry i demands quantity $f_{ji}(p)$ of industry j's output to produce one unit of its own output:

$$[f_{1i}(p), f_{2i}(p), f_{3i}(p)] = \arg\min_{x_1, x_2, x_3} \Big[p_1 \cdot x_1 + p_2 \cdot x_2 + p_3 \cdot x_3 \Big] \quad \text{such that} \quad g_i(x_1, x_2, x_3) = 1$$

For ease of notation, I write f_{ji} below with the understanding that, at any particular time, the specific input quantities demanded depend on the current vector of prices.

For the three-industry model, I collect these f_{ji} values into a 3×3 matrix, which I call T:

$$T = \begin{bmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{bmatrix} \tag{1}$$

This matrix describes the fundamental structure of the production network at a given time under given prices. The matrix T can also be interpreted as an adjacency matrix for a directed graph over the set of nodes $N = \{1, 2, 3\}$. A nonzero value f_{ji} in T indicates a weighted edge from node j to node i, where the weight is given by f_{ji}. In this way, the production processes of industries combine with price vectors to create specific instantiations of this directed network.

In addition to prices p and final demand d as described above, I define p^{-1} as a vector of the inverses of the prices:

$$p^{-1} = \begin{bmatrix} 1/p_1 \\ 1/p_2 \\ 1/p_3 \end{bmatrix}$$

Finally, I also define diagonalizations of p and p^{-1}:

$$diag(p) = \begin{bmatrix} p_1 & 0 & 0 \\ 0 & p_2 & 0 \\ 0 & 0 & p_3 \end{bmatrix}, \qquad diag(p^{-1}) = \begin{bmatrix} 1/p_1 & 0 & 0 \\ 0 & 1/p_2 & 0 \\ 0 & 0 & 1/p_3 \end{bmatrix}$$

The first step in deriving the model is to construct the direct requirements table, which has as its columns the expenditures of each industry (in dollars) to produce one dollar of its own output. To do this, I first multiply each row of T by the associated input's price (multipliers of p_1, p_2, and p_3 for the first, second, and third rows, respectively). Then, I divide each column by the associated industry's output price (divisors of p_1, p_2, and p_3 for the first, second, and third columns, respectively). This yields the matrix A:

$$A \quad = \quad [diag(p) \cdot T] \cdot diag(p^{-1}) \quad = \quad \begin{bmatrix} f_{11} & (\frac{p_1}{p_2})f_{12} & (\frac{p_1}{p_3})f_{13} \\ (\frac{p_2}{p_1})f_{21} & f_{22} & (\frac{p_2}{p_3})f_{23} \\ (\frac{p_3}{p_1})f_{31} & (\frac{p_3}{p_2})f_{32} & f_{33} \end{bmatrix} \tag{2}$$

Next, I find the Leontief inverse—given by $(I - A)^{-1}$—and then, as in standard input-output analysis, I solve for the vector of industry outputs (in dollars) as: $x = (I - A)^{-1} \cdot d$ (please see the appendix for details of the derivation). From here on, I consider the simplified case where there are no industry self-dependencies (that is, $f_{11} = f_{22} = f_{33} = 0$ and the matrix $(I - A)$ has ones on the main diagonal). This yields the following for each industry's output:

$$x_1 = \left[d_1 - f_{23} f_{32} d_1 + \left(\frac{p_1}{p_2} \right) [f_{13} f_{32} + f_{12}] d_2 + \left(\frac{p_1}{p_3} \right) [f_{12} f_{23} + f_{13}] d_3 \right] \bigg/ det(I - A)$$

$$x_2 = \left[\left(\frac{p_2}{p_1} \right) [f_{23} f_{31} + f_{21}] d_1 + d_2 - f_{13} f_{31} d_2 + \left(\frac{p_2}{p_3} \right) [f_{13} f_{21} + f_{23}] d_3 \right] \bigg/ det(I - A)$$

$$x_3 = \left[\left(\frac{p_3}{p_1} \right) [f_{21} f_{32} + f_{31}] d_1 + \left(\frac{p_3}{p_2} \right) [f_{12} f_{31} + f_{32}] d_2 + d_3 - f_{12} f_{21} d_3 \right] \bigg/ det(I - A)$$

where

$$det(I - A) = 1 - f_{12} f_{21} - f_{23} f_{32} - f_{13} f_{31} - f_{12} f_{23} f_{31} - f_{32} f_{21} f_{13}.$$

Note that if there are no cycles in the graph of the production network—such as industry 1 depends on industry 2, which depends on industry 1; or industry 1 depends on industry 2, which depends on industry 3, which depends on industry 1—then the determinant equals one.

Assuming a uniform price for all of an industry's output, the physical quantity of output for each industry, z, is computed by dividing industry output by industry price:

$$z_1 = \left[q_1 - f_{23} f_{32} \cdot q_1 + (f_{13} f_{32} + f_{12}) \cdot q_2 + (f_{12} f_{23} + f_{13}) \cdot q_3 \right] \bigg/ det(I - A)$$

$$z_2 = \left[(f_{23} f_{31} + f_{21}) \cdot q_1 + q_2 - f_{13} f_{31} \cdot q_2 + (f_{13} f_{21} + f_{23}) \cdot q_3 \right] \bigg/ det(I - A)$$

$$z_3 = \left[(f_{21} f_{32} + f_{31}) \cdot q_1 + (f_{12} f_{31} + f_{32}) \cdot q_2 + q_3 - f_{12} f_{21} \cdot q_3 \right] \bigg/ det(I - A)$$

where I have substituted $q_i = \frac{d_i}{p_i}$ and where $det(I - A)$ is defined as above. In words, these equations state that each industry i produces a total quantity (z_i) to satisfy its own final demand (q_i) as well as the final demand of the other industries (q_j, $j \neq i$) as channeled through connections in the production network. For each industry i, this total quantity is adjusted in the numerator both by the potential presence of a cycle between the other two industries and by the paths (direct and indirect) from i to the other two industries. The total quantity is also adjusted in the denominator by the presence of cycles anywhere in the economy.

Finally, given the assumption of constant returns to scale, industry total profits, π, can be calculated as the product of unit profits and total quantity produced:

$$\pi_1 = \Big[(p_1 - p_2 f_{21} - p_3 f_{31} - c_1)\Big(q_1 - f_{23}f_{32}\cdot q_1 + (f_{13}f_{32} + f_{12})\cdot q_2 + (f_{12}f_{23} + f_{13})\cdot q_3\Big)\Big] \Big/ det(I - A)$$

$$\pi_2 = \Big[(p_2 - p_1 f_{12} - p_3 f_{32} - c_2)\Big((f_{23}f_{31} + f_{21})\cdot q_1 + q_2 - f_{13}f_{31}\cdot q_2 + (f_{13}f_{21} + f_{23})\cdot q_3\Big)\Big] \Big/ det(I - A)$$

$$\pi_3 = \Big[(p_3 - p_1 f_{13} - p_2 f_{23} - c_3)\Big((f_{21}f_{32} + f_{31})\cdot q_1 + (f_{12}f_{31} + f_{32})\cdot q_2 + q_3 - f_{12}f_{21}\cdot q_3\Big)\Big] \Big/ det(I - A)$$

where $det(I - A)$ is again defined as above and where c_1, c_2, and c_3 have been added to capture any unit costs incurred by industries one, two, and three, respectively, that are outside of the three-industry economy.

1.2.2 Industry Configurations

There are 64 potential configurations of this three-industry network if we exclude any configurations with self-dependencies. Each of these configurations corresponds to choosing specific (non-diagonal) elements of the matrix T (equation 1) to be nonzero.

As with the computational model later in the chapter, my aim with the analytical variant is to explore the potential impacts of a shock to an industry j on some other focal industry i. As described above, the matrix T represents a directed graph such that arrows point from suppliers to their customers. Here and throughout the chapter, a supplier is considered to be "upstream" of an industry it supplies, while a customer is considered to be "downstream" of an industry from which it purchases. This terminology applies both to an industry's immediate suppliers and customers as well as those more removed, such that an industry's supplier's supplier is upstream of the industry, and an industry's customer's customer is downstream of the industry. In this way, tracing the path of arrows flows downstream in the network. An industry i may be downstream of a shocked industry j, upstream of a shocked industry j, or both.[6]

To develop intuition and motivation for the indicators to be proposed in section 1.5, I explore a particular configuration that illustrates the key transmission mechanisms of interest.

[6] Specifically, industry i is downstream of industry j if there is at least one directed path from j to i in the network. Similarly, industry i is upstream of industry j if there is at least one directed path from i to j in the network. If an industry i is both upstream and downstream of industry j, there is at least one cycle that includes industries i and j.

Example: Shock to the Oil Industry

I consider the effects of a supply-side shock to a hypothetical oil industry on which the three-industry economy in the model will be partially dependent. I choose to focus on the oil industry and its connections to the economy to provide some context for the empirical exercise in Chapter 2. Note that although I consider profits (π_1, π_2, and π_3 above) for illustrative purposes here, I consider both profits and revenues (x_1, x_2, and x_3 above) using the computational variant of the model.

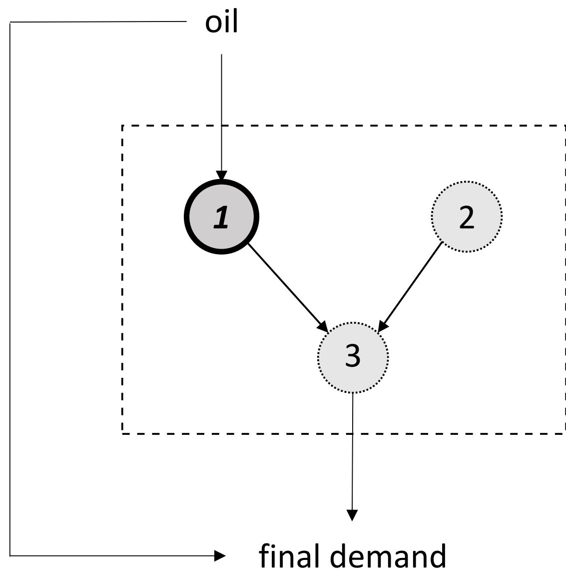
Suppose that industry three uses the outputs of industry one and industry two (and only those outputs) to make its product, and industries one and two do not depend on any industries within the economy (Figure 1.1):

$$T = \begin{bmatrix} 0 & 0 & f_{13} \\ 0 & 0 & f_{23} \\ 0 & 0 & 0 \end{bmatrix}$$

Take the final demand vector as $d' = [0 \quad 0 \quad d_3]$, such that industry three's good is a final good, and industries one and two produce intermediate goods. Although industries one and two may rely on inputs outside of the three-industry economy, they sell their outputs only to industry three. In addition, given that consumers have a limited budget and purchase some amount of oil in addition to industry three's good, q_3 is a function of p_{oil} in addition to p_3: $q_3(p_3, p_{oil})$. In this way, a supply-side oil shock can also have ramifications for final demand. Lastly, suppose that industry one directly uses some amount of oil while industries two and three do not.

In the derivation in the previous section, I abstracted from the price-setting mechanism and took the initial price vector, $p' = [p_1 \quad p_2 \quad p_3]$, as given. I retain that approach both here and in the computational variant of the model, and I take price-updating for each industry to be entirely dependent on the change in its total input cost, such that it passes through some portion of this input-cost change (whether positive or negative). Combined with the assumptions above, this implies that: p_1 is adjusted based on changes to p_{oil} and the details of industry one's production process (the function $g_1(\cdot)$ in the previous subsection); p_2 remains static; and p_3 is adjusted based on changes to p_1 and p_2 along with the details of industry three's production process (the function $g_3(\cdot)$ in the previous subsection). I also take the price elasticity of demand for industry three's good to be neither completely elastic nor completely inelastic, such that any adjustment to p_3 will result in some amount of positive, finite final demand for industry three's good.

Figure 1.1: Example Three-Industry Network Configuration.



Given that T (equation 1) has no cycles, A (equation 2) will also have no cycles, and so $det(I - A) = 1$. Using the results developed above, we have the following for industry profits:

$$\pi_1 = (p_1 - c_1) \cdot f_{13} \cdot q_3$$

$$\pi_2 = (p_2 - c_2) \cdot f_{23} \cdot q_3$$

$$\pi_3 = (p_3 - p_1 f_{13} - p_2 f_{23}) \cdot q_3$$

where f_{13} and f_{23} are functions of both p_1 and p_2 and where, following the discussion above, c_1 is a function of p_{oil} but c_2 is not. Following directly from the matrix T and the functions q_i, final demand for good three drives total production in this sub-economy, and industries one and two produce only to satisfy industry three's input demand. Financial flows enter the economy at industry three, which extracts profits equal to π_3, and then the remainder passes to industries one and two.

Suppose that p_{oil} increases exogenously by some small amount. We have the following for the change in industry one's profits:

$$\frac{\partial \pi_1}{\partial p_{oil}} = \underbrace{\frac{\partial p_1}{\partial p_{oil}} \cdot f_{13} \cdot q_3}_{\text{price pass-through effect}} + \underbrace{(p_1 - c_1) \cdot \frac{\partial f_{13}}{\partial p_1} \cdot \frac{\partial p_1}{\partial p_{oil}} \cdot q_3}_{\text{input substitution effect}} + \underbrace{(p_1 - c_1) \cdot f_{13} \cdot \frac{\partial q_3}{\partial p_{oil}}}_{\text{indirect final demand effect}} - \underbrace{\frac{\partial c_1}{\partial p_{oil}} \cdot f_{13} \cdot q_3}_{\text{direct effect}}$$

where I have simplified using $\frac{\partial p_2}{\partial p_{oil}} = 0$. There are potential effects due to: (1) industry one's pass through of the oil price; (2) industry three's changing usage of industry one's output based on industry one's price adjustment; (3) change in final demand for industry three's output; and (4) the direct effect of the shock.

We have the following for industry two's change in profits:

$$\frac{\partial \pi_2}{\partial p_{oil}} = \underbrace{(p_2 - c_2) \cdot \frac{\partial f_{23}}{\partial p_1} \cdot \frac{\partial p_1}{\partial p_{oil}} \cdot q_3}_{\text{input substitution effect}} + \underbrace{(p_2 - c_2) \cdot f_{23} \cdot \frac{\partial q_3}{\partial p_{oil}}}_{\text{indirect final demand effect}}$$

where I have simplified using $\frac{\partial p_2}{\partial p_{oil}} = 0$ and $\frac{\partial c_2}{\partial p_{oil}} = 0$. There are potential effects due to: (1) industry three's changing usage of industry two's output based on industry one's price adjustment; and (2) change in final demand for industry three's output.

Lastly, we have the following for industry three's change in profits:

$$\begin{aligned} \frac{\partial \pi_3}{\partial p_{oil}} = & \underbrace{\frac{\partial p_3}{\partial p_1} \cdot \frac{\partial p_1}{\partial p_{oil}} \cdot q_3}_{\text{price pass-through effect}} + \underbrace{\frac{\partial q_3}{\partial p_3} \cdot \frac{\partial p_3}{\partial p_1} \cdot \frac{\partial p_1}{\partial p_{oil}} \cdot (p_3 - p_1 \cdot f_{13} - p_2 \cdot f_{23})}_{\text{direct (substitution-type) final demand effect}} \\ & + \underbrace{\frac{\partial q_3}{\partial p_{oil}} \cdot (p_3 - p_1 \cdot f_{13} - p_2 \cdot f_{23})}_{\text{direct (income-type) final demand effect}} \underbrace{- q_3 \cdot \frac{\partial p_1}{\partial p_{oil}} \cdot \left[f_{13} + p_1 \cdot \frac{\partial f_{13}}{\partial p_1} + p_2 \cdot \frac{\partial f_{23}}{\partial p_1} \right]}_{\text{input substitution effect}} \end{aligned}$$

where I have simplified using $\frac{\partial p_2}{\partial p_{oil}} = 0$. There are potential effects due to: (1) industry three's pass-through of industry's one's price change; (2) change in final demand for industry three's output based on a substitution-type effect; (3) change in final demand for industry three's output based on an income-type effect; and (4) industry three's changing usage of the other two industries' outputs based on industry one's price adjustment.

It is worth considering in detail some of the implications of these results. First, given that consumers have a limited budget and purchase some amount of oil directly, suppose that final demand for industry three's good decreases in response to the shock via an income-type effect, such that $\frac{\partial q_3}{\partial p_{oil}} < 0$. In addition, suppose that consumers will substitute away from industry three's good (to final goods outside of the three-industry economy) so that $\frac{\partial q_3}{\partial p_3} < 0$.

With regards to price pass-through, suppose that industry one increases the price of its good by some amount when the oil price increases, such that $\frac{\partial p_1}{\partial p_{oil}} > 0$. The extent to which industry three increases its output price in response to an increase in the price of industry one's good is dependent on the price of industry two's good and its own production process. For simplicity, suppose that industry three cannot completely compensate for an increase in p_1 by shifting towards industry two's good (assuming the two goods are used as substitutes in industry three's production process; more below) and that the demand for its own good is not completely elastic. In this case, there will be some amount of price pass-through such that $\frac{\partial p_3}{\partial p_1} > 0$.

A final critical element is whether industry three uses the outputs of industries one and two as substitutes or as complements in its production process. If they are substitutes, then industry three will use less of industry one's output and more of industry two's output when producing a single unit of its own output, such that $\frac{\partial f_{13}}{\partial p_1} < 0$ and $\frac{\partial f_{23}}{\partial p_1} > 0$. If industry three instead uses these inputs as complements, then $\frac{\partial f_{13}}{\partial p_1} = \frac{\partial f_{23}}{\partial p_1} = 0$ because the quantities of each good required to produce a single unit of output remain fixed.

Overall, these dynamics can be summarized as follows: the impacts to industries one and three are a combination of interrelated supply and demand effects. The supply effect is that the increase in p_{oil} causes industry one to increase its output price, which in turn raises input costs for industry three. Although industry three may substitute towards industry two's output (assuming that industry three's production process exhibits substitutability in its inputs), it may also pass through some of the price increase to its own output price. All else equal, this will induce a demand effect, whereby final demand for industry three's product will decrease, in turn reducing the demand for industry one's product. Industry two will also experience this negative demand effect.

However, if industry two's output serves as a substitute to industry one's output in industry three's production process, then it will also experience a counteracting (positive) substitution effect as industry three shifts away from industry one's higher-priced output. If industry two's output is instead a complement to industry one's output in industry three's production process, then it will not experience this counterbalancing effect.

Table 1.1: Summary of Marginal Shock Effects for the Example Three-Industry Economy.

Effect	Industry One	Industry Two	Industry Three
Price Pass-Through	+	0	+
Input Substitution	– or 0	+ or 0	+ or 0
Final Demand	–	–	–
Direct Effect	–	0	0

See Table 1.1 for a summary of these dynamics, where +, –, and 0 denote whether each effect is marginally positive, negative, or not applicable, respectively, for the profits of each industry.

1.2.3 Key Transmission Mechanisms

This simple example illustrates several mechanisms by which a shock can have economic effects, which echoes some of the findings in the earlier literature:

- The supply effect traces forward through the production network from the shocked industry to final demand. This reflects pass-through of prices, and the magnitude of the propagation at each level is at least partially dependent on the ability of industries to shift away from their higher-priced inputs.

- An increase in the price of a final good may lead to both a consumption substitution effect and an income effect. This can be triggered directly if the shocked industry produces a final good, but can also be triggered indirectly by any final goods produced by industries on directed paths leading away from the shocked industry (assuming there is some amount of price pass-through along those paths). In all cases, the demand effect back-propagates through the network.

- An increase in the price of an intermediate good may also lead to substitution and income-type effects. If the buyer of the intermediate good is able to switch inputs to another intermediate good, then the producer of the first good experiences a negative substitution effect. At the same time, if the buyer of the intermediate good passes through the price increase, then any reduction in demand for the buyer's good also leads to a decrease in demand for all of its intermediate inputs.

- Industries that do not lie along directed paths from the shocked industry may experience counteracting forces. Of these industries: (1) those that produce a final good may experience a positive substitution effect if consumers switch towards their good; however, there may also be a negative income effect, especially if the shocked industry produces a final good; and (2) those that produce an intermediate good may experience a positive substitution effect as their customers substitute away from industries on the directed paths; at the same time, to the extent that the downstream buying industries experience a negative demand effect, the intermediate industries will also experience such an effect.

From the perspective of a specific industry i, these transmission mechanisms can be roughly classified along two dimensions. The first is the transmission of shocks from upstream (which impact entities via changes in input costs) versus the transmission of shocks from downstream (which impact entities via changes in demand). As has been discussed previously in the literature, supply shocks[7] tend to

[7] Here and in the computational model, I introduce a supply shock via an exogenous change in an industry's output price. In the literature, this type of shock is frequently modeled as a change in the productivity of a producing entity.

be propagated downstream while demand shocks tend to be propagated upstream (e.g., Acemoglu, Akcigit, and Kerr, 2016). At the same time, a supply-side shock for one industry can have—through the process of input substitution—a demand effect on its suppliers, which from the perspective of those suppliers appears to originate downstream.

The second dimension mirrors the discussion in the introduction, which breaks shock effects into two parts: those dealing with connectivity elements of the network (i.e., which industries are connected, directly and indirectly, to which others) versus those dealing with the evolutionary elements of the network (i.e., how existing connections change, especially as a short-term response to price variation). The connectivity side is most closely reflected in the pass-through and final demand effects in the analytical model, whereas the evolutionary side is most closely tied to the input substitution effect.

Altogether, combining these two dimensions suggests four different channels through which shocks may impact any given industry i. I propose and then explore indicators in each of these categories in section 1.5.

1.3 Production Processes as Networks

Before turning to the computational variant of the model, I propose a conceptualization of an industry's production process as a network of production steps. Although this conceptualization is potentially useful in a variety of contexts, for the purposes of this chapter, it facilitates the computational model in two ways.

First, it provides a framework that I can use to generate a random production process for each industry with an arbitrary number of inputs. This is the analog of the $g_i(\cdot)$ functions in the analytical model. I provide a solution algorithm that takes a price vector, p, and combines it with a production process to yield the required input quantities for the associated industry. This is the analog of the cost-minimization operation in the analytical model, which, when applied across the production processes of all industries, results in the matrix of $f_{ji}(\cdot)$ values.

Second, this framework allows every pair of inputs to be unambiguously classified as substitutes or complements, which is foundational for both the consideration of the indicators in section 1.5 and the evaluation of the input-categorization approach in section 1.6.

1.3.1 Framework

The production-process-as-network is captured by a directed, acyclic graph (G), where each node in the graph ($n \in N$) represents a production step (i.e., a scalar-valued function) that transforms multiple inputs into an output, and the graph edges ($e \in E$) capture which steps use the outputs of which other steps:

$$G = (N, E), \quad E \subseteq N \times N$$

$$\text{Node } n \in N \quad \sim \quad f_n : \mathbb{R}^k \to \mathbb{R} \ (k \geq 2)$$

$$\text{Edge } (\alpha, \beta) \in E \quad \Longleftrightarrow \quad f_\alpha \text{ is an argument to } f_\beta$$

Given that G is directed and acyclic, there is a set of one or more nodes that have in-degree of zero and another set of one or more nodes that have out-degree of zero. The former node(s) receive(s) the "primary" inputs to the production process (those inputs that originate external to the process), while the latter nodes(s) produce(s) the final output(s) of the process. The overall production of a process—and in turn, the primary inputs it requires—is driven by the total demand for the final output(s).

For the purposes of this chapter, I impose three main restrictions on this production process structure:

1. All nodes have only two inputs (i.e., $k = 2$ above);
2. The output of each node is used by one—and only one—other node (i.e., all nodes have out-degree of one);[8] and
3. The relationship between inputs and output at each node is captured by one of two types of constant-returns-to-scale production function: Cobb-Douglas or Leontief (i.e., the potential forms of the functions $f(\cdot)$ above are restricted; see below).

I make the first restriction for simplicity and the second to ensure that the overall production process has only one final output. As I discuss in more detail below, I make the third restriction for clarity in labeling each pair of inputs in the production process as either substitutes or complements.[9]

[8] With the exception of the node that produces the final output, which has out-degree of zero.

[9] There is also a computational side benefit to limiting nodes to be of these two types; see below.

The first and third restrictions imply that the function represented by a node α will be captured either by a two-input Cobb-Douglas formulation:

$$f_{\alpha,CD}(x_1, x_2) = A \cdot x_1^{a_1} \cdot x_2^{a_2}$$

where $a_1 + a_2 = 1$ so that the production node exhibits constant returns to scale, or by a two-input Leontief formulation:

$$f_{\alpha,L}(x_1, x_2) = B \cdot \min[b_1 \cdot x_1, b_2 \cdot x_2]$$

where $b_1 > 0$ and $b_2 > 0$. The network structure of the process leads to a composition of functions. As an example:

$$f_{\gamma,CD}(x_3, f_{\beta,L}(x_1, x_2))$$

where Leontief node β takes quantities x_1 and x_2 of its two inputs to produce a single output, and Cobb-Douglas node γ takes quantity x_3 of its first input and combines it with the output of node β (as its second input) to produce the final output of the process.[10]

I note that if the production function for each node has constant returns to scale, then the production process itself will have constant returns to scale (please see the appendix for a more detailed discussion). In turn, assuming constant returns to scale for each production node and therefore for the entire production process, I am able to focus on the ratio of inputs used to produce a single unit of output under given prices, similar to the approach in the analytical model.

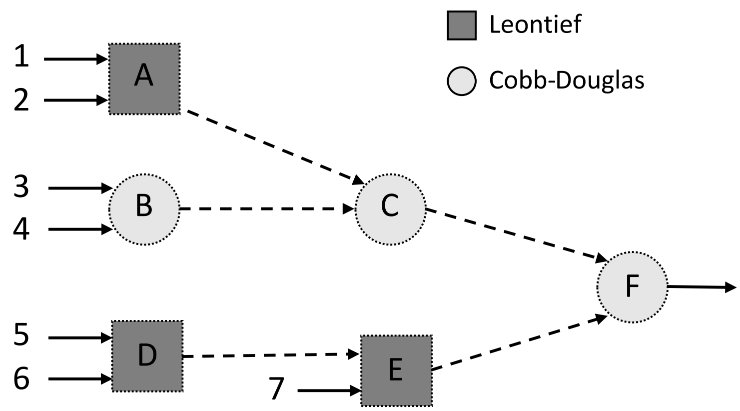
As mentioned above, I restrict the form of nodes' production functions to the Cobb-Douglas and Leontief types for clarity in labeling each pair of inputs in the production process as either substitutes or complements. Specifically, I define substitutability and complementarity of two inputs based on the type of the node that either: (1) uses the two inputs directly, or (2) is the first node, when traversing forward in the production process, that uses inputs that can be traced back to the two inputs in question (i.e., the highest common descendant node of the two production steps that use the inputs directly). If this node is of Cobb-Douglas type, I say that the two inputs are substitutes; if this node is of Leontief type, I say that the two inputs are complements.

As an example, take the production process of seven inputs and six production steps shown below (Figure 1.2). Inputs 1 and 2 are complements, as they are both used directly by the Leontief

[10] I describe in the appendix the specific parameter distributions I use to operationalize these forms in the computational model.

production step labeled A. Inputs 3 and 4 are substitutes, as they are both used directly by the Cobb-Douglas production step labeled B. Inputs 1 and 3 are substitutes, as their highest common descendant is Cobb-Douglas node C. Inputs 1 and 5 are also substitutes, as their highest common descendant is Cobb-Douglas node F.

Figure 1.2: Example Production Process with Seven Inputs and Six Production Steps.



Given that substitutability and complementarity are defined by the structure of the production process network, this approach removes any ambiguity in the classification of each pair of inputs. This is the case for any arbitrary process (within the restrictions outlined above) and regardless of the prices, or changes in prices, to which processes are exposed.

1.3.2 Solving for Input Quantities

As mentioned above, an industry's total output is driven by demand. In turn, I propose a solution algorithm that identifies required input quantities by searching over input combinations to find the minimum-cost option—under current prices—that satisfies given demand.

This is accomplished by working backwards recursively: the algorithm first chooses "trial" input quantities for the final production node that yield the required output of the production process; given these quantities, which now become the required outputs of the production nodes in the previous layer, the algorithm chooses trial input quantities for those nodes. This repeats until the beginning of the production process has been reached (that is, until all inputs have been assigned trial values). The overall cost associated with this trial input bundle is then calculated using the current vector of prices, where all intermediate inputs in the production process are assigned a cost of zero. If this is the first input bundle to be tested, it is tentatively chosen as the lowest-cost option. If not, the current cost is compared to the previously found lowest cost, and if the former is lower, the current bundle is tentatively chosen as the cost-minimizing set of inputs (otherwise, the previously found lowest-cost set of inputs remains the tentative best choice). This process repeats until all trial quantities have been checked, and whichever set of inputs was found to have the lowest overall cost is chosen as the minimum-cost option.

I note that, given the nature of the Leontief production function, the inputs to any Leontief node (regardless of whether its inputs are primary or not) can be calculated analytically based on the node's required output and the parameters of the Leontief function (B, b_1, and b_2 above) without reference to prices. In other words, for each output amount, there is only one pair of trial values for a two-input Leontief node: the quantities of the first and second inputs that yield that amount. This significantly helps to reduce the computational burden.

In addition, any Cobb-Douglas nodes at the very beginning of the process—those that use only primary inputs and no intermediate inputs—can be solved analytically without iterating over trial values because these nodes' inputs are all associated with prices. This also helps to reduce the computational burden.

This is not the case, however, for Cobb-Douglas nodes with one or more intermediate inputs, as those inputs are not associated with prices per se (these nodes are internal to the overall production process). In other words, unlike the Leontief case and the Cobb-Douglas case with only primary inputs, the entire set of trial input quantities for an internal Cobb-Douglas node must be tested for each of the particular trial values of the nodes downstream of it in the process. In turn, for a given number of these internal Cobb-Douglas nodes, the number of combinations tested for the whole production process grows exponentially with the number of trial values for each node. There are at least two options to limit this computational burden: one is to restrict the number of Cobb-Douglas nodes that appear internal to the process, and the other is to reduce the number of trial quantities tested at each node. I discuss both of these in more detail in the next section.

It is also necessary to determine what the specific trial values for these nodes will be. One complication is that the range of reasonable input values will differ across nodes depending on the specific parameters of their functional forms. In turn, I propose an approach that generates a list of trial values for each node such that the same total number of values are tested for every node but the range of those values varies by node. In my proposed approach, the extremities of these ranges are determined by the ratio of a node's inputs when making one input much smaller (or larger) than the other while holding output constant.

The general process for a node α with inputs x_1 and x_2 and production function $f_\alpha(x_1, x_2)$ is as follows (I present a simplification for two-input Cobb-Douglas nodes below):

1. Choose a value for x_1 and determine the value for x_2 implied by $f_\alpha(x_1, x_2) = 1$, and then calculate the ratio $\frac{x_2}{x_1}$; repeat this process using successively smaller values of x_1 until the ratio $\frac{x_2}{x_1}$ exceeds a pre-defined threshold; call the x_1 value at which this first occurs $x_{1,min}$;

2. Choose a value for x_1 and determine the value for x_2 implied by $f_\alpha(x_1, x_2) = 1$, and then calculate the ratio $\frac{x_1}{x_2}$; repeat this process using successively larger values of x_1 until the ratio $\frac{x_1}{x_2}$ exceeds some pre-defined threshold; call the x_1 value at which this first occurs $x_{1,max}$;

3. Finally, divide the interval $[x_{1,min}, x_{1,max}]$ into a pre-specified number of smaller intervals.

The endpoints of all of these intervals (including $x_{1,min}$ and $x_{1,max}$) are the set of x_1 values to be tested for the node α. Each trial x_1 value implies a trial x_2 value given $f_\alpha(x_1, x_2) = 1$. In an actual implementation, the input quantities and associated input prices would be scaled up based on the constant-returns-to-scale assumption. I describe in the appendix the specific thresholds and number of intervals I use in this algorithm to generate the results for the chapter.

Finally, I note that, in the case of a two-input Cobb-Douglas function with constant returns to scale (i.e., $f(x_1, x_2) = x_1^{a_1} \cdot x_2^{a_2}$ where $a_1 + a_2 = 1$), achieving the ratio $\frac{x_1}{x_2} = r$ has an analytical solution of $x_1 = r^{1-a_1} = r^{a_2}$ (and $x_2 = r^{-a_1}$). This expression for x_1 directly yields $x_{1,max}$ in step 2 above, while an analogous expression would yield $x_{1,min}$ in step 1. In turn, such expressions could replace the processes in steps 1 and 2, which would result in the same values for $x_{1,min}$ and $x_{1,max}$ but would require less computation.

1.3.3 Possible Extensions

Before concluding this section, I note that the overall approach could be extended in several ways, many of which correspond to relaxing one or more of the restrictions outlined above.

The first is to allow each production node to have more than two inputs. A second is to allow the overall production process to have more than a single output. In this latter case, a similar solution algorithm could be used, but some production nodes' output would be shared as inputs to multiple downstream production nodes.

A third is to specify the relationship between the inputs and output of a node in ways beyond the two-input Cobb-Douglas and Leontief formulations. Given that the solution algorithm is demand-based, any specification that allows for a mapping from required output to a comprehensive set of potential inputs could be used,[11] though this could significantly increase the time required to solve for input quantities.

Along these lines, a natural extension could be to replace the two-input Cobb-Douglas and Leontief production functions with two-input CES production functions. This would allow for a greater variety in the elasticity of substitution, and input pairs would be identified as having degrees of substitutability or complementarity based on the particular substitution parameter of the relevant descendant node. I discuss in the context of the input-categorization approach how this could facilitate an expansion of the method beyond binary classification of input pairs.

Lastly, it would be possible to build production processes where (in an analog of non-separability in traditional production functions) a particular input, such as x_1, enters the process at multiple nodes, though the substitutability/complementarity of inputs may not be well defined in such cases. I leave all of these modifications as potential avenues for future work.

1.4 Computational Model

The analytical variant of the model highlights the key channels through which a shock can differentially affect producing entities in a network context. However, it captures the dynamics among only three industries, and there are practical limitations to the number of network configurations and production processes that it can be used to explore.

In this section, I describe a computational variant of the model that is able to simulate economies of various sizes, network configurations, and production processes. Beyond exploring this variety, the computational model—by virtue of incorporating a larger number of industries—allows for shock dynamics to play out over multiple levels in the production network, which is a key focus of the chapter and cannot be adequately explored using the three-industry analytical version. As mentioned above, the computational variant takes the same basic form as the analytical variant but uses a range of specific functional forms and industry shocks to produce synthetic datasets that can be leveraged for additional analysis.

I describe the general features and algorithms of the model in the subsections below. I leave for the appendix the specific parameters of the model that generated the synthetic data on which the results in the next two sections are based.

1.4.1 Model Overview

At a high level, the model operates as follows:

[11] Theoretically, this includes algorithmic (rather than functional) definitions of nodes' production possibilities.

1. For each industry, randomly determine which other industries' goods it uses as inputs; this determines which elements of the matrix T (equation 1) will be non-zero. Then, using the production-process-as-network framework, randomly generate a production process for each industry. Combined with a specific vector of prices, p, this production process yields an industry's actualized values in the matrix T.

2. Assign initial final demand to each industry, which yields the final demand vector, d. Final demand for each industry responds to changes in that industry's output price.

3. For the economic configuration constructed in steps 1 and 2, repeat the following for a pre-specified number of rounds:

    (a) Choose a random price vector for the outputs of the industries in this economy. Calculate industry-level profits (some of which may initially be negative), and then repeatedly update pass-through prices and input demand until the round-to-round change in industry prices falls below a pre-defined threshold. In each round, if pass-through of cost changes for an industry is not sufficient to yield positive profits including a small markup, then increase the output price for the industry so that this condition is met.

    (b) Generate a supply-side shock, which also has a demand-side effect. For the supply-side, this translates into an exogenous change in the output price for a particular upstream industry. For the demand-side, this is an exogenous decrease in final demand for a specific downstream industry, which represents an income-type effect.

    (c) Allow the shocks to propagate through the network. To do this, again update pass-through prices and input demand over multiple rounds until the round-to-round change in industry prices falls below a pre-defined threshold, but do not require industries' profits to be positive.

    (d) Record changes to industry profits, revenues, input quantities, and output quantities alongside other key model parameters that generated this run. In addition, calculate and record a number of industry-level indicators (to be described in section 1.5) that correspond to features of the production network.

This overall process is re-run many times for various economic configurations ("configuration rounds") and initial price vectors ("price rounds"), which produces a synthetic dataset that describes industry changes along with a set of related parameters and indicators.

Before turning to the details of the model, I first describe the number of industries and the possible number of industry inputs I use to construct the synthetic data on which the empirical analyses in this chapter are based.

1.4.2 Number of Industries and Industry Inputs

As a reference point, I begin by examining the number of industries, the number of suppliers per industry, and the number of customers per industry in the U.S. Bureau of Economic Analysis' 3-digit-NAICS input-output tables. These tables reveal that the distribution of the number of suppliers tends to be normal while the distribution of the number of customers is skewed to the right.

For instance, in the year 1997, the number of suppliers ranged from 9 to 39 with a mean of about 25, while the number of customers ranged from 0 to 71 with (by construction) the same mean of approximately 25 (Table 1.2).[12] In other words, all industries have suppliers, and the number of these suppliers is grouped around a central value, whereas some industries have no intermediate customers (they sell only to final demand) and others have many customers (they are particularly important as an intermediate input). For the 1997 3-digit-NAICS data, the correlation between these two variables is low (-0.06).

Table 1.2: Number of Suppliers/Customers for 3-digit-NAICS Industries in 1997.

	# of Obs.	Mean	Std. Dev.	Min.	Max.
# of Suppliers	71	25.30	5.88	9	39
# of Customers	71	25.30	21.65	0	71

Based on this analysis, in the computational model I aim for a ratio of about 25-to-70 (approximately 0.35) between the average number of inputs per industry and the total number of industries, with a distribution that is normal. I also aim for a distribution in the number of customers that is skewed to the right (see the next subsection and the appendix for more details).

Linking the average number of industry inputs to the total number of industries introduces an important tradeoff between increasing the number of industries (for realism) while also constraining the computational burden. Specifically, per the framework introduced in the previous section, increasing the number of inputs for an industry increases the depth of the industry's production process network, and this depth increases the solution time for each industry quadratically rather than linearly (to the extent that the process contains internal Cobb-Douglas nodes).

[12] Note that in calculating these summary statistics, I only include each supplier-customer link if the link represents at least 0.5 percent of the customer's overall intermediate input purchases.

Therefore, for the purposes of the results in this chapter, I limit the size and the configuration of production processes in three ways:

1. Nodes have either two primary inputs or two internal inputs (and not one of each), which maximizes the number of nodes at each level of an industry's production process network;

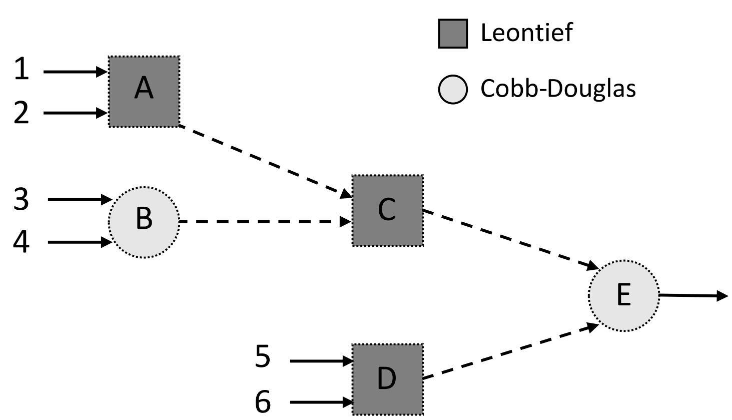
2. Processes have a maximum of eight inputs, which limits both (a) the overall depth of each process network to three and (b) the number of internal nodes to three; and

3. Processes can only contain one internal Cobb-Douglas node, which—combined with #1 and #2—transforms the relationship between the number of trial values per node and the solution computation time from quadratic to linear.

These restrictions greatly reduce the computational burden in producing synthetic datasets but also allow for reasonable variety in production processes and the number of industry inputs.

Figure 1.3 illustrates what one of these production processes might look like. This example process has six inputs and five production nodes, which are arranged in such a way to satisfy the three criteria described above. Specifically: (1) each node has either two primary inputs (nodes A, B, and D) or two internal inputs (nodes C and E); (2) the overall process has six inputs, yielding a network depth of three with two internal nodes (nodes C and E); and (3) only one of those internal nodes (E) is of Cobb-Douglas form.

Figure 1.3: Example Production Process with Six Inputs and Five Production Steps.



Under restriction #2 above, the possible number of inputs for any particular industry is two to eight, with a mean of five. In turn, for the results in this chapter, and using the ratio of 0.35 as described above, I set the computational model to simulate economies with a core set of 5 / 0.35 = 14 industries.

To represent the broader economy beyond this set of 14 core industries, I also include two additional industries that have no suppliers within the model economy, such that their output prices are fully exogenous. One of these industries I take to be the upstream "shocked" industry (via an exogenous change to its output price). The other industry is representative of the remaining part of the upstream, non-shocked economy, and in turn, I leave its output price constant.

Overall, the result is a model economy composed of 16 networked industries, 14 of which may depend on other industries within the model.

1.4.3 Model Details

I describe each part of the model in detail below.

Step 1: Economic Configuration and Production Processes

The economic configuration is determined by randomly choosing: (1) how many suppliers each industry will have; and (2) who those individual suppliers will be. I provide an overview of this process here and leave the details for the appendix.

At a high level, I first randomly determine how central each industry will be as a supplier by taking a draw from a shifted Pareto distribution. In doing so, I aim for there to be a larger number of industries with a few customers and a small number of industries with many customers. Using this draw, I assign each industry a threshold value (for use later in the process), where the higher an industry's centrality, the lower its threshold. The assigned centrality and threshold values for each industry remain constant throughout the remainder of the supplier/customer assignment process.

For each industry i, I then randomly determine how many suppliers it will have (for a total of two to eight). I always assign the upstream shocked industry to be one of i's suppliers for two reasons. The first is that I wish to focus on impacts to industries as they relate to the production network, and assigning the shocked industry as a supplier for every industry reduces the role of the direct effect as a source of variation. The second is that some of the indicators I consider are non-zero only for industries that have a direct dependence on the shocked industry. Assigning the shocked industry as a supplier for every industry allows for any differential effects predicted by these indicators to be interpreted as related to the indicators' magnitude rather than as a reflection of industries' direct dependence on or independence from the shocked industry.

Finally, I assign the remaining suppliers for each industry i by repeatedly selecting random

Algorithm 1 Industry Supplier/Customer Assignment

Parameters: $shape\_param$, $range\_end_{(1) \text{ to } (n)}$, $threshold\_val_{(1) \text{ to } (n)}$, avg_inputs, var_inputs, min_inputs, max_inputs, $candidate\_sep$, $candidate\_var$, and $num\_industries$.

1: For each industry i, determine centrality:
  a: Draw a random value, l_i, from the Pareto Type II distribution with shape parameter $shape\_param$.
  b: Assign a threshold value, t_i, to industry i as follows:
    If $l_i \in [range\_end_{(1)}, range\_end_{(2)})$, set t_i to $threshold\_val_{(1)}$;
    If $l_i \in [range\_end_{(2)}, range\_end_{(3)})$, set t_i to $threshold\_val_{(2)}$;
    ...
    If $l_i \geq range\_end_{(n)}$, set t_i to $threshold\_val_{(n)}$.
2: For each industry i, assign suppliers:
  a: Draw a random value, s_i, from the distribution $N(avg\_inputs, var\_inputs)$.
    [Check the following, and redraw if needed: $min\_inputs \leq s_i \leq max\_inputs$.]
  b: Set the industry *shocked* to be one of industry i's suppliers.
  c: Initialize a counter, $assigned\_suppliers_i$, to one.
  d: While $assigned\_suppliers_i < s_i$:
    I: Draw a random value, j, from the distribution $N(i - candidate\_sep, candidate\_var)$.
      [Check the following, and redraw if needed: $0 \leq j \leq num\_industries$; $j \neq i$.]
    II: If j was already assigned to be one of i's suppliers, go back to step (I).
    III: Draw a random value, u, from the uniform distribution over $[0, 1)$.
    IV: If $u \geq t_j$:
      i: Assign industry j to be one of industry i's suppliers.
      ii: Increment $assigned\_suppliers_i$ by one.

"candidates." For each candidate, I assign it as an actual supplier to i if: (1) it is not already a supplier to i; and (2) an additional random draw meets or exceeds the candidate's threshold value, which (as described above) is inversely proportional to how central the candidate was determined to be at the outset. In other words, the more central an industry is determined to be at the beginning of the process, the more likely it is that the industry will be assigned as a supplier when randomly chosen during another industry's assignment process.

Overall, this algorithm attempts to produce a network structure with a normal distribution in the number of suppliers and a right-skewed distribution in the number of customers. See Algorithm 1 for the pseudocode version of this process (additional details, as well as the parameters I use to generate the synthetic datasets for the chapter, are in the appendix).

With this done, the final part of this step is to generate each industry's production process. For a process with n inputs, I do this by generating $n-1$ production nodes, each of which is randomly of Cobb-Douglas or Leontief type and has parameters randomly drawn from pre-specified distributions (see the appendix for details). These nodes are arranged into a production process in accordance with the restrictions outlined in the previous subsection.

With the economic configuration and production processes in place, I am able to take a given price vector, p, and calculate the matrices T (equation 1), A (equation 2), and the Leontief inverse $(I - A)^{-1}$.

Step 2: Final Demand

I take each industry's initial final demand, d_i, as a random value drawn from a pre-specified uniform distribution (please see the appendix for details). After the economy is shocked, I adjust each industry's final demand based on the change in its output price:

$$d_{i,new} = d_i/(1 + \Delta p_i)^2$$

where $d_{i,new}$ is industry i's new final demand and Δp_i is the change in its output price pre-shock to post-shock. This yields an own-price elasticity of demand that is somewhat elastic and where $\partial d_i/\partial p_i < 0$.

With this in place, I am able to take any change to the price vector, Δp, and combine it with the old final demand vector, d, to a produce a new final demand vector, d_{new}.

Step 3(a): Initial Price Vector

I perform steps 3(a) to 3(d) in a loop, the result of which is a set of synthetic observations for the particular economic configuration created in steps 1-2. In this way, steps 1-2 create a "configuration round" that is associated with multiple "price rounds." The entire process (steps 1 through 3) can be repeated to generate observations across different economic configurations.

In step 3(a), I find a suitable price vector that, via the mechanics of the model, implies values for industry profits and industry revenues. These serve as the baseline against which the effects of the supply shock and demand shock will be compared.

To find this price vector, I first choose a random price for each industry, p_i, drawn from a pre-specified uniform distribution. I use the vector of prices to solve for each industry's minimum-cost set of inputs, and then if necessary, I update each industry's output price so that it covers its total input cost plus a pre-specified markup. Using this new price vector, I repeat the cost-minimization and pass-through process until the round-to-round change in industries' output prices falls below a given threshold, or after a maximum number of rounds, whichever comes first. I verify at the end that all industries have positive profits. See Algorithm 2 for the pseudocode version of this process (the parameters I use for the results are shown in the appendix).

The notion behind this approach is to explore the n-dimensional price space without restricting

Algorithm 2 Creation of Initial Price Vector

Parameters: p_{min}, p_{max}, max_rounds, $pass\_through\_rate$, $markup$, and $change\_threshold$.

1: Initialize a price vector p with elements p_i that are drawn randomly from $[p_{min}, p_{max}]$.
2: Initialize a binary flag, $all\_below\_threshold$, to FALSE.
3: Initialize a counter, $round\_count$, to zero.
4: While ($all\_below\_threshold$ = FALSE) and ($round\_count < max\_rounds$):
    a: Use p to solve for industries' required inputs.
    b: Save each industry's current total cost as c_i.
    c: If $round\_count = 0$:
        I: If $p_i < c_i \cdot (1 + markup)$:
            Set $p_{i,new} = c_i \cdot (1 + markup)$.
    d: If $round\_count > 0$:
        I: Calculate the change in each industry's cost from the previous round and store it as Δc_i.
        II: Set industry i's new price, $p_{i,new}$, to $p_i + pass\_through\_rate \cdot \Delta c_i$.
        III: If $p_{i,new} < c_i \cdot (1 + markup)$:
            Set $p_{i,new} = c_i \cdot (1 + markup)$.
        IV: Calculate each industry's price change as $\Delta p_i = (p_{i,new} - p_i)/p_i$.
        V: If $\Delta p_i < change\_threshold$ for all i:
            Set $all\_below\_threshold$ to TRUE.
    e: Replace the old price vector, p, with the new price vector, p_{new}.
    f: Increment $round\_count$ by one.
5: Use p to solve for industries' required inputs and generate the matrix T.
6: Use T to construct the matrix A, which can be used to calculate $(I - A)^{-1}$.
7: Calculate industry outputs/revenues as $X = (I - A)^{-1} \cdot d$.
8: Use X to calculate industry profits, Π.

prices to only those that represent equilibria. Over multiple iterations of steps 3(a)-3(d), the aim is to determine how various shocks propagate through and affect the network when beginning at various locations within the price space.

Step 3(b): Supply-Side and Demand-Side Shocks

I generate both a supply-side shock and a demand-side shock. The supply-side shock is an exogenous change in the output price of a single upstream industry (called it *shocked*) by a factor chosen randomly from {2,3}.

The demand-side shock is an exogenous change in final demand for a single downstream industry (call it *downstream*), which is proportional to the upstream shock. Specifically, I calculate the new final demand by reducing the original final demand as follows:

$$d_{downstream,new} = d_{downstream} \cdot (1 - (\Delta p_{shocked}/4))$$

where $d_{downstream,new}$ is the new final demand, $d_{downstream}$ is the original final demand, and $\Delta p_{shocked}$ is the random change in the shocked industry's output price described above. This results

Algorithm 3 Shock Propagation

Parameters: max_rounds, $pass\_through\_rate$, and $change\_threshold$.

1: Initialize a binary flag, $all\_below\_threshold$, to FALSE.
2: Initialize a counter, $round\_count$, to zero.
3: While ($all\_below\_threshold$ = FALSE) and ($round\_count < max\_rounds$):
  a: Use p to solve for industries' required inputs.
  b: Save each industry's current total cost as c_i.
  c: Calculate the change in each industry's cost from the previous round and store it as Δc_i.
  d: Set industry i's new price, $p_{i,new}$, to $p_i + pass\_through\_rate \cdot \Delta c_i$.
  e: Calculate each industry's price change as $\Delta p_i = (p_{i,new} - p_i)/p_i$.
  f: If $\Delta p_i < change\_threshold$ for all i:
    Set $all\_below\_threshold$ to TRUE.
  g: Replace the old price vector, p, with the new price vector, p_{new}.
  h: Increment $round\_count$ by one.
4: Use p to solve for industries' required inputs and generate the matrix T.
5: Use p to update final demand and generate the vector d.
6: Use T to construct the matrix A, which can be used to calculate $(I - A)^{-1}$.
7: Calculate industry outputs/revenues as $X = (I - A)^{-1} \cdot d$.
8: Use X to calculate industry profits, Π.

in an income-type effect where $\partial d_{downstream}/\partial p_{shocked} < 0$.

Step 3(c): Shock Propagation

In this step, I allow the supply-side shock to propagate through the network, resulting in a new set of industry prices. In essence, step 3(a) locates the networked economy at a certain point within the price space, step 3(b) directly perturbs one of these prices (and adjusts final demand for a single downstream industry), and this step moves the economy to a new point in price space given the general logic of the model as well as the particular elements of the current configuration. The movement of the economy from one set of prices to another is accompanied by changes in final demand, industry profits, and industry revenues.

To accomplish this updating, I perform the same process as described in step 3(a) but with two differences: (1) instead of being created, the initial price vector is now given; and (2) I do not require that industries' profits remain positive. After the round-to-round change in industries' output prices falls below a pre-specified threshold, or after a maximum number of rounds of updating (whichever comes first), the economy is at a new point in the price space. I use the new set of prices to update final demand, which can then be used to calculate industry revenues and profits. See Algorithm 3 for the pseudocode version of this process (the parameters I use to generate the synthetic datasets for the chapter are shown in the appendix).

Step 3(d): Synthetic Data Generation

This step records details of the previous steps to create a synthetic dataset. The information stored falls into two categories: (1) details of the economic configuration created in steps 1-2, including indicators that describe how industries are related to one another (to be described in the next section); and (2) details about the current shocks as simulated in steps 3(a)-3(c), including the change in profits and revenues experienced by each industry due to the shocks.

Over many rounds, the storage of this information results in a synthetic dataset that describes the impacts of shocks for various economic configurations and for multiple trial price vectors as simulated within each configuration.

1.5 Indicators

The analytical model suggests that, from the perspective of a specific industry i, shocks are transmitted via several mechanisms that can be roughly classified along two dimensions: (1) those that transmit shocks from upstream versus those that transmit shocks from downstream; and (2) those dealing with the connectivity elements of the network versus those dealing with the evolutionary elements of the network.

The intersection of these two dimensions suggests potential industry indicators across four categories:

1. Upstream/connectivity: the number, type, and/or configuration of the connections upstream of each industry i;

2. Downstream/connectivity: the number, type, and/or configuration of the connections downstream of each industry i;

3. Upstream/evolutionary: the manner in which the connections upstream of each industry i evolve in response to a shock, and especially the substitutability and complementarity of the inputs of those upstream industries; and

4. Downstream/evolutionary: the manner in which the connections downstream of each industry i evolve in response to a shock, and especially the substitutability and complementarity of the inputs of those downstream industries.

The first and second types of indicators capture how supply and demand shocks propagate downstream and upstream, respectively, to industries via some combination of direct, pass-through, and final demand effects. The third and fourth types of indicators reflect how substitution effects, on the upstream and downstream sides of industries, respectively, combine to affect industries over multiple levels in the production network (see more below).

In subsection 1.5.2, I propose six indicators that aim to capture these dynamics: one each for categories #1 and #2 and two each for categories #3 and #4. As mentioned earlier in the chapter, the indicators I propose are related directly and indirectly to the Leontief inverse, and I describe them in those terms as well as in terms of a network view of the economy.

1.5.1 Matrices P, R, and C

As a precursor to constructing the six indicators, I define three matrices—P, R, and C—that are calculated either from the direct requirements table (A as defined by equation 2 in the analytical model) or from the non-price-adjusted "quantity" requirements table (T as defined by equation 1 in the analytical model). See the left side of Figure 1.4 for a diagram of how P, R, and C relate to the matrices A and T. I describe the construction of these three matrices briefly here and provide details in the appendix.

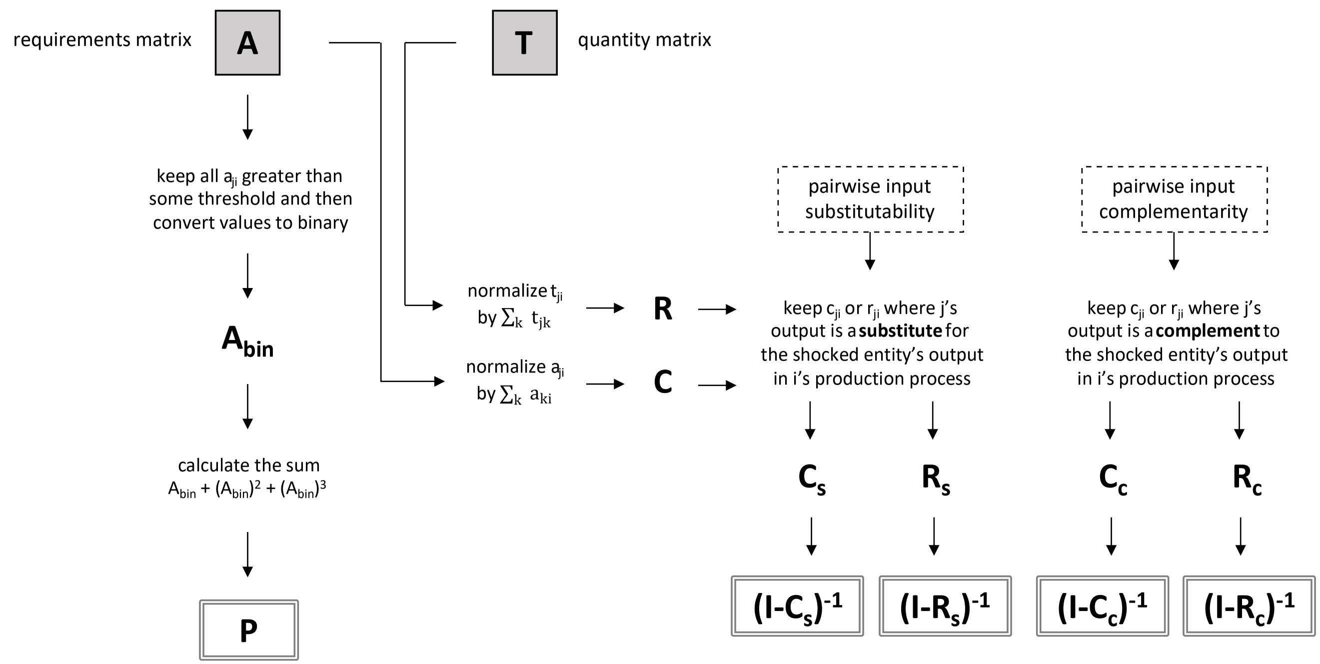
The first is P, with elements p_{ji}, that represent the number of walks[13]—of length one, two, or three—from each industry j to every other industry i. In determining the number of walks, I consider only those elements of A that exceed a pre-defined threshold. The matrix P intends to capture, for each industry i, the economically relevant connectivity to every other industry j in both the upstream direction (p_{ji}) and the downstream direction (p_{ij}). Said another way, the elements of this matrix roughly capture the "number of ways" it is possible to traverse the network to reach one industry from another.

The second matrix is R, with elements r_{ji}, that reflect the relative proportion of industry i's output (in quantity terms) used by industry j, based on the quantity of industry i's output used by each of its customers to produce one unit of their own output (i.e., R is a row-normalized version of T). The matrix R intends to capture, for each industry i, the importance of every other industry j as a direct customer of i. I base the matrix R on T (rather than on A) because it is the relative quantities purchased by each of i's customers that are most relevant from i's perspective.

The last matrix is C, with elements c_{ji}, that represent the relative expenditure by industry i (in dollar terms) on the output of industry j, based on i's expenditures on all of its suppliers' outputs to produce one dollar of its own output (i.e., C is a column-normalized version of A). The matrix C intends to capture, for each industry i, the importance of every other industry j as a direct supplier

[13] Note that a walk is distinguished from a path in that the former allows for repeated nodes in traversing the graph of the production network from one industry to another.

Figure 1.4: Calculation of Indicator Matrices from Requirements/Quantity Matrices.



Note: this figure illustrates the construction of the matrices P, $(I - C_s)^{-1}$, $(I - R_s)^{-1}$, $(I - C_c)^{-1}$, and $(I - R_c)^{-1}$ from which the proposed indicators are constructed. Specifically: (1) the matrix P is used to create the supply and demand walks indicators; (2) the matrices $(I - C_s)^{-1}$ and $(I - C_c)^{-1}$ are used to construct the upstream substitute and complement indicators, respectively; and (3) the matrices $(I - R_s)^{-1}$ and $(I - R_c)^{-1}$ are used to construct the downstream substitute and complement indicators, respectively.

to i. I base the matrix C on A (rather than on T) because it is i's relative expenditures on its inputs (rather than the quantities of those inputs) that are most important from its perspective.

1.5.2 Indicator Definitions

In this subsection, I propose six indicators to be constructed for each industry i, which together look to reflect both the connectivity and the evolutionary aspects of shock dynamics in networked economies. For illustrative purposes, I refer to the upstream shocked industry as *shocked*. I refer to the downstream industry that experiences an income-type final demand effect as *downstream*.

Before describing these indicators, I first define two additional indicators based on the classic Leontief inverse, which I use for comparison purposes in the analysis.

Upstream and Downstream Elements of Leontief Inverse

These two indicators aim to capture the upstream supply shock from *shocked* and the downstream demand shock from *downstream* via the linkages represented within the Leontief inverse. These indicators are similar to those used in Acemoglu, Akcigit, and Kerr (2016), with the exception that I do not sum industry i's column in the Leontief inverse when creating the upstream indicator nor do I sum its row when creating the downstream indicator.[14] I instead take a single element from the column and a single element from the row because, in the computational model, the connections between the shocked industries (*shocked* and *downstream*) and all other industries are already explicitly accounted for within the model.

In turn, for the upstream indicator, I take the element of industry i's column in the Leontief inverse associated with *shocked*:

$$\text{up\_leontief\_inv}_i = \Big[(I-A)^{-1}\Big]_{shocked,i}$$

where A is the direct requirements table for the economy before the upstream and downstream shocks have been applied.

For the downstream indicator, I take the element of industry i's row in the Leontief inverse associated with *downstream*:

$$\text{down\_leontief\_inv}_i = \Big[(I-A)^{-1}\Big]_{i,downstream}$$

[14] Note that I use the terms "upstream" and "downstream" here to denote the origin of the shock, whereas Acemoglu, Akcigit, and Kerr (2016) use these terms to indicate where the shock effect is experienced.

where A is again the direct requirements table before shocks have been applied.

I use these indicators for two purposes in the analysis. The first is as a comparison to the connectivity upstream and downstream indicators I propose below. The second is as a "control" for the substitutability and complementarity indicators (also proposed below), in the sense that I want to test if substitutability and complementarity input dynamics are statistically significant even when controlling for these elements of the Leontief inverse.

Upstream and Downstream Walks to Shocked Industries

The two walks indicators I propose here are alternatives to the indicators constructed from the Leontief inverse. As described above, the Leontief inverse captures the direct and indirect connections (of all lengths) between every pair of industries in the economy, taking into account the magnitudes of these connections as present in the table from which the inverse is calculated.

Here I make two modifications to this approach. The first is to convert the direct requirements table from numerical values to binary values, keeping only those links that exceed a pre-defined threshold. This yields a matrix that captures whether each industry is, or is not, connected directly to every other industry in an economically relevant way. The second is to use this matrix to consider walks of length three or less rather than walks of any length. This yields the matrix P. I limit the walk length so that the values in P reflect the immediate connections in the vicinity of each industry and with less cyclicality than would be incorporated with walks of longer length.

The first indicator is then the same as the upstream Leontief indicator above, except using the matrix P rather than the Leontief inverse constructed from the original direct requirements table:

$$\text{supply\_walks}_i = p_{shocked,i}$$

This indicator aims to differentiate industries by serving as a proxy for the transmission of the supply shock from the upstream direction. It does so by reflecting the relative "number of ways" that a shock to *shocked* can propagate when tracing arrows downstream in the network.

Analogously, the second indicator is the same as the downstream Leontief indicator above, except using the matrix P rather than the Leontief inverse based on the original direct requirements table:

$$\text{demand\_walks}_i = p_{i,downstream}$$

This indicator aims to differentiate industries by serving as a proxy for the transmission of the demand shock from the downstream direction. It does so by reflecting the relative number of ways that i is a supplier, directly or indirectly, to *downstream* when tracing arrows in the network.

The overarching purpose in proposing these two connectivity indicators is to present an alternative way of reflecting, in a reduced-form manner, the static structure of the network at any given point in time. In being constructed from a matrix of binary values, these indicators emphasize the role of the existence or non-existence of connections as a transmitter of shocks. They also have the advantage of being intuitive and requiring only binary information to calculate.

Upstream Substitutability and Complementarity of Inputs

These indicators intend to capture how input substitutability and complementarity on the upstream side of each industry i either dampens or transmits the supply shock as it propagates downstream towards i. Said another way, these indicators reflect how input substitution moderates the pass-through effect as a shock moves from upstream to downstream.

Although there are potentially many ways in which these types of upstream evolutionary indicators could be constructed, in this chapter, I specifically consider how upstream industries use *shocked* alongside their other inputs, and how patterns in such usage combine over multiple levels in the production network to influence the transmission of the shock.

As an illustration, suppose that industry i uses two inputs that are substitutes for one another: *shocked* and the output of one other industry. In the case that $p_{shocked}$ increases, industry i is able to (all else equal) substitute towards its other input, reducing the magnitude of the increase in its total production cost and, in turn, reducing the amount of its resulting price pass-through. If industry i instead uses *shocked* as a complement to its other input, this method of cost reduction is not available.

A similar dynamic holds for industry i's suppliers' usage of their own inputs (vis-à-vis *shocked*) as well as for other suppliers further upstream. The notion is that, as these interactions layer one on top of another, the initial upstream shock will either be dampened by chains of substitutability or, conversely, transmitted by chains of complementarity. The two indicators I propose in this subsection aim to measure the magnitude of each of these types of chains.

To construct the indicators, I create two intermediate matrices based on the matrix C defined above (recall that C reflects the relative importance of j as a direct supplier to i). The first, which I call C_s, takes C and retains only those elements where *shocked* is used as a substitute for other inputs (setting the other elements to zero). For example, if industry i uses industry j's output as a substitute for *shocked*, then the element c_{ji} is retained in the matrix C_s. If industry i instead uses industry j's output as a complement to *shocked*, then the element c_{ji} is set to zero in the matrix C_s.

I construct the second matrix, C_c, in an analogous fashion, except that the conditions are reversed: an element c_{ji} is retained in the matrix C_c if industry i uses industry j's output as a complement to *shocked*. Otherwise, the corresponding element in C_c is set to zero. See the right side of Figure 1.4 for a diagram of how C_s and C_c relate to C.

Taking these components together, the "upstream substitutes" indicator is then the weighted product of all walks on the upstream side of industry i, considering only the edges in the network that represent inputs used as substitutes for *shocked*, and where the weights are the relative importance of inputs viewed from the perspective of each customer. In terms of the Leontief inverse, the indicator is the sum of the i-th column (excluding the entry for i) of the Leontief inverse constructed using the matrix C_s:

$$\text{up\_subs}_i = \sum_{k \neq i} \left[(I - C_s)^{-1} \right]_{k,i}$$

As described above, this indicator aims to differentiate industries based on the substitutability of *shocked* and other inputs on the upstream side of each industry. I expect that a greater degree of substitution would lead to less shock transmission on the upstream side.

The "upstream complements" indicator is the weighted product of all walks on the upstream side of industry i, considering only the edges in the network that represent inputs used as complements to *shocked*, and where the weights are (again) the relative importance of inputs viewed from the perspective of each customer. In terms of the Leontief inverse, the indicator is the sum of the i-th column (excluding the entry for i) of the Leontief inverse constructed using C_c:

$$\text{up\_comp}_i = \sum_{k \neq i} \left[(I - C_c)^{-1} \right]_{k,i}$$

This indicator aims to differentiate industries based on the complementarity of *shocked* and other inputs on the upstream side of each industry. I expect that a greater degree of complementarity would lead to relatively greater shock transmission on the upstream side.

I note that, by construction, these indicators will be non-zero only for industries that have a direct dependence on *shocked*.

Overall, these indicators move from a connectivity view to an evolutionary view of the network by calculating the Leontief inverse not from a single (original or modified) input-output table—as did the Leontief and walks indicators above—but from proxies of the "marginal changes" to the entries of the input-output table (as reflected in C_s and C_c) in the presence of a shock.

Downstream Substitutability and Complementarity of Inputs

The final two indicators intend to capture how input substitutability and complementarity on the downstream side of each industry i either increases or decreases demand for industry i's output. Although changes in prices flow from upstream to downstream in the network, these indicators capture how input substitution back-propagates to effect demand moving from downstream to upstream. Similar to above, I consider how downstream industries use *shocked* alongside their other inputs, and how patterns in such usage combine over multiple levels in the production network to influence demand for industry i's output.

As an example, suppose that industry i has one customer and that customer uses industry i's output as a substitute for *shocked*. All else equal, an increase in $p_{shocked}$ will cause the customer to shift towards industry i's output, resulting in a positive demand effect. If, instead, the customer uses industry i's output as a complement to *shocked*, this effect is not present.

A similar dynamic holds for this customer's customers' usage of inputs (vis-à-vis *shocked*) as well as for other customers further downstream. The notion is that, as these interactions layer one on top of another, demand for industry i's output will either be relatively increased by chains of substitutability or, conversely, relatively decreased by chains of complementarity. The indicators I propose here aim to measure the magnitude of each of these types of chains.

To construct these indicators, I create two intermediate matrices based on the matrix R defined above (recall that R reflects the relative importance of j as a direct customer of i). The first, which I call R_s, is constructed in the same way as the matrix C_s described above, except that the source matrix is R rather than C. Similarly, the second matrix, R_c, is constructed in the same way as the matrix C_c, except that the source is R rather than C. See the right side of Figure 1.4 for a diagram of how R_s and R_c relate to R.

The "downstream substitutes" indicator is then the weighted product of all walks on the downstream side of industry i, considering only the edges in the network that represent inputs used as substitutes for *shocked*, and where the weights are the relative importance of customers' usage of outputs viewed from the perspective of each supplier. In terms of the Leontief inverse, the indicator is the sum of the i-th row (excluding the entry for i) of the Leontief inverse constructed using R_s:

$$\text{down\_subs}_i = \sum_{k \neq i} \left[(I - R_s)^{-1} \right]_{i,k}$$

This indicator aims to differentiate industries based on the substitutability of *shocked* and other inputs on the downstream side of each industry. I expect that a greater degree of substitution would lead to a greater (beneficial) effect on the demand side.

The "downstream complements" indicator is the weighted product of all walks on the downstream side of industry i, considering only the edges in the network that represent inputs used as complements to *shocked*, and where the weights are (again) the relative importance of customers' usage of outputs viewed from the perspective of each supplier. In terms of the Leontief inverse, the indicator is the sum of the i-th row (excluding the entry for i) of the Leontief inverse constructed using R_c:

$$\text{down\_comp}_i = \sum_{k \neq i} \left[(I - R_c)^{-1} \right]_{i,k}$$

This indicator aims to differentiate industries based on the complementarity of *shocked* and other inputs on the downstream side of each industry. I expect that a greater degree of complementarity would lead to a lesser counteracting (or possibly negative) effect on the demand side.

I note that, by construction, these indicators will be non-zero only for industries that have at least one customer that depends directly on *shocked*.

As with the upstream substitutability and complementarity indicators, these downstream indicators capture an evolutionary view of the network by calculating the Leontief inverse from the anticipated changes to the input-output table (as proxied by R_s and R_c) in the presence of a shock.

1.5.3 Approach to Indicator Assessment

The goal of this subsection and the next is to leverage synthetic data from the computational model to assess if and how each indicator predicts differences in outcomes among industries that are subjected to the same shock. To accomplish this, I analyze the data using a fixed-effects linear regression model that groups synthetic observations into panels based on the combination of a particular configuration round and a particular price round.

The idea is that each configuration round-price round combination is analogous to a single shock in a real-world economy at a particular point in time. The coefficients in the regression model reflect how the indicators—by capturing various connectivity and evolutionary aspects of the network—differentiate industries from one another in the presence of the same shock. In this way, the approach is focused on heterogeneity of outcomes within a shock event rather than attempting to identify the effects across various shocks per se.

The regression framework is as follows:

$$
\begin{aligned}
y_{ir} = &\alpha + \beta_{up\_leontief} \cdot \log(\text{up\_leontief\_inv}_{ir}) + \beta_{down\_leontief} \cdot \log(\text{down\_leontief\_inv}_{ir}) + \\
&\beta_{supply} \cdot \log(\text{supply\_walks}_{ir}) + \beta_{demand} \cdot \log(\text{demand\_walks}_{ir}) + \\
&\beta_{up\_subs} \cdot \log(\text{up\_subs}_{ir}) + \beta_{up\_comp} \cdot \log(\text{up\_comp}_{ir}) + \\
&\beta_{down\_subs} \cdot \log(\text{down\_subs}_{ir}) + \beta_{down\_comp} \cdot \log(\text{down\_comp}_{ir}) + \\
&\beta_{\lambda} \cdot \lambda_{ir} + \gamma_r + \epsilon_{ir}
\end{aligned}
$$

where y_{ir} is the percent change in profits or revenues for industry i within configuration round-price round r, α is a constant, the variables in logs are the indicators for industry i as defined above (within the configuration round-price round r), λ_{ir} is a vector of two controls (to be discussed below), γ_r is a fixed-effect for configuration round-price round r, and ϵ_{ir} is the error term.

I take logs of the indicators because most of their distributions are skewed to the right (see below).[15] I leave the dependent variable, y_{ir}, as it is because it contains both positive and negative values.

I run the regression twice for each dependent variable (profits and revenues): once with the Leontief indicators and the substitutability/complementarity indicators together, and once with the walks indicators and the substitutability/complementarity indicators together.

Controls

I propose two additional, binary regressors to control for the direct effect of the upstream shock and for two simplifications made within the computational model. These simplifications are: (1) that the substitutability and complementarity of inputs are completely random, and in turn, are unrelated both to the relative importance of inputs (from the perspective of each industry) and to inputs' probability of being shocked; and (2) that industries pass through their cost increases at a constant rate regardless of the elasticity of demand for their output.

The two regressors defined below attempt to control for these dynamics. In the next subsection, I present the results including these controls, and in the appendix, I present the same regressions with and without the controls. Although the results are qualitatively similar in both cases, the controls aid in the interpretability of the results by separating out these factors from the indicators.

[15]Similar to MaCurdy and Pencavel (1986), I add a small constant to the independent variables to ensure that they are positive before taking the log. For each variable, I set this constant equal to the smallest non-zero value divided by eight. This allows for the retention of the zero-valued observations in the regressions while not substantially changing their economic meaning (given that, in a real-world application, we would expect zero and close-to-zero values for these indicators to reflect similar relationships).

The first control reflects whether or not an industry has a substantial, direct dependence on the upstream shocked industry. Specifically, I consider an industry to have such a dependence if its entry for *shocked* in the direct requirements table is greater than (or equal to) the average of such entries across all observations in the synthetic dataset:

$$\text{direct} \geq \text{average}_i = \mathbf{1}[a_{shocked,i} \geq \overline{a_{shocked}}]$$

where A is the direct requirements table for each artificial economy (within each price round) before the upstream and downstream shocks have been applied. By dividing the observations into two groups (low dependence and high dependence on the shocked industry), this control allows for easier interpretation of the evolutionary indicators (see the appendix for additional discussion).

The second control is a proxy for whether the intermediate demand for an industry is relatively elastic. Specifically, I set this control to one for any industry that has zero downstream chains of complementarity (i.e., has only downstream chains of substitutability):

$$\text{zero down\_comp}_i = \mathbf{1}[\text{down\_comp}_i = 0]$$

Although these chains directly reflect complementarity only with regards to the shocked industry, they indirectly reflect other patterns of complementarity due to the way that production processes are constructed. In this way, the control serves as a proxy for the ability of each industry's customers to substitute away from its output.

1.5.4 Results of Indicator Assessment

Table 1.3 presents summary statistics for the synthetic dataset. This dataset was generated by running the computational model for 300 configuration rounds of 2 price rounds each (for 14 firms), yielding 8,400 observations. I choose a large ratio of configuration rounds to price rounds so that the synthetic dataset reflects a variety of economic configurations rather than many initial price vectors within each configuration. The indicators tend to be right-skewed, with fewer industries exhibiting the largest values.

Below I present the regression results (Table 1.4). To aid in comparing the indicators, I standardize all variables (including the dependent variables) before running the regressions. In turn, each coefficient can be interpreted as the effect on profits or revenues—in fractions of one standard deviation—based on a one-standard-deviation change in the (log of the) associated indicator.

I also present the linear combinations of log(up_subs) - log(up_comp) and log(down_subs) -

Table 1.3: Summary Statistics for the Synthetic Dataset.

Variable	Mean	Std. Dev.	Min.	Max.
up_leontief_inv	0.107	0.108	0.004	0.959
down_leontief_inv	0.053	0.213	0.000	6.817
supply_walks	2.800	2.373	0.000	14.000
demand_walks	1.175	1.649	0.000	11.000
up_subs	0.801	0.425	0.000	2.438
up_comp	0.519	0.479	0.000	5.322
down_subs	1.177	1.142	0.000	19.000
down_comp	0.728	1.189	0.000	19.000

log(down_comp) based on the values of those coefficients in each regression. These linear combinations reflect how differences in chains of substitutability and complementarity on industries upstream and downstream sides, respectively, contribute to the effects they experience in the presence of the shocks.

Upstream and downstream elements of Leontief inverse. The coefficients for these indicators are negative and statistically significant in all of the regressions in which they appear. This is consistent both with theory and with previous research. The magnitudes of the *up_leontief_inv* coefficients are larger than those for *down_leontief_inv*, especially in the profit regression. This may be due to the fact that the magnitude of the supply shock is larger than the magnitude of the demand shock (in terms of percentage change; see section 1.4.3), though more research would be needed to understand this fully.

Number of supply walks to shocked *and number of demand walks to* downstream. The coefficients on these indicators are negative and statistically significant in three of the four regressions in which they appear. Unlike the Leontief inverse indicators, the statistically significant *supply_walks* coefficient is closer in magnitude to the coefficients for *demand_walks*.

When comparing the results for the Leontief indicators to the results for the walks indicators, I note that the magnitudes of the *down_leontief_inv* and *demand_walks* coefficients are similar to one another, while the *up_leontief_inv* coefficients (and especially the one in the profit regression) are quite a bit larger in magnitude than the statistically significant *supply_walks* coefficient. As above, more research would be needed to understand these differences.

Upstream substitute and complement walks. I am interested in upstream chains of substitutability and complementarity both separately and together, where the latter is captured by the linear combination of log(up_subs) - log(up_comp).

For profits, these indicators reveal that a greater number/magnitude of upstream substitutability

Table 1.4: Regression Results Using the Synthetic Dataset.

	(1) profits	(2) profits	(3) revenues	(4) revenues
log(up_leontief_inv)	-0.484***		-0.133***	
	(0.0225)		(0.0168)	
log(down_leontief_inv)	-0.0341***		-0.0531***	
	(0.00977)		(0.0159)	
log(supply_walks)		0.0101		-0.0320**
		(0.0100)		(0.0147)
log(demand_walks)		-0.0535***		-0.0932***
		(0.0128)		(0.0128)
log(up_subs)	0.0585***	0.0848***	0.000434	0.00708
	(0.0202)	(0.0219)	(0.0123)	(0.0125)
log(up_comp)	-0.149***	-0.119***	-0.000712	0.00524
	(0.0120)	(0.0132)	(0.0133)	(0.0134)
log(down_subs)	-0.00107	0.00135	0.294***	0.287***
	(0.0102)	(0.0100)	(0.0174)	(0.0167)
log(down_comp)	0.0395	0.0276	-0.136***	-0.150***
	(0.0351)	(0.0377)	(0.0352)	(0.0352)
direct $\geq$ average	-0.117***	-0.751***	-0.0866**	-0.239***
	(0.0288)	(0.0439)	(0.0345)	(0.0296)
zero down_comp	0.0869	0.0548	-0.578***	-0.598***
	(0.0829)	(0.0899)	(0.0820)	(0.0819)
log(up_subs) - log(up_comp)	0.207***	0.204***	0.00115	0.00185
	(0.0196)	(0.0212)	(0.0150)	(0.0151)
log(down_subs) - log(down_comp)	-0.0406	-0.0263	0.430***	0.437***
	(0.0314)	(0.0340)	(0.0312)	(0.0312)
N	8,400	8,400	8,400	8,400
adj. R^2	0.199	0.101	0.114	0.112

Standard errors clustered by configuration round-price round.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

chains is associated with greater profits, while a greater number/magnitude of upstream complementarity chains is associated with smaller profits. The linear combination is also positive and statistically significant for both profit regressions. The magnitude of this difference is larger (in absolute value) than the coefficients for $down\_leontief\_inv$ and $demand\_walks$, and about half the magnitude of the coefficient for $up\_leontief\_inv$.

For revenues, the upstream substitutability/complementarity indicators are all statistically insignificant. This is true for the linear combinations as well.

Downstream substitute and complement walks. As with the upstream substitutability and complementarity indicators, I am interested in downstream chains of substitutability and complementarity both separately and together, where the latter is captured by the linear combination of log(down_subs) - log(down_comp).

For revenues, these indicators suggest that a greater number/magnitude of downstream substitutability chains is associated with greater revenues, while a greater number/magnitude of downstream complementary chains is associated with smaller revenues. The combination log(down_subs) - log(down_comp) is also positive and statistically significant for both revenue regressions. The magnitude of this difference is larger (in absolute value) than the coefficients for all of the connectivity indicators: $up\_leontief\_inv$, $down\_leontief\_inv$, $supply\_walks$, and $demand\_walks$.

For profits, the downstream substitutability/complementarity indicators are all statistically insignificant. This is the case for the associated linear combinations as well.

Controls. Finally, the indicator for substantial direct dependence on the upstream shocked industry is negative and statistically significant in all regressions. The proxy for elastic intermediate demand is negative and statistically significant for both of the revenue regressions. As might be expected, this latter finding suggests that industries are less able to capture the benefits of output price increases when their downstream customers can easily substitute to other inputs.

Summary

The patterns in these results suggest a few takeaways. The first is that the numbers of upstream and downstream walks can serve as a proxy for industries' structural exposure to shocks on their upstream and downstream sides, respectively. In this sense, they may be able to play a differentiating role similar to that of elements of the Leontief inverse.

I note that one potential drawback of these connectivity indicators as compared to the Leontief indicators is that—in relying on binary information—the former may be more sensitive to the inclusion or exclusion of certain edges. This might be particularly relevant in cases where the production network exhibits clustering and where the clusters are connected by a small number of links. In these cases, measurement error or other factors that lead to the inclusion or exclusion of certain bridging links could have a substantial impact on the indicator values. Additional research, including larger simulated networks that exhibit more variation in structure, could provide insight into the robustness of the results for these indicators. Such analyses may also shed light on the differences in coefficient magnitudes described above.

A second takeaway is that chains of substitutability and complementarity on industries' upstream and downstream sides can combine, over multiple levels in the network, to significantly contribute to the effects that industries experience in the presence of upstream and downstream shocks, respectively. The analysis provides evidence that the part of the shock effect due to these short-term evolutionary dynamics can be similar in magnitude to the part of the effect resulting from the connectivity aspects of the network, even when controlling for those latter aspects. I also note that the coefficient estimates for these indicators tend to be stable regardless of whether I include the Leontief indicators or the proposed connectivity indicators. Additional research would be needed to understand why the difference between the substitutability and complementarity coefficients on the downstream side is about twice the corresponding difference on the upstream side.

The results for the evolutionary indicators also echo the analytical model in suggesting which mechanisms might be at work. Specifically, given that (1) industries' profits are a function of both supply- and demand-side factors (i.e., input costs as well as demand), whereas (2) revenues are a function of only demand, we would expect the pass-through effect to be relatively greater for profits and the demand effect to be relatively greater for revenues. This is what the results suggest: the upstream substitutability and complementarity indicators are statistically significant for profits but not for revenues, which provides evidence that these indicators are reflecting how input substitution on the upstream side of industries is moderating pass-through; conversely, the downstream substitutability and complementarity indicators are statistically significant for revenues but not for profits, which suggests that these indicators are capturing how input substitution is being back-propagated through the network to impact demand.

Lastly, given that all of the evolutionary indicators are based on a binary categorization of input pairs as substitutes or complements, their significance provides evidence that such binary information can be relevant even without regard to the degree of substitutability or complementarity (i.e., without information about elasticity).

1.6 Input-Categorization Approach

In this section, I use the computational model to test an approach that categorizes each pair of inputs in an industry's production process as substitutes or complements.[16]

Specifically, during step 3(d) of each model run, I record the input quantity changes for each industry (based on the shocks that occurred during that run) alongside the substitutability or complementarity of each pair of inputs in industries' production processes as defined earlier in the chapter. I then use the approach to infer, using only the quantity change data, the substitutability or complementarity of inputs. I evaluate the method by comparing the categorizations with the true relationship for each input pair.

1.6.1 Description of Approach

The data source for the input-categorization approach (the "training dataset") is a group of observations that specifies for each industry within each combination of configuration round and price round: (1) how much (in percentage terms) the industry's use of its inputs changed in that price round; and (2) how much (also in percentage terms) its output quantity changed in that price round. Given that production processes change from one configuration round to the next, the training dataset for an industry is the set of quantity changes across all of the price rounds within a particular configuration round. For example, if there are two industries, two configuration rounds, and ten price rounds, there will be four training datasets of ten observations each: one for each industry from the first configuration round and one for each industry from the second configuration round.

Given that all production processes exhibit constant returns to scale, I first subtract the percentage change in an industry's output quantity within a price round from the input percentage changes in that price round. The notion is that an increase or decrease in demand for an industry's output will—in the absence of changes to the prices of its inputs—affect its usage of inputs in an equal way. Subtracting the output quantity change from each of the input quantity changes aims to provide the approach with the relative changes among the inputs. In the next subsection, I also present results without this adjustment (using the same synthetic dataset) as a point of comparison.

Each of the industry training datasets is used as the input to the approach as outlined in Algorithm 4 (the parameters I use for testing purposes are shown in the next subsection). To determine whether each pair of inputs are substitutes or complements, the algorithm repeats the same process multiple times, each time taking a different input x as the "focus" input.

[16] As described in more detail below, the algorithm may not (in certain cases) classify an input pair as either substitutes or complements.

Algorithm 4 Approach to Categorizing Pairwise Substitutability/Complementarity of Inputs

Parameters: min_test_val, max_test_val, $test\_val\_increment$, and $scaling\_factor$.

1: Select a focus input $x \in X$, where X is the set of inputs for the given industry.
2: Train the XGBoost algorithm to predict the changing usage of x given the simultaneous changes in the other input quantities as well as the change in the industry's output.
3: Perform the training via cross-validation, storing the RMSE across the folds as r.
4: For each other input $y \neq x$:
  a: For each test value $t \in [min\_test\_val, max\_test\_val]$ in increments of $test\_val\_increment$:
    I: Predict the change in the industry's usage of x given a t percentage change in its usage of input y, assuming no change in its use of other inputs or of its output.
    II: Record the test value t alongside the predicted change in the usage of x.
  b: Regress the predicted values against the test values and a constant, and store the estimated coefficient c for the test values term.
  c: If $|c| \geq r/scaling\_factor$:
    I: Take the inputs x and y to be complements in the production process if $c > 0$.
    II: Take the inputs x and y to be substitutes in the production process if $c < 0$.
  d: If $|c| < r/scaling\_factor$, take y to be neither a complement nor a substitute for x in the industry's production process.
5: Repeat the steps from the beginning with another focus input x.

Specifically, for a given focus input x, I train the eXtreme Gradient Boosting (XGBoost) algorithm to predict the changing use of x based on the simultaneous changes in the other input quantities and in the output quantity. This is akin to estimating the marginal rate of technical substitution, but for multiple pairs of inputs simultaneously and where the algorithm may use the industry's changing output quantity as a predictor. This training is performed via cross-validation, storing the root-mean-square-error (RMSE) across the folds. This RMSE represents the relative strength of the trained model in understanding the relationship between x and the other inputs and output, which may be stronger (smaller RMSE value) for some focus inputs and weaker (higher RMSE value) for other focus inputs.

I choose a machine learning-based approach both (1) because my focus is on prediction and (2) to minimize assumptions about the relationships among the inputs and output (for example, that the usage of one input is a linear combination of the usage of the other inputs). I leverage gradient boosted trees, and the XGBoost algorithm in particular, for several reasons. I first note that my intention in training each model is to capture the relationships among the inputs of a particular production process rather than to predict input usage in other production processes or to learn something about production processes more generally. Given that gradient boosting works to reduce the bias of the final model by iterative refinement over multiple rounds, it is a natural fit for this purpose as compared to bagging ensemble methods, such as random forests, that are more targeted towards a reduction in variance (though all ensemble methods aim to reduce both bias and variance). I choose gradient boosted trees over artificial neural networks as the former are often used for tabular data, though it would be a valuable exercise in future work to replicate the approach using the latter. Finally, as gradient boosted trees can be prone to over-fitting, I choose the XGBoost algorithm specifically as it includes multiple regularization terms as part of its objective function. I tune the XGBoost hyperparameters using grid search.

Once I have a trained model for a given focus input x, the approach examines the pairwise relationship between x and every other input $y \neq x$. Specifically, the trained model is used to predict how the use of x will change if each other input y is used much less, is used somewhat less, is used more, is used much more, etc. The result is again loosely akin to the marginal rate of technical substitution, but now for each pairwise combination of inputs x and y in isolation. The algorithm fits a line through these predicted values and examines the resulting slope in two steps. In the first, it checks to see if the magnitude of the slope is larger than a scaled version of the RMSE. If it is, then in the second, it checks the sign of the slope and classifies x and y as complements if the sign is positive or as substitutes if the sign is negative.

The idea is that a positive slope represents a relationship between x and y where a large increase in the use of y corresponds to a large (predicted) increase in the use of x, a moderate increase in the use of y corresponds to a moderate (predicted) increase in the use of x, and so on. I interpret this as a complementary relationship between x and y. Conversely, when the slope is negative, it represents a relationship between x and y where a large increase in the use of y corresponds to a large (predicted) decrease in the use of x, a moderate increase in the use of y corresponds to a moderate (predicted) decrease in the use of x, and so on. I interpret this pattern as x being substitutable for y in this industry's production process.

In either case, the sign of the slope is only used for classification if the magnitude of the slope is large enough relative to the RMSE, which acknowledges that the predictions used in calculating the slope contain a certain amount of error. If the slope is not large enough relative to the RMSE, then the inputs x and y are not classified as either substitutes or complements. Adjusting the factor used to scale the RMSE allows the algorithm to be more or less conservative when producing its categorizations (see subsection 1.6.2 for related discussion and results).

Note that, because the XGBoost algorithm performs feature (regressor) selection in addition to estimation of a functional relationship, there are cases where a change in industry use of an input y (regardless of the magnitude of that change) leads to zero predicted change in the use of a focus input x. In these cases—along with those cases where the slope is not large enough—I do not consider input y to be either a complement or a substitute to input x in the production process.

Altogether, at the end of the algorithm, there are up to two estimates of the substitutability or complementarity between every pair of inputs: one when the first input is the focus input and one when the second input is the focus input (note there may be no estimates or only one estimate for a pair given the conditions outlined above). For the purposes of evaluation, if there is only one estimate for a pair of inputs, I take that to be the estimate for the pair (i.e., only one estimated c value is used to determine whether the inputs are substitutes or complements). If there are two estimates for a pair of inputs, I take the average of the two c values (call it c_{avg}) and then I observe whether c_{avg} is positive or negative to determine the categorization for the pair.

1.6.2 Results of Approach

Below I present results that compare the categorizations produced by the approach with the true substitutability or complementarity of inputs.

These results are based on a synthetic dataset of 30 configuration rounds composed of 20 price rounds each.[17] In this way, the approach produces its categorizations based on 20 input/output quantity-change observations for each industry. I present below the parameters I use within the algorithm for the test (Table 1.5). As in the previous two sections, the number of inputs for each industry ranges from two to eight.

Table 1.5: Parameter Values for Testing the Input-Categorization Approach.

Parameter	Value
min_test_val	-0.5
max_test_val	0.5
$test\_val\_increment$	0.02
$scaling\_factor$	4

As mentioned above, I test the approach using two versions of the same synthetic dataset: one with the original quantity change values, and one where the input quantity changes for an industry (within each price round) are adjusted by that industry's output quantity change in that round.

Results without Adjustment for Output Quantity Change

Of the total number of input pairs across all industries across all rounds, the approach categorized approximately 58 percent of the pairs as being either substitutes or complements. The remainder of the input pairs were not categorized one way or the other because, as described above, the condition

[17]This yields the same number of observations as the synthetic dataset used to the test the indicators in the previous section, but with a much larger number of price rounds within each configuration round.

$|c| \geq r/scaling\_factor$ did not hold or because of regressor selection (see Algorithm 4 for details). Of the input pairs that were identified as being either substitutes or complements, about 75 percent were categorized correctly.

If the input pairs are grouped by the total number of inputs in the production processes of which they are a part, the percentage categorized decreases with the total number of inputs (Table 1.6). However, the percentage categorized correctly does not exhibit a clear downward pattern, with values in the range of about 65-85 percent. In other words, a greater number of total inputs for an industry appears to decrease the algorithm's ability to identify input pairs as being either substitutes or complements, but when it does categorize input pairs, the algorithm's accuracy is relatively consistent regardless of a production process' total number of inputs.

Table 1.6: Percentage Categorized/Correct by Industry Number of Inputs (Without Adjustment).

Industry # of Inputs	% Categorized	% Correct
2	100.0%	86.2%
3	85.5%	78.7%
4	77.8%	66.7%
5	57.7%	80.2%
6	53.2%	76.4%
7	52.6%	70.6%
8	44.6%	82.0%

Finally, I note that when there are two coefficient estimates for an input pair (one when the first is the focus input and the other when the second is the focus input), the two coefficients have the same sign over 99 percent of the time. These input pairs (with two estimates) make up about 53 percent of all input pairs with at least one estimate. When considering only those pairs where the two estimates have the same sign, approximately 76 percent of the pairs were categorized correctly.

Results with Adjustment for Output Quantity Change

When using the output-quantity-adjusted version of the synthetic dataset, the approach identified about 42 percent of the input pairs as being either substitutes or complements. Of these input pairs, approximately 94 percent were categorized correctly. This suggests that adjusting for output quantity change decreases the ability of the algorithm to categorize input pairs one way or another, though it increases the rate at which categorizations are correct (see additional discussion below).

As above, if the input pairs are grouped by the total number of inputs in the production processes of which they are a part, the percentage categorized decreases with the total number of inputs (Table 1.7). Reflecting an overall categorization rate of 42 percent here versus 58 percent using the nonadjusted version of the dataset, the percent categorized for each group is also lower here than the corresponding value in the table above (e.g., 100 percent of input pairs in 2-input processes were categorized when not adjusting for output quantity, whereas 86.2 percent of input pairs in 2-input processes were categorized when making such an adjustment).

Table 1.7: Percentage Categorized/Correct by Industry Number of Inputs (With Adjustment).

Industry # of Inputs	% Categorized	% Correct
2	86.2%	100.0%
3	56.6%	98.9%
4	48.1%	92.9%
5	42.2%	99.5%
6	41.9%	91.1%
7	37.9%	91.4%
8	33.3%	89.9%

The percentage categorized correctly exhibits somewhat of a downward trend, with higher values at the top of the table and lower values at the bottom. However, as compared to the percentage categorized correctly when not adjusting for output quantity change, the percentages presented here are more closely grouped and are substantially higher: approximately 90-100 percent.

Lastly, when there are two coefficient estimates for a given input pair, those estimates have the same sign over 99 percent of the time. These input pairs (with two estimates) make up about 47 percent of all input pairs with at least one estimate. When considering only those pairs where the two estimates have the same sign, approximately 94 percent were categorized correctly.

Results with Variations to the Scaling Factor

The scaling factor in Algorithm 4 reflects how conservative the algorithm will be in determining whether an input pair has a large enough regression coefficient to be considered an actual estimate. Specifically, the smaller the value for $scaling\_factor$, the larger the coefficient c will need to be—relative to the RMSE—for c's sign to be used to categorize an input pair as either substitutes or complements. If the coefficient c is not large enough, then the algorithm does not categorize the associated input pair as either substitutes or complements.

As shown above, I use a scaling factor of four for the main results. Here, I present the performance of the algorithm—as measured by the number of input pairs categorized and, of those, the number categorized correctly—when using scaling factors that range from 0.5 to 32 (Table 1.8).

As expected, increasing the scaling factor from 0.5 to 32 increases the percentage of input pairs that are categorized but also decreases the algorithm's accuracy. Specifically, over these values

Table 1.8: Approach Results Using Various Scaling Factors.

	Without Adjustment		With Adjustment	
Value of $scaling\_factor$	% Categorized	% Correct	% Categorized	% Correct
0.5	44.0	78.9	36.1	96.4
1	50.7	77.0	38.8	95.4
2	54.8	76.2	40.6	94.5
4	57.6	75.4	41.9	93.7
8	59.8	75.1	43.0	93.0
16	61.6	74.9	43.6	92.5
32	62.9	74.7	44.3	92.0

for the scaling factor, the percentage categorized ranges from about 44 to 63 percent when not adjusting for output quantity change and from approximately 36 to 44 percent when making such an adjustment. The percentage categorized correctly ranges from about 79 down to 75 percent in the former case and from about 96 down to 92 percent in the latter case.

Overall, these numbers suggest two takeaways: (1) that the magnitude of the scaling factor appears to have a greater impact on percentage categorized than on percentage categorized correctly; and (2) that the impact on percentage categorized is more important when not making an adjustment for output quantity change.

1.6.3 Possible Extension: Elasticities of Substitution

In section 1.3 on production processes as networks, I note that the framework could be extended to allow for more variety in the elasticity of substitution by, for example, allowing nodes to be of CES form rather than only Cobb-Douglas and Leontief. The degree of substitutability or complementarity for each pair of inputs would then correspond to the CES substitution parameter for the relevant node (either the node that uses the two inputs directly or the highest common descendant).

The input-categorization algorithm presented above could likewise be modified to produce an elasticity estimate for each pair of inputs rather than just a binary categorization of pairs as substitutes or complements. Although there are potentially many ways to introduce such a modification, one natural extension could be to interpret the slope coefficient c in Algorithm 4 as reflecting the degree of substitutability or complementarity between each pair of inputs. For instance, a negative slope would still indicate that two inputs are substitutes, but the steepness of the slope (in absolute value) would be taken to reflect the elasticity of substitution.

This interpretation could be tested as follows: if the production-processes-as-networks framework were extended as described above, the synthetic dataset produced by the model could record the elasticity parameter associated with every pair of inputs. The input-categorization approach could then be applied to the quantity-change datasets in the same way as in section 1.6.2, but the accuracy of the results would be evaluated based not only on the sign of c for each pair but also on its magnitude vis-à-vis the recorded elasticity for that pair. I leave this extension for future work.

1.6.4 Summary

The results indicate that when the approach categorizes input pairs as either substitutes or complements, it does so with an accuracy rate of about 65-85 percent when not adjusting by output quantity change and with an accuracy rate of approximately 90-100 percent when making such an adjustment. In the first case and somewhat in the second case, these rates are generally uncorrelated with the total number of inputs in a production process.

At the same time, the results provide evidence that the approach is better able to categorize inputs one way or another (as opposed to not at all) when: (1) production processes have fewer total inputs; and (2) when not adjusting industries' input quantity changes by their output quantity change. There are at least three potential reasons for these findings, two of which are related to properties of the computational model and the synthetic datasets it produces.

The first reason—unrelated to the synthetic data—is that as the number of inputs in a production process grows, the larger the number of input pairs there are (ranging from one pair in a two-input process up to 28 pairs in an eight-input process). In turn, as the total number of inputs in a process grows, the greater the quantity of information that needs to be extracted from the observations to categorize the same percentage of input pairs. However, in the results above, there are 20 observations for every industry regardless of its total number of inputs, which is likely one reason why the percent categorized decreases in the number of inputs.

The second reason is that some input pairs are more exposed to price variation than others. Specifically, any input pair that includes the shocked industry is (by construction) exposed to price variation resulting directly from the shock, whereas all other input pairs are exposed to price variation only to the extent that prices are transmitted through the network. In turn, we might expect the algorithm to be able to more easily categorize input pairs that include the shocked industry, and this is the case: when considering only such pairs, the algorithm categorizes 88 percent and 81 percent when not adjusting and when adjusting, respectively, for output quantity change.

The third reason is that, due to the fact that the synthetic dataset contains no measurement error, the adjustment by output quantity change produces some input quantity changes that are zero. For example, take a simple, two-input Leontief production process: in this case, the input quantities will change by exactly the same amount as the output quantity, and so the adjustment will transform the original input quantity changes from non-zero values to zeroes. In turn, the approach is not able to identify a relationship between such inputs.

In sum, all three of these factors likely contributed to the results regarding the percentage of input pairs categorized, though the latter two may be less relevant in an empirical context.

I also note that, as mentioned above, when there are two coefficient estimates for an input pair, these two estimates nearly always have the same sign. When considering only those pairs where there are two estimates and where they have the same sign, the accuracy rate is very close to the rate when considering all pairs. This suggests that the approach's performance is similar regardless of whether it produces one or two estimates for a given input pair, though more research would be needed to understand why there are two estimates (versus one) in about 50 percent of cases.

Finally, it is also important to note that the external validity of the results in this section (in terms of correlating with the approach's performance in empirical settings) is dependent on the extent to which the framework and the synthetic data mirror actual production processes. For instance, one additional dynamic that could be integrated into the computational model—and then used to further test the input-categorization approach—is to allow production nodes that do not exhibit constant-returns-to-scale, which could impact the effect of adjusting input quantity changes by output quantity change.

Overall, given the results presented above, the approach appears to be relatively accurate in categorizing input pairs as either substitutes or complements, and especially so when the input quantity changes have been adjusted by the output quantity change. Although adjusting input quantities in this way appears to reduce the number of input pairs that are categorized one way or another, this is at least partially due to features of the synthetic dataset. For all of these reasons, and to prioritize accuracy when taking the approach to actual data, I use the variant with input quantities adjusted in Chapter 2.

1.7 Conclusion

In this chapter, I take a computational approach to exploring the propagation of shocks in production networks with a particular emphasis on considering, and distinguishing, the connectivity and the evolutionary components of shock effects. In doing so, I make three main contributions to the existing literature.

The first is to develop a computational model of production networks that can simulate a range of artificial economies with a focus on heterogeneity in producing entities' usage of inputs. A key part of the model is a conceptualization of production processes as networks, which provides a method to generate a variety of random production processes while also explicitly capturing the substitutability or complementarity between every pair of inputs. I use the model in the current chapter as a way to evaluate the indicators and the input-categorization approach, though the model has potential applications in other contexts as well.

The second main contribution of the chapter is the indicators themselves, which aim to capture both the connectivity and the evolutionary parts of shock effects in networked economies. I consider the connectivity portion via two walks indicators: one reflecting the transmission of supply-side shocks from upstream to downstream, and the other reflecting the transmission of demand-side shocks from downstream to upstream. These walks indicators can be thought of as variations on the classic Leontief inverse, in the sense that they are based on the input-output table but consider connections in a binary (rather than numerical) way and incorporate walks of length three or less (rather than walks of any length). Intuitively, these indicators represent the "number of ways" that a supply- or demand-side shock can be transmitted to industries.

I consider the evolutionary portion of shock effects via four indicators: two that reflect patterns of input substitutability and complementarity on the supply side, and two that reflect patterns of input substitutability and complementarity on the demand side. As with the walks indicators, these evolutionary indicators can also be thought of as modifications to the Leontief inverse. However, instead of considering the arrangement of network connections at a given point in time, these indicators aim to capture the potential marginal changes to the entries of the underlying input-output table in the presence of shocks. Intuitively, the upstream indicators represent the ability (or inability) of industries to shift away from their higher-priced inputs as a supply shock propagates downstream, while the downstream indicators represent the positive (or negative) back-propagation of demand as entities adjust their usage of intermediate inputs.

Using synthetic data produced from the computational model, I conduct an analytical exercise to assess the degree to which these indicators differentiate industries in the presence of shocks. The results suggest that the indicators are able to capture both the connectivity and the evolutionary components of shock effects, while simultaneously providing evidence that the evolutionary portion can be significant even when controlling for the connectivity portion.

The final main contribution of the chapter is to propose an approach to estimating the pairwise substitutability or complementarity of producing entities' inputs. I test the approach using synthetic data from the computational model and find that it is generally accurate in its categorizations.

This approach may have a wide variety of applications, including the construction of the proposed indicators in an empirical context.

Overall, via the model I construct, the indicators I propose, and the input-categorization approach I develop, I attempt to bridge the static Leontief inverse on one hand with the role of input substitution and complementarity on the other. I do so using a computational framework that is both conceptual and methodological, with the overarching goal of producing results that have empirical applicability.

Chapter 2

Oil Price Episodes and the U.S. Economic Production Network

2.1 Introduction

Oil price fluctuations have received substantial attention from researchers, policymakers, and the public. One of the reasons is that energy price increases—including shocks to the price of crude oil—are seen as fundamentally different than increases in the prices of other goods and services (Kilian, 2008a). This difference in perception is partially due to aspects of energy price fluctuations themselves, such as the rapidity and duration of increases, as well as attribution of increases to events beyond the U.S. economy. Such a view is also due to the associated consumer and economic effects, including impacts resulting from relatively inelastic consumer demand for energy, and to the apparent negative, causal relationship between energy price increases and macroeconomic indicators.

A large literature has investigated the economic impacts of oil price fluctuations. Although our understanding of these historical episodes and our methods for analyzing them have improved over the years, research continues to shed light on the relationship between oil prices and the broader economy. Baumeister and Kilian (2016) state the following:

> "[Oil price transmission mechanisms] also have implications for the debate about climate change and for environmental policies. Of particular interest is the question of how the effects of oil price shocks vary across industries, plants, and households, how long these effects last, and how they may have changed over time. One important insight of the recent literature has been that these questions cannot be answered without taking account of the underlying causes of the oil price shock" (pp. 157-158).

In this chapter, I use a production networks view of the economy to explore the relationship between manufacturing industries' economic performance and oil price fluctuations in the United States over the half-century between 1968 and 2018. In doing so, I aim to investigate two primary questions. First, when an oil price episode occurs, how does an industry's place in the production network mediate the impacts it feels? Second, can we "see" oil price episodes through the lens of differentiated industry impacts, and if so, what does this tell us about the episodes themselves?

I make two main contributions to the literature in exploring these questions. The first is to consider oil price episodes within a framework that explicitly takes account of the direct and indirect linkages among industries. This builds on the existing literature not only by providing additional evidence about the heterogeneity of industry-level effects, but by considering the specific, network-based mechanisms through which oil-related supply and demand shocks propagate in the U.S. economy. I find that the production network is relevant for industry outcomes and that these differentiated outcomes provide insight into oil episodes (see below).

The second main contribution of the chapter is to empirically test a set of seven industry-level indicators that aim to capture, and distinguish, two parts of shock effects in networked economies. Specifically, three of these indicators reflect the part of a shock effect due to the "connectivity" aspects of the network—which industries are linked, directly and indirectly, to which others—by measuring the number of paths by which each manufacturing industry is connected to petroleum refineries, oil and gas extraction, and automobile manufacturing. These indicators vary over the study period based on three snapshots of the U.S. production network at the 6-digit-NAICS level: one in 1967, one in 1987, and one in 1997.

The other four indicators capture the part of a shock effect due to the "evolutionary" aspects of the network: patterns in industries' input substitution in response to price variation over multiple layers in the network. For the current chapter, these indicators specifically reflect how petroleum products are used as a substitute for or a complement to other inputs in each industry's production process. These indicators vary annually based on changes to the network at the 3-digit-NAICS level.[18]

Altogether, these seven indicators differ from previous research both conceptually and in terms of their construction. This is especially true of the evolutionary indicators, which are based on categorizing individual input pairs in industries' production processes as either substitutes or complements, rather than estimating a single substitution parameter (i.e., elasticity) across multiple inputs as has often been done in previous work.

I evaluate the relevance and predictive power of the indicators by examining their statistical significance for industries' year-to-year output and value added within a regression framework. I find that the results reflect substantial industry differentiation during many of the U.S. oil price episodes

[18] More precisely, I estimate the substitutability/complementarity relationships among inputs based on the 3-digit-NAICS data from 1966-2016. I take these classifications to be static throughout the study period, but the indicators constructed from them vary for each year.

we know about from the existing literature. At the same time, the indicators also provide information about these episodes by illuminating the mechanisms through which shocks were transmitted in the economy—upstream versus downstream, and connectivity versus evolutionary—and how such dynamics mediated the impacts that industries felt.

For instance, consider one of these indicators: the number of network paths that connect each industry, on its supply side, with the petroleum refineries industry. The findings in Chapter 1 suggest that the greater the number of these paths, the greater the impact industries will experience—as compared to other industries—during a particular shock. This indicator becomes negative and statistically significant (at the 95 percent level) several times during the study period, including during two of the biggest and most well known oil price episodes: those in 1973-1974 and 1979-1980. Looking across the study period, the results suggest that—when comparing industries at the 25th and 75th percentiles of this indicator—those in the latter group had outcomes about 0.6-4.7 percent lower than those in the former group during such episodes.

Given that this indicator captures the number of paths to petroleum refineries on industries' upstream sides, these results provide evidence that industries were significantly impacted by supply-side shock propagation during these times. Moreover, two of the evolutionary indicators—which capture substitution possibilities in the upstream direction of industries—not only confirm this finding at a number of points but suggest that such propagation is a direct result of industry-level differences in the ability to shift away from higher-priced petroleum products. This provides an explanation for why the supply-side channel may have been important during certain episodes even though the cost share of oil in production is relatively small.

As a second example, consider an indicator that captures how an industry's customers (and those customers' customers, and so on) tend to use more of an industry's output when using less petroleum products. Moving from the 25th to the 75th percentile in this customer-demand measure corresponds with approximately 0.8-1.2 percent higher outcomes in 1975-1976 and 1981-1982, suggesting a (positive) counteracting effect on the demand side for some industries in the aftermath of the 1973-1974 and 1979-1980 episodes. In other words, as industries substituted away from petroleum products, the industries to which they substituted tended to do better than others (all else equal) during these post-periods. Interestingly, the reverse pattern emerges during the episodes themselves, suggesting that industries were actually substituting towards—and not away—from petroleum products as prices began to rise, which provides evidence that these price increases were substantially driven by demand-side factors.

Overall, the patterns in the results suggest three key takeaways: (1) that oil price episodes are associated with significant heterogeneity in outcomes across industries; (2) that an industry's place in the production network and its relationships to its suppliers and customers, both immediate and removed, are relevant for its economic outcomes; and (3) that oil price episodes are a complex synthesis of multiple factors, in the respect that they appear to have had both supply- and demand-side—as well as connectivity and evolutionary—causes and consequences.

Finally, I note that the approach in this chapter, as well as its particular findings, may be able to inform both the retrospective evaluation of past policies and/or the development of future policies related to energy and environment, including those that aim to address climate change. Specifically, as described in Chapter 1, the indicators represent a loose correspondence between shocks and a set of heterogeneous industry impacts: on one hand, they represent a way to differentiate industries in the presence of various types of shocks, where such shocks might be known or anticipated; on the other hand, they provide a method of inferring the presence of different types of shock transmission given the observed (and varying) economic performance of industries.

I use the indicators for the latter purpose in this chapter by focusing on oil episodes and relevant shocked industries, though the approach could also be applied to other energy or environmental contexts with their own set of shocked industries. In addition, the estimates produced in this chapter, or in similar research, could likewise be used for the former purpose, such as in predicting which industries would be most affected, and how, by the enactment of an upstream carbon tax.

Related Literature. This study is primarily related to previous work in the areas of oil price shocks and economic production networks, while also having connections to research on general energy price increases, pass-through of input prices, and economic impacts of regulation.

Oil Price Shocks. The early literature on oil price shocks, which emerged in the 1980s, suggested at least two stylized facts regarding the connection between oil prices and the U.S. economy: (1) that historical increases in the price of crude oil have often preceded reductions in U.S. output; and (2) that price increases have frequently coincided with significant events arising outside of the U.S. economy (Hamilton, 1983; Hamilton, 1985). These early papers—in addition to stimulating a large literature on oil price shocks—were influential in suggesting the view that oil price episodes were (at times) a consequence of foreign events and simultaneously a cause of negative outcomes for the domestic economy.

Baumeister and Kilian (2016) discuss how the oil shocks literature has evolved since the 1980s and provide a thorough review of oil price episodes from the 1970s on. They note several important ways in which the more recent literature has modified the view of oil price episodes. One of these is the finding that exogenous geopolitical events (relating to the supply of crude oil) are only one cause of oil price fluctuations, and may not even be the most important. For instance, they argue that most major oil price changes since 1973 are primarily explained by demand shifts, and most importantly, by changes in oil demand driven by the global business cycle. They also note that shifts in demand for oil inventories—which can be triggered by changes in perception, geopolitical events, and/or precautionary demand—have likewise been important (Baumeister and Kilian, 2016).[19]

Along these lines, we might divide the mechanisms by which oil price shocks affect the economy into two broad categories: supply-side and demand-side (Kilian, 2008a). The supply-side viewpoint considers oil as an input to domestic production and, in turn, sees oil price fluctuations as influencing production decisions. In contrast, the demand-side viewpoint examines how oil price fluctuations motivate changes in the spending of consumers and firms.

Kilian (2008a) notes that there are well-known issues with using the supply-side channel to explain macroeconomic impacts of oil price changes. One problem is that the cost share of oil in domestic output is small, and so "...standard production-based models of the transmission of energy price shocks are not capable of explaining large fluctuations in real output" (Kilian 2008a, p. 880). The author refers to several studies that have attempted to deal with this issue by introducing one or more structures or mechanisms, including substantial markups that vary over time, complementarities between energy and capital in production, and energy as an essential prerequisite for the use of capital. He notes that although these models emphasize the importance of the supply channel in transmitting energy price shocks, it is not yet clear which, if any, have empirical support.

Another strand of the literature has instead focused on the demand side. Kilian (2008a) draws on this literature to describe several channels through which energy price changes can have economic effects, all of which act through a reduction in aggregate demand. These include a discretionary income effect, whereby discretionary income is reduced when energy prices rise (especially if energy demand is relatively inelastic), as well as an operating cost effect, whereby demand decreases for energy-intensive durables (such as automobiles) when energy prices rise. The author also points out the potential importance of indirect mechanisms, which may—through the reallocation of capital and labor among industries in the economy—lead even small changes in energy prices to have substantial effects on output and employment.

Researchers have studied industry-specific outcomes as one way to explore these supply- and demand-side dynamics. For instance, Lee and Ni (2002) use a vector autoregression model to identify

[19] The authors note: "It is important to keep in mind that oil inventory demand depends on the shortfall of expected supply compared with expected demand rather than just one side of the market. Historical evidence suggests that, in practice, shifts in inventory demand tend to arise only when geopolitical turmoil coincides with expectations of strong demand for crude oil and tight oil supplies" (Baumeister and Kilian, 2016; p. 156).

how oil price shocks affected the output and prices of 14 industries. If an industry's output and price move in the same direction, the authors call it a demand-side shock; if output and price move in opposite directions, they call it a supply-side shock. They find that the two oil-intensive industries in their study—petroleum refining and industrial chemicals—are affected through the supply-side channel, while most of the other industries (and especially the automobile industry) are affected on the demand side. They conclude that although oil price shocks appear to affect industries' supply as well as their demand, such shocks primarily reduce the supply of oil-intensive industries while mostly reducing the demand of many other industries. The authors also note that these patterns reflect industry experiences as detailed in the trade press.

Another industry-focused study was completed by Davis and Haltiwanger (2001), who investigate the impacts of oil price shocks on job creation and destruction. Treating oil price shocks as exogenous, they estimate a vector autoregression model using detailed U.S. manufacturing employment data over the period 1972-1988. They find heterogeneous shock responses across industries, and when regressing those responses against industry-specific characteristics, identify capital intensity, product durability, and energy cost share as important determinants of the magnitude of the employment response. The authors also find an asymmetric effect of oil price shocks, with much bigger impacts following increases as compared to decreases.

More generally, there is evidence that the magnitude of price impacts has decreased over time, perhaps as a result of changes to the structure of the U.S. economy (Kilian, 2008a). One potential set of factors relates to the U.S. automobile industry, which has become (1) relatively more competitive in its production of energy-efficient vehicles (as compared with foreign automakers) but also (2) relatively less important to the overall economy, in terms of both consumer expenditures and employment. Another possible factor is the nature of shocks themselves, which may have transformed from being primarily supply-driven to mostly demand-driven.

Overall, several themes emerge from this literature that are relevant for the current chapter:

- Oil price shocks have likely had heterogeneous impacts across industries;

- Some industries are more likely to experience, or be conduits for, supply-side effects (such as the petroleum refining industry) while others are more likely to face changes in demand (such as the automobile industry), and the differences may be partially explained by the manner in which industries use their oil and other energy inputs;

- Oil prices are not just a cause of economic outcomes (for instance, when price increases are precipitated by geopolitical turmoil) but also an effect of economic trends (such as when growth leads to strong demand); and

- Indirect effects may be just as important as direct effects when assessing and explaining impacts to industries.

Previous studies have often taken a macro view of the effects of oil price shocks, and when research has considered sub-national effects—such as at the industry level—these studies have not (to my knowledge) explicitly incorporated how industries are related to one another. In this way, the current chapter builds on the existing literature by explicitly considering the potential relevance of inter-industry linkages and the connectivity/evolutionary dynamics of shocks over time.

Economic Production Networks. This study is also related to a literature that considers production network models. The papers I discuss below fall into two broad categories: those that are primarily theoretical and those that take theory to data.

Theoretical papers include early work by Long and Plosser (1983) as well as more recent studies by Gabaix (2011), Acemoglu et al. (2012), Acemoglu, Ozdaglar, and Tahbaz-Salehi (2017), Baqaee (2018), and Baqaee and Farhi (2018), among others. A common theme across these papers is an examination of how industry or firm interconnections can lead micro-level shocks to influence aggregate outcomes. One exception is Baqaee and Farhi (2018), who instead consider the impacts of micro-level shocks on other micro entities (i.e., on firms and industries), while also explicitly allowing for heterogeneity in production functions and consumption preferences. I describe in more detail in Chapter 1 the similarities and differences between the current approach and this work by Baqaee and Farhi.

Empirical papers in the production networks literature have considered volatility at the aggregate level as well as impacts at the industry and firm levels. Some of the findings of this research are:

- At the economy-wide scale, that the estimated importance of sectoral shocks in aggregate volatility is significantly affected by the substitution possibilities of industries, with greater input complementarity implying a larger role for industry-specific shocks (Foerster, Sarte, and Watson, 2011; Atalay, 2017);

- At the industry level, that supply shocks are more powerfully propagated downstream while demand shocks tend to be propagated upstream (Acemoglu, Akcigit, and Kerr, 2016); and

- At the firm level, that input specificity is a key factor in the propagation of idiosyncratic shocks through the economy (Barrot and Sauvagnat, 2016), that the effect of a shock tends to decrease with network distance (Carvalho et al., 2021), and that the correlation between firm growth and foreign-country GDP growth is significantly correlated with international trade linkages, both for directly-connected firms as well as for those that purchase from them (Di Giovanni, Levchenko, and Mejean, 2018).

The empirical analysis in this chapter is most similar to Acemoglu, Akcigit, and Kerr (2016), as I use an industry-level dataset and consider upstream and downstream shock propagation simultaneously. The current study differs in its specific focus on oil price episodes, as well as its inclusion of indicators that intend to capture the substitutability and complementarity of inputs. With regards to this latter aspect, the current chapter is similar to Atalay (2017) and Barrot and Sauvagnat (2016) in its consideration of substitution possibilities, though it differs (1) from the first in that it allows such patterns to vary at the industry and input levels and (2) from the second in that it uses an input-categorization approach to estimate the pairwise substitutability and complementarity of industry inputs while also considering these dynamics over multiple levels in the production network.

Other Related Literatures. This study is also broadly related to several other literatures, which I briefly mention here.

Heterogeneity of energy price impacts: Research has found that energy prices can differentially affect industries based on industry characteristics, with energy intensity being an important factor. In particular, energy intensity appears to be a partial determinant of the magnitude of production and/or employment impacts experienced by industries in response to regulation or energy price increases, with greater energy-intensity leading to greater impacts (Curtis, 2018; Aldy and Pizer, 2015; Fowlie, Reguant, and Ryan, 2016). There is also evidence that industry energy-intensity affects industry geographic location and employment (Kahn and Mansur, 2013).

Pass-through of input prices: One key mechanism by which energy prices may propagate through the network is pass-through of input price increases from one industry to another. Research has shown that pass-through can be significant, though the magnitude depends on numerous factors, including market power, demand elasticity, and the form of the shocks, whether idiosyncratic (local) or shared (Stolper, 2016; Muehlegger and Sweeney, 2017; Ganapati, Shapiro, and Walker, 2018).

Economic impacts of regulation: Finally, there is a broad literature that considers the economic impacts of energy and environmental regulation. Although the current study is focused specifically on oil price episodes, a better understanding of the U.S. economic production network could potentially be relevant for regulatory assessment, especially if the production network were considered at a regional, rather than a national, level. As one example, the NOx Budget Trading Program was a cap-and-trade system in effect from 2003 to 2008 that aimed at NOx emissions reductions from sources (certain electricity generators and industrial plants) in a subset of U.S. states. Related research includes studies by Curtis (2018) and Deschênes, Greenstone, and Shapiro (2017), who analyzed employment, pollution, health outcomes, and defensive expenditures associated with the program. Given that the program exhibited temporal, spatial, and industry heterogeneity, a network approach in this context could provide evidence as to how and why region-specific or industry-specific effects may have had a broader economic impact.

Overall, the current research builds on these studies by leveraging a network view of the economy to investigate how supply-side, demand-side, connectivity, and evolutionary factors impact industries in an oil episodes setting. In particular, the economic production network can be viewed as an additional lens through which to view industry heterogeneity, where an industry's dependence on energy inputs (such as on the utilities and/or petroleum products industries) are a subset of its inputs in the larger production network, and where the connections an industry has to its suppliers and customers are just one level of its broader connectedness to other industries.

The remainder of this chapter proceeds as follows: in section 2.2, I briefly discuss several oil price episodes as background; in section 2.3, I review the model developed in Chapter 1; in section 2.4, I describe the empirical analysis in detail, including the construction of the indicators; in section 2.5, I present and discuss the main results, as well as some robustness results that consider an information technology context rather than an oil episodes context; and in section 2.6, I end with some concluding thoughts.

2.2 Historical Oil Price Episodes

Before turning to the model, I briefly review several specific oil price episodes that provide background and context for the empirical analysis.

The 1973-1974 oil price crisis has traditionally been explained as a supply shock triggered by the 1973 Arab-Israeli War (Baumeister and Kilian, 2016). Barsky and Kilian (2002) highlight several other factors that may have been at play, including economic considerations related to the Tehran/Tripoli agreements, U.S. dollar deflation, rising U.S. inflation, and strong worldwide demand for oil. In this latter interpretation, the oil price increase of this period was endogenous to the global macroeconomy (Baumeister and Kilian, 2016). In support of this view, Kilian (2008b) found that perhaps a quarter of the 1973-1974 price episode was due to an exogenous OPEC supply shock.

Similarly, although the traditional view of the 1979-1980 oil price crisis is that it was due to a reduction in Iranian oil production following the Iranian Revolution, the timing of the geopolitical events leaves a temporal gap before the largest price increase (Baumeister and Kilian, 2016). Kilian and Murphy (2014) argue that a change in expectations, rather than interference with production, was the cause. Specifically, they find that about one-third of the price increase was due to inventory demand, presumably as a reaction to the geopolitical situation.

From September 1980 to January 1981, the WTI price increased from $36 to $38 due to Iraq invading Iran and the destruction of Iranian oil infrastructure. Baumeister and Kilian (2016) note that this incident presents an example of a supply shock without a demand shock.

The oil price reached a peak in April 1980 and then generally fell through the early to mid-1980s. There were many reasons for this, including an increase in non-OPEC production (Baumeister and Kilian, 2016). Saudi Arabia attempted to stabilize the price of oil by reducing its own production, but by the end of 1985, its financial losses were significant and it was forced to reverse course. This led to a sharp drop in the oil price in 1986. Baumeister and Kilian (2016) point out that this decline was also caused, importantly, by lowered inventory demand for oil, as people learned that OPEC was unable to prop up the oil price.

OPEC reached a new accord in late November 1988 that reduced their overall production ceiling by about 18 percent. There was a sense by some analysts at the time that the agreement would hold in a way that it had not in recent years, given the end of the Iran-Iraq war and a ceasefire that had been observed since August.[20] In combination with strong oil demand from industrialized countries, the oil price—which had been stagnant or declining since April—changed trajectory, increasing through the first part of the next calendar year.

In August 1990, an oil price spike followed the invasion of Kuwait and the disruption in Iraqi and Kuwaiti oil production. In addition to the conflict itself, equally important was increased inventory demand in anticipation of potential disruption to Saudi oil production. Once enough troops had moved into Saudi Arabia by late 1990, fears subsided and inventory demand—and prices—fell sharply. Baumeister and Kilian (2016) note that the quick decline in prices in 1991 lends strength to the expectations/inventory explanation, as actual production in Iraq and Kuwait was slower to recover than were prices.

There were two exogenous oil supply disruptions in late 2002/early 2003: one due to civil unrest in Venezuela and the other a result of the 2003 Iraq War. However, there was a relatively small shift in inventory demand (due to modest fears regarding the impact on Saudi oil production), and so the price of oil only briefly spiked by about $6 (Baumeister and Kilian, 2016).

[20]For instance, see: Phillips, John. November 29, 1988. OPEC accord could boost oil prices. United Press International Archives. Available at: https://www.upi.com/Archives/1988/11/29/OPEC-accord-could-boost-oil-prices/7426596782800/.

A U.S.-specific price incident occurred in late 2005 due to the effects of Hurricanes Rita and Katrina. As Kilian (2008a) describes:

> "Whereas the reduction of U.S. crude oil supply associated with these exogenous events was negligible on a global scale, the reduction in refining capacity was not. It constituted both a decline in the demand for crude oil (associated with a fall of crude oil prices) and a decline in the supply of gasoline and other refined products, reflected in a sharp increase in gasoline prices" (p. 876).

The WTI price climbed from $28 to $134 between mid-2003 and mid-2008. There is agreement that this substantial increase was caused by strong demand rather than by supply disruptions (Baumeister and Kilian, 2016).

Finally, Kilian and Lee (2014) estimate that the Libyan crisis in 2011 led to a price increase of $3-$13, depending on the specification. They find that tensions with Iran the following year led to a $0-$9 increase.

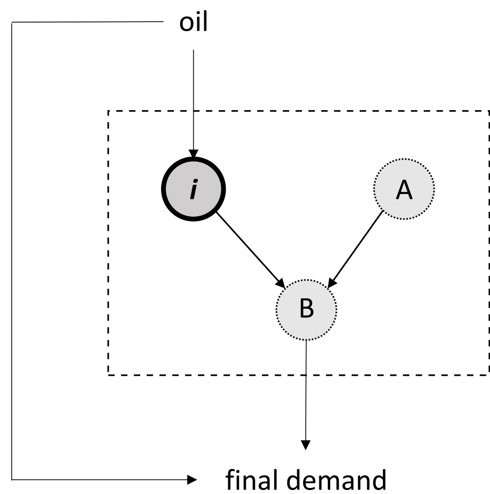
2.3 Model

The results in Chapter 1 provide evidence that the effects experienced by industries in the presence of shocks have two parts: (1) the portion due to the connectivity aspect of the network, such as which industries are directly and indirectly connected to which others; and (2) the portion due to the evolutionary aspect of the network, which is a function of how industries change their usage of inputs in response to varying input prices. In the next section, I draw on these findings to propose seven indicators relevant for oil price episodes in the U.S. economy: three to capture connectivity aspects of the network and four to reflect evolutionary aspects.

Before turning to those indicators, I present here a short example based on the model from Chapter 1. Consider a simple three-industry economy, where industry B uses industry i and industry A to make its product, and industries i and A do not depend on any industries within the economy (Figure 2.1). In this model, as in Chapter 1, each industry produces its output in the least-cost manner to satisfy the sum of intermediate and final demand. Suppose that:

- Industry i uses some amount of oil, such that its output price is a function of the oil price;
- Industries i and A produce intermediate goods, while industry B produces a final good (such that there is only final demand for industry B's good); and
- Final demand for industry B's good is a function of the oil price in addition to good B's price, given that consumers have a limited budget and purchase some amount of oil in addition to the good.

Figure 2.1: Illustration of a Three-Industry Economy.



Consider the impact of an exogenous, supply-side shock to the oil price, and as an illustrative case, consider specifically the impact on industry i. The results of the model in Chapter 1 suggest that the change in industry i's profits/revenues is determined by a combination of up to four factors: the direct effect of the supply shock; a pass-through effect as industry i increases its output price to accommodate an increase in input costs; an input substitution effect caused by industry B's substitution possibilities; and a final demand effect in response to industry B potentially changing its own output price.

In the language of Chapter 1, there is both an upstream connectivity factor (how i is connected to the shocked industry) and a downstream connectivity factor (how i is connected to the change in final demand). There is also a downstream evolutionary factor: how demand for i's output is affected by industry B's substitution possibilities. Although not explicitly part of this example, there is also an upstream evolutionary factor: in the same way that industry B may increase its output price depending on its ability to substitute away from industry i's higher-priced output, such dynamics would also be relevant for i as it attempts to substitute away from oil.

In Chapter 1, I propose and then explore indicators that fall into each of these four categories (upstream connectivity, downstream connectivity, upstream evolutionary, and downstream evolutionary). In the next section, I adapt these indicators to the oil episodes context and then examine them empirically.

2.4 Empirics

Applying the results of Chapter 1 to the oil episodes context, I propose the following seven indicators for each industry i:

1. The number of directed paths from the petroleum refineries industry to industry i;

2. The number of directed paths from industry i to the oil and gas extraction industry;

3. The number of directed paths from industry i to the automobile manufacturing industry;

4. A measure that captures the relative substitutability of petroleum products and other inputs in i's production process, as well as the substitutability of petroleum products and other inputs in the production processes of i's suppliers, and those suppliers' suppliers, and so on;

5. A measure that captures the relative complementarity of petroleum products and other inputs in i's production process, as well as the complementarity of petroleum products and other inputs in the production processes of i's suppliers, and those suppliers' suppliers, and so on;

6. A measure that captures the relative substitutability of petroleum products and industry i's output in the production processes of i's customers, as well as the substitutability of petroleum products and the outputs of i's customers in the production processes of those customers' customers, and so on; and

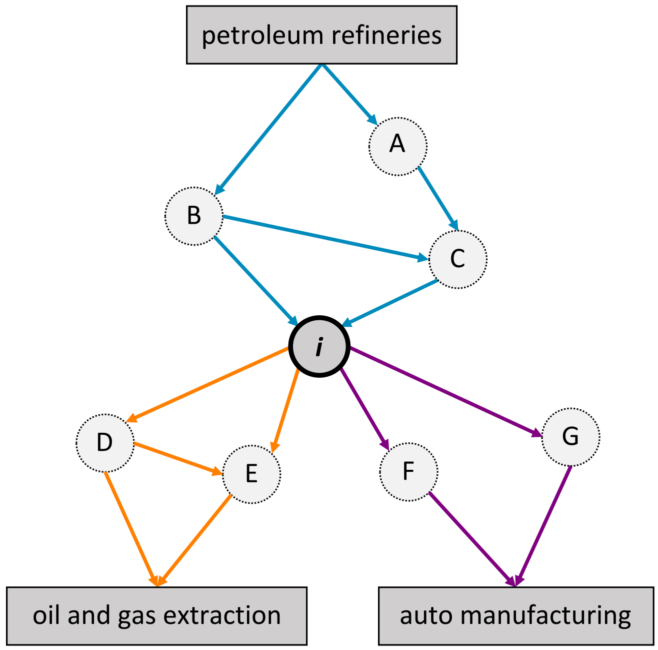
7. A measure that captures the relative complementarity of petroleum products and industry i's output in the production processes of i's customers, as well as the complementarity of petroleum products and the outputs of i's customers in the production processes of those customers' customers, and so on.

I calculate the first three of these indicators for industries at the 6-digit-NAICS level. Due to data availability, the last four are estimated at the 3-digit-NAICS level. For this reason, I follow the 6-digit-NAICS and 3-digit-NAICS definitions and refer to the petroleum industry as "petroleum refineries" in the 6-digit context and as "petroleum products" in the 3-digit context.[21] I describe the conceptual framework for the indicators here and then discuss their empirical construction in the following subsections.

Given the focus of this chapter on oil price episodes, it is natural to consider the relationship of industries to petroleum refineries (indicator 1 above) and oil and gas extraction (indicator 2 above).

[21] I note that the 3-digit-NAICS industry 324 includes coal products in addition to petroleum products. For simplicity, I refer to this industry in the 3-digit context as "petroleum products."

Figure 2.2: Illustration of Connectivity Indicators.



Note: the three connectivity indicators are: (1) supply paths to petroleum refineries (in blue); (2) demand paths to oil and gas extraction (in orange); and (3) demand paths to auto manufacturing (in purple). Arrows point from suppliers to their customers.

Specifically, I include the petroleum refineries industry on the upstream side to account for shocks to crude oil production/prices that we would expect to directly affect petroleum refineries and, in turn, to directly or indirectly affect each successive layer of industries moving downstream in the production network.

Although the oil and gas extraction industry is upstream of the petroleum refineries industry in the broader network (see footnote 22 below), I include an indicator for oil and gas extraction on the downstream side of each industry i for two reasons. The first and primary reason is to account for the increasingly important role of energy production in the U.S. economy, and in particular, to capture the fact that this capital-intensive industry creates demand for various manufacturing industries upstream of it. The second reason is that I assume any effects that propagate from oil and gas extraction to downstream manufacturing industries are first channeled through the petroleum refineries industry. In this way, including an upstream oil and gas extraction indicator would provide information similar to that already provided by the petroleum refineries indicator.

I include automobile manufacturing (indicator 3 above) as a demand industry of interest for two reasons. One is that previous research has indicated that gasoline prices may have an impact on automakers through consumer purchasing decisions (see the literature discussion above and, for example, Langer and Miller, 2013). Another is that, in the context of economy-scale production networks, the automobile manufacturing industry is particularly downstream. For example, Antràs et al. (2012) discuss two measures of industry upstreamness in the production network, showing that they are theoretically equivalent and discussing some of their economic interpretations. They calculate the measures using the benchmark 2002 input-output table for the United States and find that automobile manufacturing is the most downstream industry.[22]

Overall, I intend for these three indicators to account for the most salient upstream and downstream connectivity relationships in an oil episodes context (Figure 2.2).

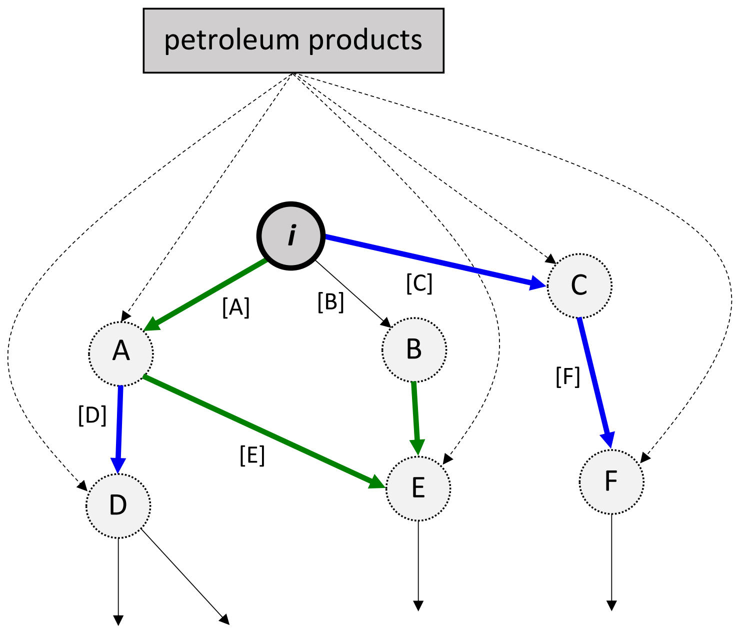
The evolutionary (shock response) relationships considered in Chapter 1 are captured by the remaining four indicators. Rather than focusing on the structural relationships among industries per se—that is, which industries are connected to one another, and how—these indicators intend to reflect how short-run adjustments throughout the network can reinforce one another to impact industries. Two of these indicators (numbered 4 and 5 above) account for such dynamics on the upstream side of an industry i, while the other two (numbered 6 and 7) account for these effects on the downstream side.

As an illustration, consider the two downstream evolutionary indicators for a hypothetical industry industry i (Figure 2.3). In this example, industry i has three customers: A, B, and C. Industry i provides some fraction of its output (in dollars) to A, which is designated in the diagram by $[A]$. Industry A itself has two customers: D and E. Industry A provides some fraction of its output (again in dollars) to D and E, which is designated in the diagram by $[D]$ and $[E]$, respectively.

The colors of the edges indicate how an industry uses each of its inputs in relation to its usage of petroleum products. Specifically, the green edge marked $[A]$ signifies that industry A generally uses less (more) of industry i's output when it uses more (less) petroleum products. In this sense, I label industry i's output a substitute for petroleum products in industry A's production process (see Chapter 1 for the details of this interpretation). A similar relationship is reflected by the edge marked $[E]$; industry E tends to use less (more) of industry A's output when it uses more (less) petroleum products. The opposite relationship is signified by the blue edge directed from industry i to industry C. In this case, industry C generally uses more (less) of industry i's output when it uses

[22] These scores range from a minimum of 1.0003 for the automobile manufacturing industry (most downstream) to a maximum of 4.6511 for the petrochemical manufacturing industry (most upstream). The score for the petroleum refineries industry is 2.3961, and the score for the oil and gas extraction industry is 3.3468. Note that the authors also calculate the measures for 16 European countries and the United States as presented in the OECD STAN database. They find that the rank correlations among the countries are high and statistically significant, indicating that the measures are stable across countries.

Figure 2.3: Illustration of Downstream Evolutionary Indicators.



Note: the two downstream evolutionary indicators are: (1) chains of substitutability with petroleum products (in green); and (2) chains of complementarity with petroleum products (in blue). Arrows point from suppliers to their customers.

more (less) petroleum products. In this way, I label industry i's output a complement to petroleum products in industry C's production process.

Suppose there is an exogenous shock to the oil price, which causes petroleum products to become more expensive. As described in Chapter 1, the notion is that—as the price rises—industry A will tend to use more of industry i's output. At the same time, industry E will also tend to use more of industry A's output. These effects combine to result in increased demand for industry i's output, and we would expect such effects to be greater the larger the values $[A]$ and $[E]$. Following this logic, I create a downstream indicator for industry i by taking the product of the values along each continuously green path (e.g., $[A]$, $[E]$, etc.) leading away from i and then summing across those paths (for illustration, only one such path is pictured). I call these chains of input substitutability "substitute paths." Similarly, taking the product of the values along each continuously blue path (e.g., $[C]$, $[F]$, etc.) leading away from i, and then summing across those paths, yields another downstream indicator for industry i. We expect these "complement paths" to have an opposite (dampening) effect on demand for industry i's output in the presence of oil price increases.

I create two upstream indicators for each industry i in an analogous fashion, except that the values used are the fraction of an industry's total inputs it receives from each supplier, rather than the fraction of its total output that it provides to each customer. The intuition for these two upstream indicators is as follows: continuously green paths on the upstream side of industry i represent a line of suppliers that use more of their other inputs when using less petroleum products, suggesting that these suppliers may be able to substitute away from petroleum products in the event of a price shock. All else equal, such substitution would allow these suppliers to moderate an increase in their input costs, resulting in a lesser degree of price pass-through. As such, we would expect a greater number and/or magnitude of these paths to, on average, reduce the effect of an upstream price shock on industry i. Conversely, continuously blue paths on the upstream side of industry i represent a line of suppliers that use less of their other inputs when using less petroleum products, suggesting that these suppliers may be relatively restricted in their ability to substitute away from petroleum products in the event of a price shock. Therefore, we might expect, on average, such paths to amplify (or at least transmit) the effects of an upstream price shock for industry i.

Although I have discussed these downstream and upstream indicators in terms of exogenous price increases, industries would also experience effects—that is, opposite effects—in the presence of exogenous price decreases. At the same time, and as discussed in the literature review above, it is also likely that during some (or possibly many) oil episodes, price fluctuations were due not only to exogenous (i.e., supply-related) factors but also to changing demand. This is particularly relevant for the downstream evolutionary indicators, which—in addition to reflecting how intermediate input demand, vis-à-vis petroleum products, changes in response to fluctuating oil prices—may also signal how the oil price is responding to changing demand. In this sense, the downstream evolutionary indicators can provide a proxy both for the role of prices on demand and for the role of demand on prices. I discuss these dynamics in greater detail in section 2.5 in the context of the main results.

Finally, given that the construction of the downstream evolutionary indicators uses edge values that sum to one across each industry's customers (e.g., $[A] + [B] + [C] = 1$ for industry i in the figure), each of the downstream indicators falls in the range [0,1], and the pair of downstream indicators for each industry sums to less than (or equal to) one. Similarly, given that the construction of the upstream evolutionary indicators uses edge values that sum to one across each industry's suppliers (not pictured), each of the upstream indicators also falls in the range [0,1], and the pair of upstream indicators for each industry sums to less than (or equal to) one. From this view, the pair of downstream indicators and the pair of upstream indicators each represent the relative percentage importance of substitute and complement relationships (vis-à-vis petroleum products) for each industry on the downstream and upstream sides, respectively.

2.4.1 Data Sources and Variables Construction

I focus on U.S. manufacturing industries' value of shipments and value added as the main outcomes of interest. The National Bureau of Economic Research (NBER) and U.S. Census Bureau's Center for Economic Studies (CES) publish a dataset that contains annual-level data on U.S. manufacturing at the 6-digit-NAICS level over the period 1958-2018 (hereafter referred to as the NBER-CES dataset), which I use to construct the dependent variables for the analysis (see below).

As mentioned above, the indicators for the regressions are constructed from two sets of data:

- For the connectivity indicators—which can be calculated from a single snapshot of the production network—I use the "benchmark" direct requirements input-output tables published by the U.S. Bureau of Economic Analysis (BEA). I use the benchmark tables for 1967, 1987, and 1997.

- For the evolutionary indicators—which require an estimate of industries' input relationships—I use the U.S. BEA's annual "use" input-output tables for the years 1966-2016, in which industries are generally aggregated at the 3-digit-NAICS level. I also use the U.S. BEA's chain-type price and quantity indexes for gross output by industry presented at the 3-digit-NAICS level.

Finally, although I estimate the production network structure and industry input substitutability and complementarity from input-output tables for industries beyond manufacturing, I only assess the outcomes for those manufacturing industries that appear in all relevant datasets (that is, in the NBER-CES data and in the datasets used to generate the indicators).

Dependent Variables

I calculate the dependent variable, y_{it}, as a percentage change in value of shipments or value added as compared to the average over the previous two years. Specifically:

$$y_{it} = (v_{it} - ((v_{i,t-1} + v_{i,t-2})/2))/((v_{i,t-1} + v_{i,t-2})/2)$$

where v_{it} is the value of shipments or the value added for 6-digit-NAICS industry i in year t.

Connectivity Indicators

Let I be the set of 6-digit-NAICS industries. Let $i \in I$ denote an arbitrary industry, and let $refining$, $extract$, and $auto$ in I denote the petroleum refineries, oil/gas extraction, and automobile manufacturing industries, respectively. Let A_t be the 6-digit-NAICS direct requirements table for the year t, where an element a_{jit} of A_t is the dollar amount of industry j's output required to produce one dollar's worth of industry i's output in the year t.

Given that industry classification codes have changed over time (and switched from the SIC system to the NAICS system in 1997), I make use of two industry concordances published alongside the NBER-CES dataset: one to convert from 1972 SIC to 1987 SIC, and the other to convert from 1987 SIC to 1997 NAICS.[23] I also create my own concordance to convert from 1967 SIC to 1972 SIC. With these, I am able to match industries in 1967 with those in 1987 and 1997.

I then calculate the following three indicators for $t \in \{1967, 1987, 1997\}$ for each industry i:

(1) The number of directed paths from the petroleum refineries industry to industry i:

$$\text{supply\_paths\_refining}_{it} = \text{num\_paths}(refining, i|A_t);$$

(2) The number of directed paths from industry i to the oil and gas extraction industry:

$$\text{demand\_paths\_extraction}_{it} = \text{num\_paths}(i, extract|A_t);\ \text{and}$$

(3) The number of directed paths from industry i to the automobile manufacturing industry:

$$\text{demand\_paths\_auto}_{it} = \text{num\_paths}(i, auto|A_t)$$

where $\text{num\_paths}(x, y|A_t)$ is the number of paths from x to y calculated using depth-first search over the directed network represented by A_t.[24] For computational reasons, and as an analog to the corresponding indicators in Chapter 1, only paths of length three or less are included in the total.

Evolutionary Indicators

As described above, the evolutionary indicators intend to capture chains of substitutability and complementarity on industries' upstream and downstream sides. In order to construct these indicators, I first leverage the input-categorization approach described in Chapter 1 to estimate the substitutability or complementarity between petroleum products and every other input within each 3-digit-NAICS industry's production process.

As described in Chapter 1, the output of this approach is a binary categorization of each pair of inputs in a production process as either substitutes or complements. These estimates are not formal estimates of production functions but are instead estimates of the relationships among changing

[23] Please see: https://www.nber.org/research/data/nber-ces-manufacturing-industry-database.
[24] I use the Python library NetworkX to determine the number of paths.

input and output quantities as determined by a machine learning algorithm. Although the approach does not impose any restrictions on the form of these relationships, per the empirical exercise in Chapter 1, the categorization for a pair of inputs can be thought of as estimating whether a particular production step—the one that uses the inputs either directly or indirectly—is of a form that takes its inputs as substitutes (e.g., a Cobb-Douglas production node) or is of a form that takes its inputs as complements (e.g., a Leontief production node). Please see Chapter 1 for additional details regarding this interpretation.

There are three main steps in constructing the evolutionary indicators:

1. Assemble a dataset for each industry that records the annual percentage changes in all of its input quantities alongside the annual percentage change in its output quantity over the years 1966-2016;

2. Use the input-categorization approach to infer, for each industry, which inputs are substitutes for petroleum products and which are complements to petroleum products in the industry's production process; and

3. Using these input relationships, generate the two upstream evolutionary indicators and the two downstream evolutionary indicators.

As I describe in more detail below, the substitute/complement categorizations produced in the second step are held constant throughout the study period, while the values of the indicators created in the third step are allowed to vary based on the annual values in the 3-digit-NAICS input-output tables.

Construction of Quantity Changes Dataset. Let B_t be the 3-digit-NAICS use input-output table in year t, where b_{jit} in B_t is the dollar amount of industry j's output used to produce industry i's total output in year t. In addition, let p_{it} be industry i's price index in year t, and let q_{it} be industry i's quantity index in year t.

For each year $t \in [1966, 2016]$ and for each pair of industries i and j, I:

- Calculate the ratio of industry i's use of industry j's output in year t compared to its use in year $t-1$: $c_{jit} = (b_{jit}/b_{j,i,t-1})$;

- Adjust this ratio by dividing by the equivalent ratio for industry j's price index, and then subtract one: $d_{jit} = c_{jit}/(p_{jt}/p_{j,t-1}) - 1$;

- Calculate the change in industry i's output quantity index between the years $t-1$ and t: $e_{it} = (q_{it}/q_{i,t-1}) - 1$; and then

- Adjust industry i's input percentage changes by its output percentage change: $f_{jit} = d_{jit} - e_{it}$.

At the end of this process, I arrange the f_{jit} values to create a dataset for each industry i, where the observations represent the years 1966-2016 and the variables are the estimated quantity changes in industry i's use of each other industry j's output (adjusted by the percentage change in i's own output[25]). I also add industry i's output change (e_{it} above) as a variable to the dataset.

Estimating Pairwise Substitutability and Complementarity of Inputs. I use the input-categorization approach proposed in Chapter 1 to estimate the pairwise substitutability or complementarity of petroleum products and every other input in each industry's production process.

For each industry i, this first involves training the eXtreme Gradient Boosting (XGBoost) algorithm to predict industry i's year-to-year change in petroleum products quantity used ($f_{oil,i,t}$) using the simultaneous changes in the other quantities (f_{jit} with $j \neq i$ and $j \neq oil$) as well as the change in industry i's output quantity index (e_{it}). In doing so, I aim to estimate a functional relationship loosely related to the marginal rate of technical substitution, but for all inputs simultaneously and where the algorithm may use the industry's changing output quantity as a predictor. I apply the algorithm to each industry's dataset separately under the assumption that each industry has a unique production process. The training for each dataset is performed via cross-validation, and I store the root-mean-squared-error (RMSE) across the folds for use in the next step.

I then use the trained model for each industry i to infer an input pattern between petroleum products and each other input. Specifically, the notion is to see how i's predicted change-in-use of petroleum products varies when each other input j goes from being used much less (change of -50 percent) to being used much more (change of +50 percent). The result is, similar to the first step, loosely akin to the marginal rate of technical substitution, but now for each pairwise combination of petroleum products and input j in isolation (within industry i's production process). I fit a line through these predicted points, and if the absolute value of its slope is large enough relative to the RMSE,[26] then I classify petroleum products and input j either as substitutes in i's production process (if the slope is negative) or as complements in i's production process (if the slope is positive). Please see Chapter 1 for additional details about the approach and discussion of its performance.

Note that, because the XGBoost algorithm performs feature (regressor) selection in addition to estimation of a functional relationship, there are cases where a change in industry i's use of industry

[25] Please see Chapter 1 for the reasoning behind this adjustment.

[26] In Chapter 1, I used a threshold of RMSE/4 to determine if a slope would be used for categorization or if the associated input pair would be left as uncategorized. Here I decrease this threshold to RMSE/12 for two reasons. The first is to minimize the number of zeros in the evolutionary indicators, which aids in their interpretation in the regressions. The second reason is that the results in Chapter 1 suggest that decreasing this threshold increases the number of input pairs that are categorized while only minimally decreasing the approach's accuracy.

j's output leads to zero predicted change in the use of petroleum products. In these cases—along with those cases where the slope is not large enough relative to the RMSE—I do not consider industry j's output to be either a complement or a substitute to petroleum products in industry i's production process.

Overall, the result of the approach is a categorization, for each industry, of whether the industry uses petroleum products as a substitute for or a complement to (or neither) each other input in its production process. These categorizations are taken to be constant throughout the study period, under the simplifying assumption that—even if industries' production processes change over time—an input that is used as a substitute for petroleum products at one time will not become a complement at another time, and vice-versa.

Construction of Indicators. With this information, I have an estimate for each industry i of the subset of its suppliers' outputs that it uses as complements to petroleum products and the subset of its suppliers' outputs that it uses as substitutes for petroleum products. Viewed another way, this also provides estimates for each industry about how its various customers use its own output as a complement to or substitute for petroleum products in their respective production processes.

Analogous to Chapter 1, and as described above, I calculate four indicators from a combination of these categorizations and the values in the original input-output tables. One difference from Chapter 1 is that here, instead of using the Leontief inverse, I calculate and then sum the weighted paths of length three or less between each industry i and every other industry j.[27] I do this primarily to focus on the substitutability and complementarity relationships in the immediate vicinity of each industry i rather than incorporating relationships of all lengths (as does the Leontief inverse). Using paths of length three or less also provides a parallel with the connectivity indicators described above.

To implement this approach, I first take the original 3-digit-NAICS production network and keep only the edges from suppliers where the suppliers have been identified as substitutes for petroleum products (the green edges previously illustrated). Call this the "substitutes network." Similarly, I again take the original production network and keep only the edges from suppliers where the suppliers have been identified as complements to petroleum products (the blue edges previously illustrated). Call this the "complements network."

Then, for each industry i and year $t \in [1966, 2016]$, I construct the two downstream evolutionary indicators by repeating the following twice, once in the substitutes network and once in the com-

[27] I note that these approaches are conceptually similar with two main differences. The first is that the Leontief inverse captures "walks" rather than "paths" in the network, where the former allows repeated visits to the same node while the latter does not. The second difference is that the Leontief inverse incorporates walks of all lengths, while here I consider paths of lengths three or less only.

plements network: (1) for each industry $j \neq i$ and $j \neq oil$, find all directed paths of length three or less from i to j; (2) for each path, take the product of the edge weights along that path, where the weight of the edge from supplier x to customer y represents the fraction of x's output purchased by y in year t; and (3) sum the products across all paths. Call the sum for the substitutes network $\text{down\_sub\_paths}_{it}$ and the analogous sum for the complements network $\text{down\_comp\_paths}_{it}$.

For each industry i and year $t \in [1966, 2016]$, I construct the two upstream evolutionary indicators by repeating the following twice (similar to above but slightly modified), once in the substitutes network and once in the complements network: (1) for each industry $j \neq i$ and $j \neq oil$, find all directed paths of length three or less from j to i; (2) for each path, take the product of the edge weights along that path, where the weight of the edge from supplier x to customer y represents the fraction of y's inputs it purchases from x in year t; and (3) sum the products across all paths. Call the sum for the substitutes network up_sub_paths_{it} and the analogous sum for the complements network $\text{up\_comp\_paths}_{it}$.

In this way, although the structures of the substitutes and complements networks remain static over the study period, the indicators produced from them vary over the years 1966 to 2016 based on the changing values in the underlying 3-digit-NAICS input-output tables.

Finally, given that these indicators are constructed at the 3-digit-NAICS level, I use the subscript $h(i)$ when industry i is instead (as below) at the 6-digit-NAICS level. The function $h(\cdot)$ maps from 6-digit-NAICS industries to their 3-digit-NAICS counterparts.

2.4.2 Summary Statistics

I calculate summary statistics for each of the indicators (Table 2.1). When viewed as a histogram, the distribution of supply_paths_refining is approximately normal. The distributions of demand_paths_extraction and demand_paths_auto are quite skewed in the positive direction. The evolutionary indicators also tend to be skewed in the positive direction. For these reasons and for interpretability of the results, I take the logarithm of all indicators in the regressions. I do not take the logarithm of the dependent variables because they contain negative as well as positive values.

I also calculate the correlation between each pair of indicators (Table 2.2). Generally, these correlations are small and suggest that each of the indicators is capturing distinct information about industries' places and relationships in the production network. One notable exception is the high, positive correlation between demand_paths_extraction and demand_paths_auto, which makes sense given that the supply chains of oil and gas extraction and automobile manufacturing may have substantial overlap.

Table 2.1: Summary Statistics for Indicators.

Variable	Mean	Std. Dev.	Min.	Max.
supply_paths_refining	28.619	15.552	0.000	96.718
demand_paths_extraction	0.397	0.967	0.000	6.000
demand_paths_auto	2.394	5.726	0.000	37.000
down_comp_paths	0.128	0.139	0.000	0.767
down_sub_paths	0.125	0.121	0.000	0.753
up_comp_paths	0.210	0.164	0.023	0.822
up_sub_paths	0.054	0.072	0.000	0.471

Table 2.2: Pairwise Correlations of Indicators.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(1) supply_paths_refining	1.00						
(2) demand_paths_extraction	0.03	1.00					
(3) demand_paths_auto	0.02	0.73	1.00				
(4) down_comp_paths	0.05	0.18	0.23	1.00			
(5) down_sub_paths	0.01	-0.10	-0.13	-0.21	1.00		
(6) up_comp_paths	-0.03	-0.02	0.11	-0.12	0.02	1.00	
(7) up_sub_paths	0.08	0.11	-0.02	0.08	-0.03	-0.28	1.00

2.4.3 Regression Approach

In the example in the model section above, I consider a single supply-side shock, which also has ramifications for a single final-demand good. In the U.S. economy, there are likely many supply-side and demand-side shocks occurring simultaneously, though we would not expect all shocks to be equally important in every time period. We would, however, expect the seven indicators defined in this section to be particularly relevant in the years where oil supply-side, extraction and auto demand-side, and petroleum products intermediate input factors are relatively strong compared to other changing factors. In a limiting situation where these are the only changing factors, they would theoretically explain all of the variation across industries' year-to-year economic outcomes.

To account for the occurrence of non-oil-related shocks, I include controls in the specification that allow for economy-wide shocks in each year and shocks to 2-digit-NAICS industries in each year. I also include changes in employment and general energy usage[28] at the 6-digit-NAICS level, which I calculate based on information provided within the original NBER-CES dataset. I do not include other economic controls, such as changes in capital and material inputs, as these are taken to be determined by or highly correlated with the indicators that are included (in the years in which the indicators are relevant).

[28]Per the NBER-CES documentation, this includes expenditures on seven types of energy: electricity, residual fuel oil, light oils, liquefied petroleum, coal, coke, and natural gas.

To account for the portion of an oil-related shock that affects industries directly (i.e., without propagation through the production network), I include a control for direct dependence on the petroleum products industry at the 3-digit-NAICS level. Analogous to the construction of the substitute and complement indicators, I measure this dependence as the fraction of an industry's inputs it purchases from the petroleum products industry in each year.[29]

To implement this approach, I pool the data across all years and generate year fixed-effects, 2-digit-NAICS-by-year fixed effects, and interactions between year dummies and each of the seven indicators. The regression specification is then as follows:

$$
\begin{aligned}
y_{it} = \alpha + Y_t + 2\text{digit}_i \cdot Y_t + \beta_{emp} \cdot \text{emp}_{it} + \beta_{energy} \cdot \text{energy}_{it} + \\
&\beta_{direct} \cdot \log(\text{direct}_{h(i),t-2}) + \beta_{direct,t} \cdot \log(\text{direct}_{h(i),t-2}) \cdot Y_t + \\
&\beta_{refining} \cdot \log(\text{supply\_paths\_refining}_{i,g(t)}) + \beta_{refining,t} \cdot \log(\text{supply\_paths\_refining}_{i,g(t)}) \cdot Y_t + \\
&\beta_{extract} \cdot \log(\text{demand\_paths\_extract}_{i,g(t)}) + \beta_{extract,t} \cdot \log(\text{demand\_paths\_extract}_{i,g(t)}) \cdot Y_t + \\
&\beta_{auto} \cdot \log(\text{demand\_paths\_auto}_{i,g(t)}) + \beta_{auto,t} \cdot \log(\text{demand\_paths\_auto}_{i,g(t)}) \cdot Y_t + \\
&\beta_{dcomp} \cdot \log(\text{down\_comp\_paths}_{h(i),t-2}) + \beta_{dcomp,t} \cdot \log(\text{down\_comp\_paths}_{h(i),t-2}) \cdot Y_t + \\
&\beta_{dsub} \cdot \log(\text{down\_sub\_paths}_{h(i),t-2}) + \beta_{dsub,t} \cdot \log(\text{down\_sub\_paths}_{h(i),t-2}) \cdot Y_t + \\
&\beta_{ucomp} \cdot \log(\text{up\_comp\_paths}_{h(i),t-2}) + \beta_{ucomp,t} \cdot \log(\text{up\_comp\_paths}_{h(i),t-2}) \cdot Y_t + \\
&\beta_{usub} \cdot \log(\text{up\_sub\_paths}_{h(i),t-2}) + \beta_{usub,t} \cdot \log(\text{up\_sub\_paths}_{h(i),t-2}) \cdot Y_t + \epsilon_{it}
\end{aligned}
$$

where y_{it} is either the percentage change in the value of shipments or the percentage change in value added, α is a constant, Y_t is a set of year fixed-effects, 2digit_i is a set of 2-digit-NAICS fixed effects, emp_{it} is the industry-level change in employment,[30] energy_{it} is the industry-level change in energy usage,[31] direct_{it} is the fraction of industry i's inputs purchased from petroleum products in year t, $g(t)$ maps year t to one of {1967, 1987, 1997} (whichever most closely precedes t), $h(i)$ maps each 6-digit-NAICS industry i to its 3-digit-NAICS counterpart, and the other regressors are calculated as in section 2.4.1.

I note that I include the evolutionary indicators with a two-year lag to ensure that the indicator values are based on data (i.e., on entries in the 3-digit-NAICS input-output tables) that were

[29] I include this control uninteracted with the year as well as interacted with the year, which yields a separate estimate for each year in the sample. As with the indicators, we would expect direct dependence on the petroleum products industry to be associated with heterogeneous impacts across different years in the study period.

[30] To mirror the dependent variables, this is the change in employment versus the prior-two-year average.

[31] To mirror the dependent variables, this is the change in energy expenditures versus the prior-two-year average.

established prior to the year in which the indicators are used to differentiate industry outcomes (though—as described above—the categorizations of input pairs as substitutes or complements are based on quantity-change observations over the entire period 1966-2016). For similar reasons as well as for consistency, I also include the control for direct dependence on petroleum products with a two-year lag.

Finally, as described above, the supply path and demand path variables are constructed for each industry at the 6-digit-NAICS level. The substitute and complement indicators instead vary at the 3-digit-NAICS level. In turn, $\beta_{refining}$, $\beta_{extract}$, and β_{auto} (and their interacted versions) are identified by variation across all 6-digit-NAICS industries in the sample. In contrast, the upstream and downstream substitutes/complements coefficients are identified by variation in the average change in outcomes for industries at the 3-digit-NAICS level.

2.5 Results and Discussion

The results for the indicators are shown below (Figures 2.4 through 2.8). In each figure, I plot against the left axis the change in value of shipments / change in value added for each year in the study period based on (1) the linear combination of the non-interacted and interacted coefficients for each indicator when (2) moving from the 25th percentile to the 75th percentile in the values for that indicator.

For instance, in Figure 2.4, I calculate $\beta_{refining}+\beta_{refining,t}$ for $t \in [1968, 2018]$ and then multiply that value by:

$$\log(\text{pct}(\text{supply\_paths\_refining}, 75)/\text{pct}(\text{supply\_paths\_refining}, 25))$$

where $\text{pct}(\cdot,\cdot)$ represents the percentile operator. Given that the 25th and 75th percentiles for demand_paths_extract are both zero and that the 25th percentile for demand_paths_auto is zero, I instead use the following multipliers for those indicators:

$$\log(\text{pct}(\text{demand\_paths\_extract}, 90)/\min_{>0}(\text{demand\_paths\_extract})) \qquad \text{and}$$

$$\log(\text{pct}(\text{demand\_paths\_auto}, 75)/\min_{>0}(\text{demand\_paths\_auto}))$$

where $\text{pct}(\cdot,\cdot)$ is again the percentile operator and where $\min_{>0}(\cdot)$ finds the smallest value greater than zero.

Analogous to Chapter 1, I also present the differences between the substitutes and complements coefficients on industries' upstream (Figure 2.9) and downstream (Figure 2.10) sides. These figures provide another lens through which to view the effects of changing intermediate input usage patterns on industry outcomes (see below).

Lastly, for clarity, I only show the linear combinations in each year that are statistically significant at the 95 percent level (standard errors are clustered at the industry level). For comparison purposes, I include in the appendix a set of figures that display linear combinations in each year that are statistically significant at the 80 percent level, which help to convey a broader sense of trends across the study period.

I plot against the right axis in all figures the percentage change in the WTI spot price (where the change is calculated against the average of the prior two years). Note that the spot price is shown only for graphical comparison purposes and was not used in the analysis in any way.

The coefficients on the non-interacted terms are generally statistically insignificant (Table 2.3). Two exceptions are demand paths to auto manufacturing for value of shipments and upstream complement paths for value added.

Table 2.3: Coefficients and Standard Errors for Non-Interacted Terms.

	Value of shipments	Value added
# of supply paths to refineries	-0.000546	0.00204
	(0.00129)	(0.00218)
# of demand paths to extraction	0.000232	0.0000914
	(0.000768)	(0.00113)
# of demand paths to auto manuf	-0.00146**	-0.000408
	(0.000685)	(0.000950)
downstream complement paths	-0.000251	0.00119
	(0.00166)	(0.00229)
downstream substitute paths	0.00103	0.000765
	(0.00173)	(0.00268)
upstream complement paths	-0.000481	0.00897*
	(0.00414)	(0.00523)
upstream substitute paths	0.000422	-0.00218
	(0.00234)	(0.00317)
N	16,677	16,677
adj. R^2	0.695	0.541

Standard errors clustered by industry.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

I first discuss the results for the connectivity indicators and then turn to the results for the evolutionary indicators.

2.5.1 Supply and Demand Paths

As described in section 2.4, the paths indicators capture the number of ways that each manufacturing industry is connected to petroleum refineries on its upstream side and to oil/gas extraction and auto manufacturing on its downstream side. In turn, and given the regression specification above, a negative (positive) value for one of these indicators in a particular year reflects that the more connected a manufacturing industry is to the associated upstream or downstream industry (as quantified by the number of paths), the smaller (larger) its year-to-year change in value of shipments or value added compared to other industries in that year (controlling for both year-specific factors and for 2-digit-NAICS-by-year-specific factors). In this way, these indicators differentiate manufacturing industries based on their structural relationships in the network vis-à-vis these three other oil-related industries.

My interpretation of these patterns through an oil episodes lens is based both on the model in Chapter 1 and on previous research, which suggest that price changes tend to be propagated from upstream to downstream in the network while demand changes tend to be propagated from downstream to upstream in the network. In turn, I take: (1) supply paths to refineries—by virtue of measuring the connections upstream of industries—as becoming relevant when there are changes to the oil price as channeled from upstream; and (2) demand paths to oil/gas extraction and auto manufacturing—through measuring the connections downstream of industries—as becoming relevant when there are changes to demand for extractive activities and automobiles, respectively, as channeled from downstream. In other words, the supply paths indicator is a proxy for shock transmission via the supply-side channel, while the demand paths indicators serve as proxies for shock transmission via the demand-side channel.[32]

I note that although the supply paths indicator reflects the supply-side transmission of an oil price change, it does not necessarily provide information about the underlying cause of the price change itself. In other words, regardless of whether an oil price change is driven by supply-side

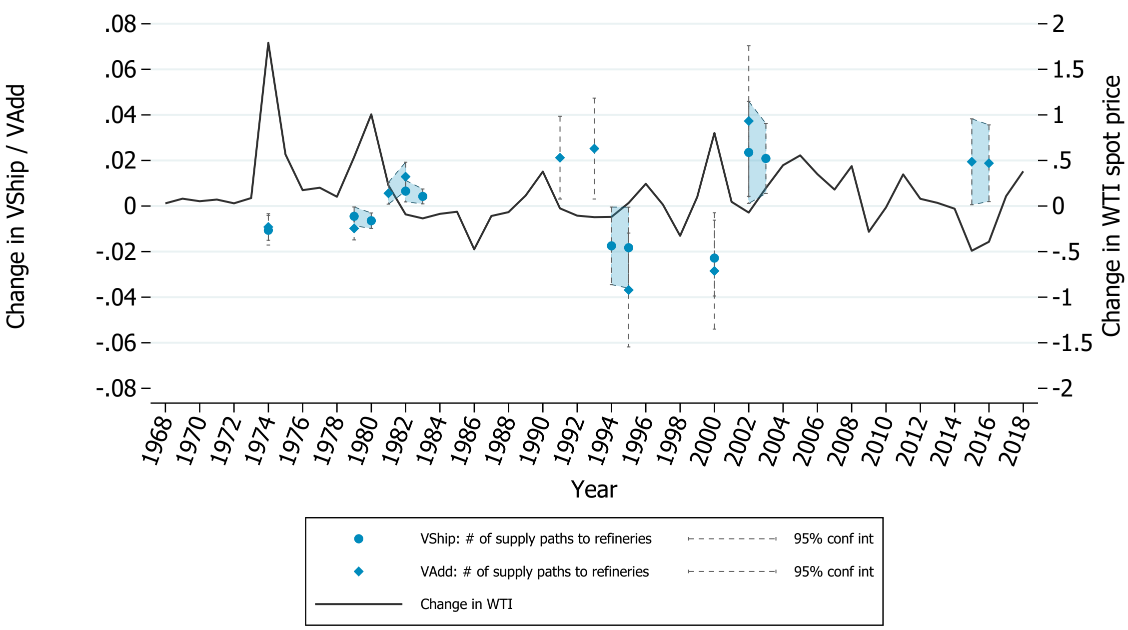
[32] There is a secondary interpretation of the supply paths indicator, which stems from its close relationship with the output of petroleum refineries. Specifically, as demand for manufacturing industries with strong dependence (direct or indirect) on petroleum refineries increases, this demand necessarily translates into increased demand for the output of petroleum refineries itself, which—all else equal—puts upward pressure on the oil price. From this view, given that the supply paths indicator measures the (relative) economic outcomes of such high-dependence industries, it can serve not only as a proxy for the effect of oil price changes on such industries through the supply-side channel (as discussed), but also as a proxy for the effect of these industries on oil prices via demand. I do not focus on this latter interpretation as it is less relevant in the context of the results.

factors (e.g., shifts in oil production), demand-side factors (e.g., macroeconomic trends), or both, it may still manifest as impacting industries through the supply-side channel. As I describe in the next subsection, the downstream evolutionary indicators provide some insight into which of these underlying causes is most relevant in each oil episode.

Supply Paths to Petroleum Refineries

We see that supply paths to petroleum refineries is negative and statistically significant at the 95 percent level during two of the most well-known price episodes: the oil shocks of 1973-1974 and 1979-1980 (Figure 2.4). Per the discussion above, this suggests that the oil price increases during these episodes affected industries by channeling the shock from upstream to downstream. The results for the upstream evolutionary indicators (next subsection) provide some evidence during the first episode (for the years 1973-1975) and leading into the second (in 1978-1979) that the pass-through effect was due, in part, to the inability of some industries to substitute away from higher-priced petroleum products. At the same time, the downstream evolutionary indicators suggest that the price fluctuations themselves were at least partially driven by changes in demand (see below).

Figure 2.4: Indicator Results: Supply Paths to Petroleum Refineries.



Note: the figure shows the change in value of shipments / value added based on linear combinations of $\beta_{refining} + \beta_{refining,t}$ for $t \in [1968, 2018]$ moving from the 25th percentile to the 75th percentile. Only estimates significant at the 95 percent level are pictured. Successive points are connected by confidence areas.

Supply paths to refineries is also negative at two other times: 1994-1995 and 2000. Each of these corresponds to a period in which the oil price had been declining but then reversed: in 1994-1995, after a period where prices were generally decreasing following the spike in 1990; and in 2000, as demand—and the oil price—recovered after the Asian financial crisis of 1997. As above, that the supply paths indicator is negative suggests that both of these reversals led to effects for industries via the supply-side channel. Although the upstream evolutionary indicators do not become statistically significant during these times, the downstream evolutionary indicators provide some evidence for a shifting away from petroleum products during 2000-2001, which suggests that the corresponding price increase may have been at least partially supply-driven.

The supply paths indicator becomes positive and statistically significant at the 95 percent level five times during the study period: 1981-1983, 1991, 1993, 2002-2003, and 2015-2016. These results are again best explained by propagation via the supply-side channel, but now in reaction to decreasing (rather than increasing) oil prices.

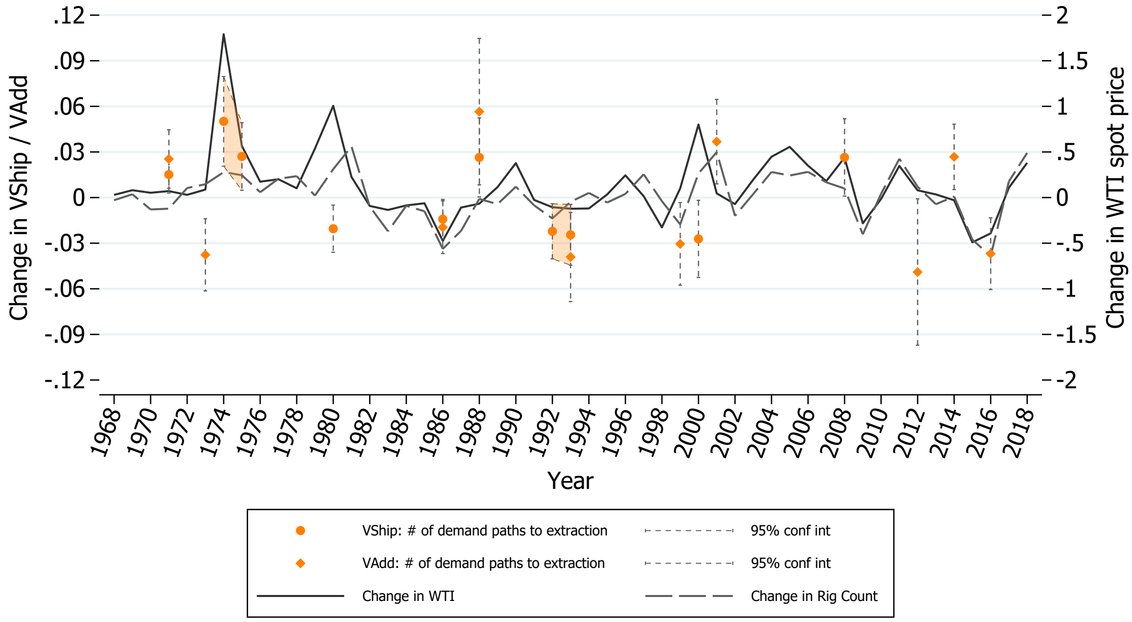
For 1981-1983, 1991, and (to a lesser extent) 2002-2003, this interpretation is further supported by the upstream evolutionary indicators, which also reflect the pass-through of decreasing prices during these years. The upstream evolutionary indicators instead reflect the pass-through of increasing—rather than decreasing—prices in 1993 and 2016. This latter finding suggests that the positive values for the supply paths indicator in 1993/2015-2016 may have involved somewhat different mechanisms than those at work in the other periods, though more research would be needed to understand these dynamics fully.

This said, the downstream evolutionary indicators suggest a shifting-away from petroleum products just before or during nearly all of these episodes, which is consistent with a demand-driven reduction in the oil price and with the fact that the United States experienced recessions in 1980-1982, 1990-1991, and 2001. Altogether, we can interpret the positive values for the supply paths indicator during these five times as follows: the oil price fell in large part due to reduced demand; this decreased price was passed through the production network from upstream to downstream, resulting in a reduction in input costs for industries; and this reduction was greater (i.e., more beneficial) the larger the number of paths an industry had to petroleum refineries on its upstream side.

Demand Paths to Oil and Gas Extraction

Demand paths to oil and gas extraction becomes negative and statistically significant at the 95 percent level at seven points during the study period: 1973, 1980, 1986, 1992-1993, 1999-2000, 2012, and 2016 (Figure 2.5). Conversely, this indicator becomes positive and statistically significant at six times: 1971, 1974-1975, 1988, 2001, 2008, and 2014.

Figure 2.5: Indicator Results: Demand Paths to Oil and Gas Extraction.



Note: the figure shows the change in value of shipments / value added based on linear combinations of $\beta_{extract} + \beta_{extract,t}$ for $t \in [1968, 2018]$ moving from the smallest value greater than zero to the 90th percentile. Only estimates significant at the 95 percent level are pictured. Successive points are connected by confidence areas.

These patterns are generally explained by variation in the U.S. oil and natural gas rig count, which has tended to follow oil prices with a lag (see the dashed line in Figure 2.5). Specifically, decreases in the rig count (or stagnating growth) are frequently associated with lower year-over-year outcomes for manufacturing industries that have greater exposure to oil and natural gas extraction on their downstream sides. Conversely, higher year-over-year outcomes for such industries tend to occur in years where the rig count is increasing.

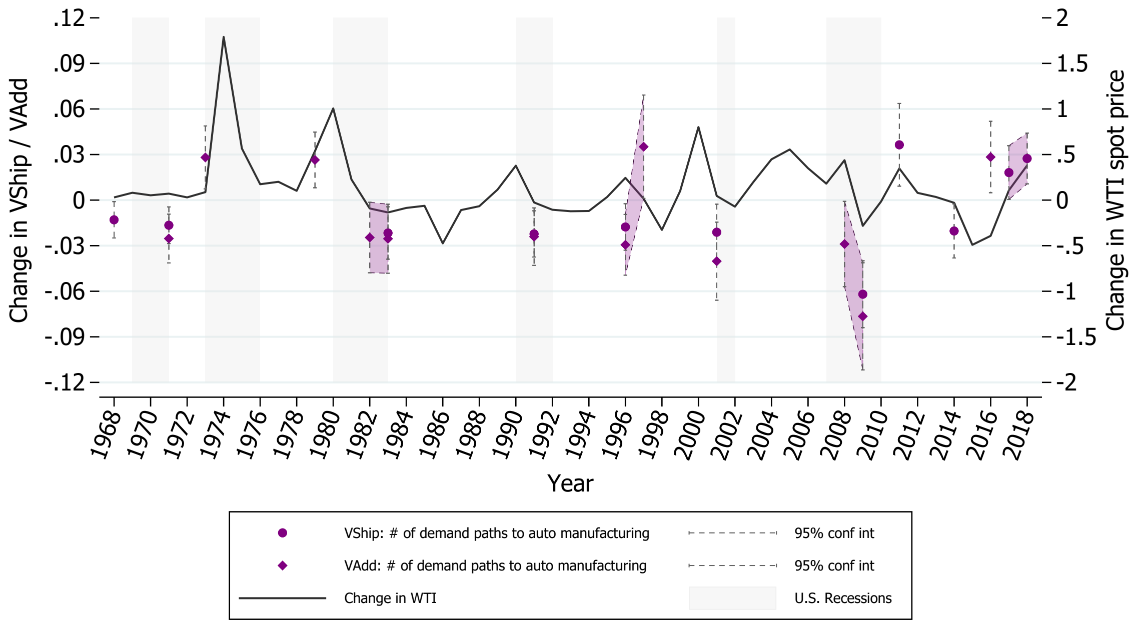
This general pattern—of positive (beneficial) differentiation during expansion of the rig count and negative (detrimental) differentiation during decreases or stagnating growth—is further illustrated by the results at the 80 percent significance level as shown in the appendix. Altogether, these findings confirm the important role of extractive activity for some manufacturing industries on the demand side, and suggest that such patterns are related to oil episodes in the respect that the oil price is a primary driver of domestic extraction.

Demand Paths to Auto Manufacturing

Demand paths to auto manufacturing becomes negative and statistically significant at the 95 percent level in several years during or after economic recessions, including 1971, 1982-1983, 1991, 2001, and

2008-2009 (Figure 2.6; U.S. recessions are highlighted in grey). This suggests that manufacturing industries with a greater amount of downstream exposure to auto manufacturing did worse, year-over-year, than other industries during these times. This also implies, in turn, that there was a relatively greater reduction in consumer spending on automobiles as compared to the outputs of other manufacturing industries during these periods.

Figure 2.6: Indicator Results: Demand Paths to Auto Manufacturing.



Note: the figure shows the change in value of shipments / value added based on linear combinations of $\beta_{auto} + \beta_{auto,t}$ for $t \in [1968, 2018]$ moving from the smallest value greater than zero to the 75th percentile. Only estimates significant at the 95 percent level are pictured. Successive points are connected by confidence areas.

As described in the introduction, some previous research has suggested that oil price episodes have affected the U.S. economy by reducing demand for automobiles. The results provide evidence for this theory to the extent that oil price increases have often preceded recessions, though the reductions in automobile demand generally appear to occur with a lag vis-à-vis the price episodes. For this reason, the results suggest that the proximate causes of these demand reductions were the corresponding recessions, even though the recessions themselves may have been triggered (in part) by the oil price.

Lastly, of the five periods the indicator becomes positive and statistically significant at the 95 percent level, two occur in the 2010s, one occurs in 1997, and the last two occur just as the 1970s oil episodes are beginning: once in 1973 and once in 1979. As will be discussed in the next subsection, this provides additional evidence that the price increases associated with the 1970s episodes were at least partially driven by demand.

2.5.2 Chains of Substitutability and Complementarity

I present the upstream and downstream substitutability/complementarity indicators on their own (Figures 2.7 and 2.8) and as differences (Figures 2.9 and 2.10).

As described in section 2.4, the upstream evolutionary indicators measure how each industry—and its suppliers, both direct and indirect—use petroleum products as a substitute for or a complement to their other inputs. In doing so, these indicators reflect the relative ability (as captured by the substitute paths indicator) or inability (as captured by the complement paths indicator) of industries and their suppliers to shift their intermediate input usage away from petroleum products.

In turn, in years with pass-through of an oil price increase, we would expect industries with a greater number/magnitude of upstream substitute (complement) paths to experience relatively higher (lower) year-over-year outcomes as compared to other industries. Relatedly, we would expect the difference between the upstream substitute and complement paths coefficients to be positive when such price increases are being passed through the network. In the case that a decrease in the oil price is being passed through the network, we would expect the reverse of these patterns. See Table 2.4 for a summary of these relationships.

Table 2.4: Interpretation of Upstream Evolutionary Indicators.

	In year t, pass-through of an:	
	oil price increase	oil price decrease
substitute paths	+ (or relatively larger)	– (or relatively smaller)
complement paths	– (or relatively smaller)	+ (or relatively larger)
difference	+	–

I note that, similar to the case of the supply paths indicator, these upstream evolutionary indicators do not necessarily reflect whether the underlying cause of an oil price change is supply-driven, demand-driven, or both, as all of these may be passed through the supply-side channel.

The downstream evolutionary indicators measure how an industry's customers—and those customers' customers, and so on—use petroleum products as a substitute for or a complement to each industry's output. In doing so, these indicators capture the extent to which the changing intermediate input use of an industry's customers, both direct and indirect, increases (as captured by the substitute paths indicator) or decreases (as captured by the complement paths indicator) demand for that industry's output.

In turn, in years where intermediate input usage is shifting away from petroleum products, we would expect industries with a greater number/magnitude of downstream substitute (complement) paths to be associated with relatively higher (lower) year-to-year outcomes as compared to other industries. Similarly, we would expect the difference between the downstream substitute and complement paths coefficients to be positive when such shifts are occurring. In the case that intermediate input usage is instead shifting towards petroleum products, we would expect the reverse of these patterns. See Table 2.5 for a summary of these relationships.

Table 2.5: Interpretation of Downstream Evolutionary Indicators.

	In year t, intermediate input use shifting:	
	away from petroleum products	towards petroleum products
substitute paths	+ (or relatively larger)	− (or relatively smaller)
complement paths	− (or relatively smaller)	+ (or relatively larger)
difference	+	−

As described briefly in section 2.4, by virtue of their close relationship with demand for petroleum products, the downstream evolutionary indicators can be interpreted in two ways. The first reflects the effect of oil prices on industries' intermediate input demand, while the second reflects the effect of input demand on the oil price. Which of these two mechanisms I take to be dominant during an episode depends on whether the indicators suggest a shifting away or a shifting towards petroleum products when prices are increasing or decreasing.

Specifically, in an approach similar to that of Lee and Ni (2002) as described in the introduction, I take these indicators to suggest: (1) the effect of prices on input demand (i.e., a supply-side driver) when increasing (decreasing) prices correspond to a shifting away from (towards) petroleum products; or (2) the effect of input demand on prices (i.e., a demand-side driver) when increasing (decreasing) prices correspond to a shifting towards (away from) petroleum products.

Altogether, while I always interpret the upstream evolutionary indicators as reflecting the transmission of prices via the supply-side channel, my interpretation of the downstream evolutionary indicators depends on whether the indicators suggest a shifting away or a shifting towards petroleum products at the same time that prices are increasing or decreasing.

Turning to the results, we see that the upstream evolutionary indicators suggest the supply-side pass-through of increasing oil prices most clearly for the 1973-1974 and 1979-1980 oil episodes—as well as two points in the 2010s—as complement paths becomes negative and statistically significant at the 95 percent level just before or during these periods (Figure 2.7). This finding is echoed by the difference between the upstream substitutability and complementarity indicators, which is positive and statistically significant in 1975, 1978, 2013-2014, and 2016-2017 (Figure 2.9). Conversely, and as mentioned above, the results suggest the supply-side pass-through of decreasing oil prices most clearly during the periods 1980-1983 and 1990-1991 (Figures 2.7 and 2.9).

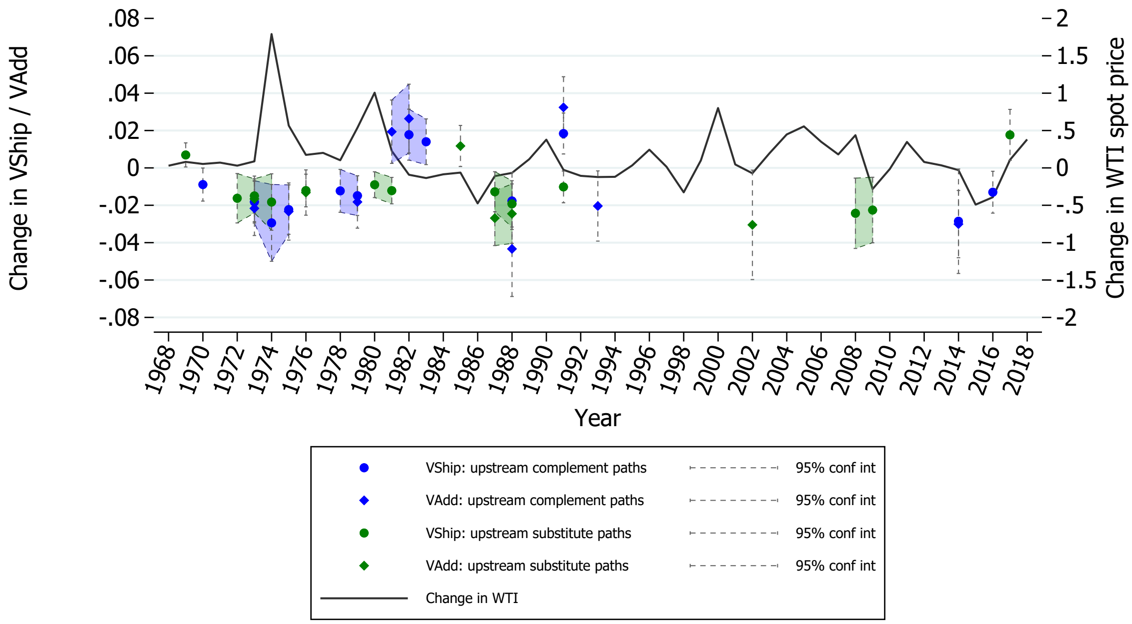
Combined with the results for supply paths to refineries, these findings provide strong support for the conclusion that industries were affected via the supply-side channel in response to (1) the price increases associated with the 1973-1974 and 1979-1980 episodes and (2) the price decreases in the aftermath of the 1979-1980 and 1990 episodes. The results for the upstream evolutionary indicators further suggest that this pass-through was at least partially the effect of heterogeneity in industries' ability to shift away from or towards petroleum products. However, as discussed above, this does not necessarily imply that the price fluctuations were themselves due to supply-side factors (such as changes in foreign oil production). This said, that the connectivity and evolutionary indicators both point to the supply side during these episodes provides the strongest evidence during the study period that such factors may have been at work at these times.

Relatedly, I note that the upstream evolutionary indicators tend to become statistically significant more often during the first half of the study period (Figures 2.7 and 2.9). This pattern is also reflected by the results at the 80 percent significance level as shown in the appendix. If these indicators do in any way reflect a supply-side cause of price fluctuations in addition to a supply-side channel, this suggests—as has some previous research—that oil price episodes may have become relatively less supply-driven over the past several decades.

In contrast, the results for the downstream evolutionary indicators suggest a prominent role for demand over the entire study period. Specifically, these results show that a number of oil price increases (or slowing decreases) may have been at least partially demand-driven, including those in 1973-1974, 1979, 1983-1984, 1987-1989, 2004-2006, 2010-2011, and 2017-2018. This is evidenced by periods (1) when the downstream substitute paths indicator becomes negative and statistically significant (Figure 2.8) and/or (2) when the difference between the downstream substitute and complement paths coefficients becomes negative and statistically significant (Figure 2.10). As described above, such patterns suggest a greater use of petroleum products as an intermediate input.

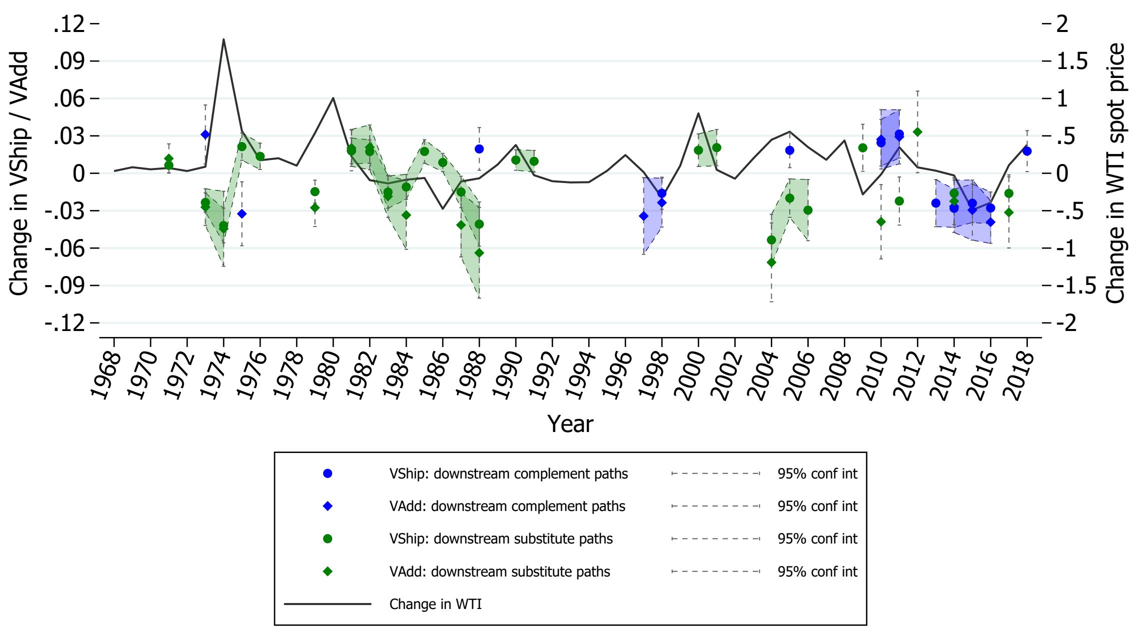
These findings corroborate some of the more recent work in the oil episodes literature that has attributed oil price increases to demand factors in addition to (or in place of) supply factors. As described in the introduction, this marks a change from the early literature—especially with regards to the 1970s episodes—that has often viewed oil price changes as exogenous and supply-driven. In contrast, the results here suggest a key role for demand-driven price increases throughout the 1970s, the 1980s, the 2000s, and the 2010s.

Figure 2.7: Indicator Results: Upstream Chains of Substitutability and Complementarity.



Note: the figure shows the change in value of shipments / value added based on linear combinations of $\beta_{ucomp} + \beta_{ucomp,t}$ and $\beta_{usub} + \beta_{usub,t}$ for $t \in [1968, 2018]$ moving from the 25th percentile to the 75th percentile. Only estimates significant at the 95 percent level are pictured. Successive points are connected by confidence areas.

Figure 2.8: Indicator Results: Downstream Chains of Substitutability and Complementarity.

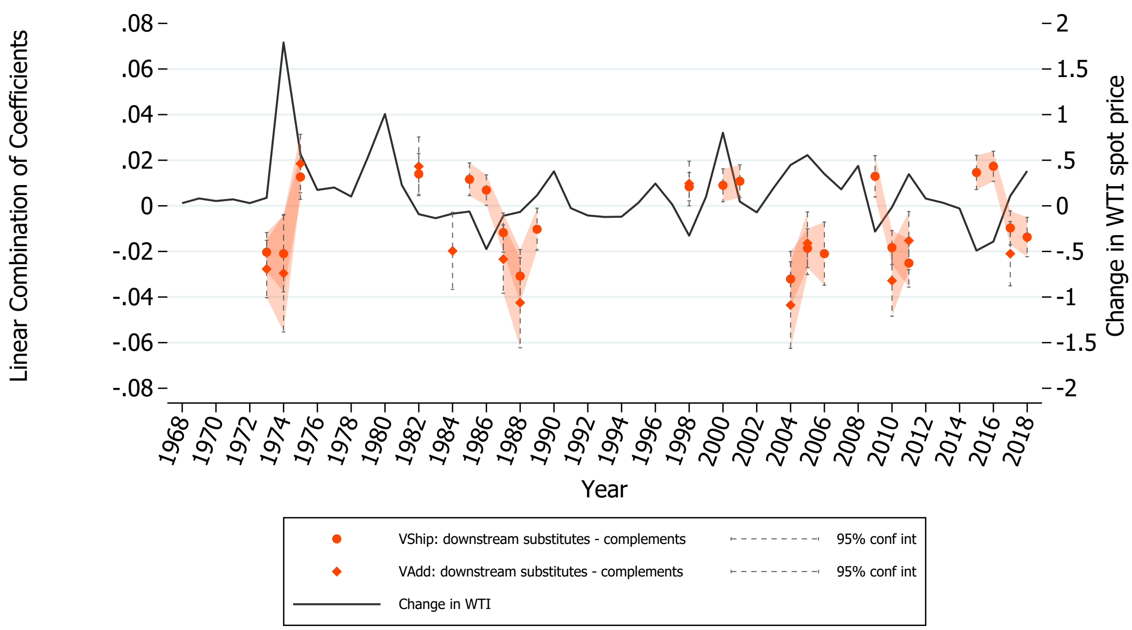


Note: the figure shows the change in value of shipments / value added based on linear combinations of $\beta_{dcomp} + \beta_{dcomp,t}$ and $\beta_{dsub} + \beta_{dsub,t}$ for $t \in [1968, 2018]$ moving from the 25th percentile to the 75th percentile. Only estimates significant at the 95 percent level are pictured. Successive points are connected by confidence areas.

Figure 2.9: Difference Between Upstream Substitutability/Complementarity Indicators.

Note: the figure shows linear combinations of $\beta_{usub} + \beta_{usub,t} - \beta_{ucomp} - \beta_{ucomp,t}$ for $t \in [1968, 2018]$. Only estimates significant at the 95 percent level are pictured. Successive points are connected by confidence areas.

Figure 2.10: Difference Between Downstream Substitutability/Complementarity Indicators.



Note: the figure shows linear combinations of $\beta_{dsub} + \beta_{dsub,t} - \beta_{dcomp} - \beta_{dcomp,t}$ for $t \in [1968, 2018]$. Only estimates significant at the 95 percent level are pictured. Successive points are connected by confidence areas.

The results show a shifting-away from petroleum products as the change in the oil price either peaked or was declining, including in 1975-1976, 1981-1982, 1985-1986, 1990-1991, 1997-1998, 2000-2001, 2009, and 2015-2016. In the converse of the above, this is evidenced by periods (1) when the downstream substitute (complements) paths indicator becomes positive (negative) and statistically significant (Figure 2.8) and/or (2) when the difference between the downstream substitute and complement paths coefficients becomes positive and statistically significant (Figure 2.10).

That these patterns occur during periods of inflection for the oil price (from increasing to decreasing) suggests that the shifting away from petroleum products can potentially be explained by both of the interpretations described above. On one hand, given that all of these years follow periods in which the oil price had been increasing (or started to decrease at a slower rate), this shifting away can be interpreted as an attempt by industries to reduce their intermediate input costs in response. At the same time, a reduction in demand for petroleum products—by reducing demand for oil—itself likely put downward pressure on the price. Taken together, it could be that this shifting-away from petroleum products was at first an effect of the previous increase, and then a cause of the subsequent decrease, in the oil price during these periods.

2.5.3 Summary of Oil Price Episodes Results

Overall, there are two main takeaways from the results: (1) that industries experienced heterogeneous outcomes during oil price episodes, and these outcomes were closely tied to both their places and their relationships in the production network; and (2) that oil price episodes have varied in terms of their causes and consequences.

With regards to the first of these, the results provide empirical evidence for the relevance of the connectivity and the evolutionary aspects of the production network in propagating oil shocks throughout the U.S. economy. Beyond this, that the connectivity and evolutionary indicators frequently become statistically significant at the same time further suggests both (1) that these indicators are capturing different types of information and (2) that there are multiple mechanisms at work even within the same episode.

In terms of magnitudes, the evolutionary indicators often differentiate industries on a scale equal to, or surpassing, the connectivity indicators. This reinforces the finding in Chapter 1 and in some previous research that substitution patterns are a primary determinant of industry outcomes in the presence of shocks.

With regards to the oil episodes themselves, the results broadly suggest two patterns. The first is that many episodes involving oil price increases appear to be driven (at least partially) by industries' intermediate input demand. In turn, the shifting-away from petroleum products on industries' downstream sides that we would expect in supply-driven oil episodes generally does not appear until prices are peaking or declining.

The second is that manufacturing industries were sometimes affected by these episodes through the propagation of price changes from upstream to downstream, even if a particular episode itself was not primarily supply-driven. In addition, there is evidence for a number of episodes that at least a portion of this supply-side pass-through was the direct result of heterogeneity in industries' ability to substitute away from petroleum products.

Considering final—rather than intermediate—demand shows a somewhat different pattern. Specifically, while intermediate demand seems to be a driver of prices, prices seem to be a driver—at least partially—of some final demand, and in particular, demand for extractive activities and for automobiles. In the first case, the oil price appears to affect demand for certain manufacturing outputs through its impact on the U.S. oil and natural gas rig count. In the second case, the oil price appears to affect consumer demand for automobiles, though in most instances, the overlap of this demand shift and the occurrence of U.S. recessions adds ambiguity to the true proximate cause.

Finally, although shock transmission via the supply-side channel during an episode does not necessarily imply a supply-related cause for the associated oil price change, to the extent we interpret the results in this way, the patterns of the evolutionary indicators suggest a decreasing role for such supply-driven price changes over the past few decades. This stands in contrast to demand-side price drivers, which appear to have remained relevant throughout the entire study period.

2.5.4 Clustering Years Based on Indicator Results

As a final way to consider the results, I use hierarchical clustering to split the years in the study period into several groups. This method works by first assigning each year to its own group, and then iteratively combining the two closest groups (based on a subset of the indicators; see below) until only a single group remains.[33] Working backwards, the algorithm is then able to specify which years belong to which groups when moving from one group to two groups, two groups to three groups, and so on. There are various techniques, also known as stopping rules, to determine when to conclude this splitting process to yield the optimal number of groups (see below).

In essence, this approach provides the clustering algorithm with the results in the previous subsections and asks: which years are most similar to one another? It does this by considering

[33]Specifically, I use the hierarchical agglomerative linkage clustering algorithm as implemented in Stata. I apply Ward's method to perform the clustering, which by default in Stata uses squared Euclidian distance.

multiple indicators—for both value of shipments and value added—all at once, which provides a complement to the indicator-by-indicator discussion above.

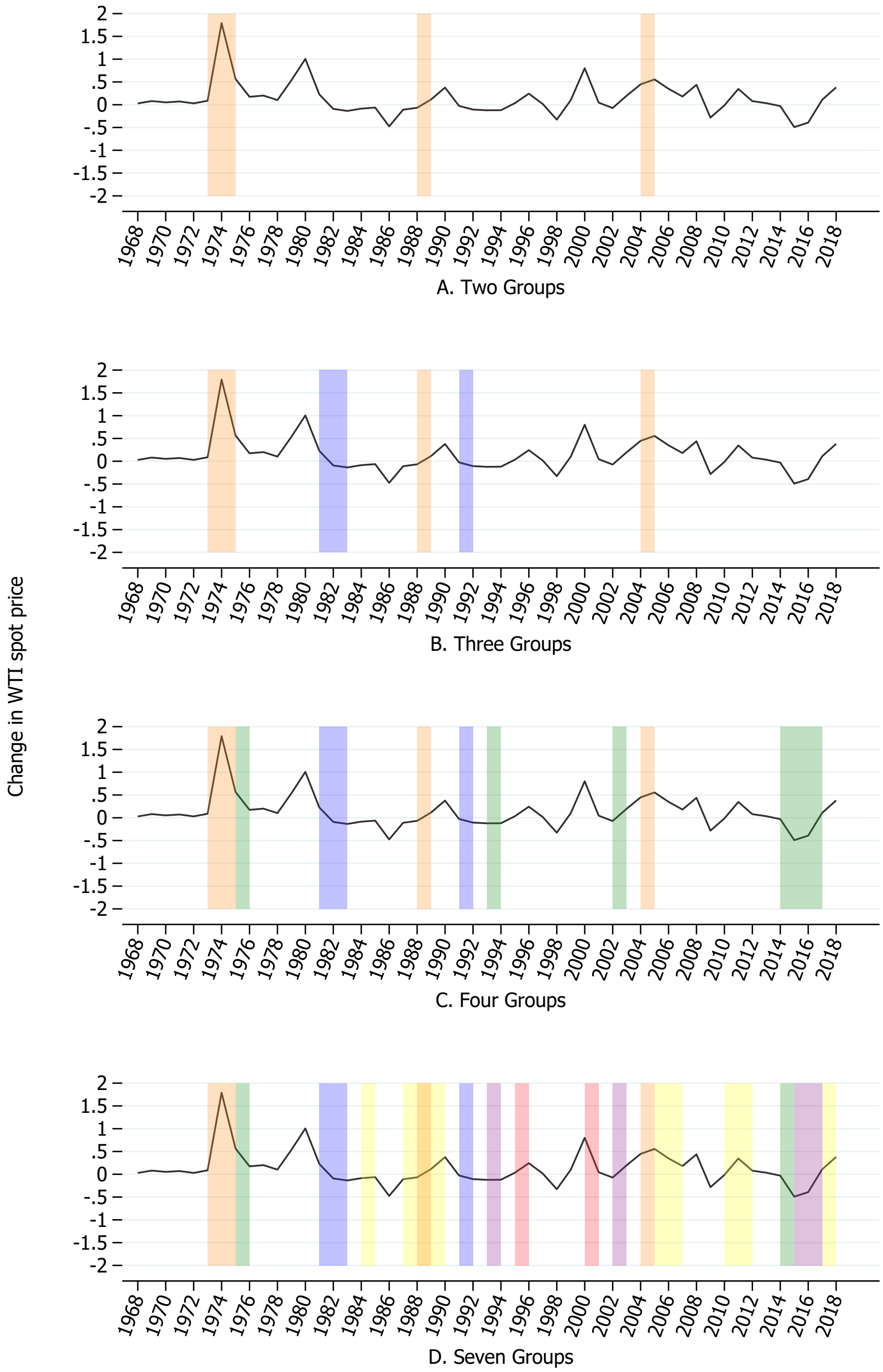
To implement this approach, I first create a dataset by (1) gathering together, for each year 1968-2018, the linear combinations for supply paths to refineries, upstream substitute and complement paths (and their differences), and downstream substitute and complement paths (and their differences) and (2) setting any combinations that are not statistically significant at the 95 percent level to zero. I exclude the demand paths indicators because, per the discussion above, they are more indirectly related to oil price episodes than are the other indicators. I use a 95 percent confidence level to match the results in the previous subsections and to filter out all but the most significant combinations across the study period.

This yields a dataset that associates each year with seven linear combinations for value of shipments and seven linear combinations for value added, some (or many) of which may be zero. I then apply hierarchical clustering to split the study period into different groups based on the (dis)similarity of these linear combinations across years. I consider both the Calinski-Harabasz and the Duda-Hart stopping-rules. The first of these is global in the sense that it considers information about all groups (regardless of where the algorithm is in the splitting process), while the second is local in that it considers the cluster structure of the next group to be split. The Calinski-Harabasz rule suggests that the study period be split into seven groups or groups of much larger size (more than 20). The Duda-Hart rule suggests that the study period be split into two, three, four, seven, or nine groups, or several groups of size 12 or larger.

For comparison purposes and for interpretability, I consider how the algorithm splits the years when I specify there to be two, three, four, or seven groups. In Figure 2.11, the two-group results are shown in panel A, the three-group results are shown in panel B, the four-group results are shown in panel C, and the seven-group results are shown in panel D. Note that, in each panel, one of the groups is represented by the non-highlighted years, such that panel A has one highlight color (orange), panel B has two highlight colors (orange and blue), and so on. I call these non-highlighted years the "reference group" within each panel.

We see that if the algorithm is directed to split the years into two groups, it distinguishes 1973-1974, 1988, and 2004 as being different from the remainder of the years in the study period (Figure 2.11, panel A). The key common factor across these years is a substantial increase in the oil price compared to the preceding period. Per the results above, all three also correspond to strong patterns of intermediate input use shifting towards petroleum products. That these periods are grouped together echoes the previous discussion that the 1973-1974 episode was likely similar to

Figure 2.11: Study Period Years Grouped by Hierarchical Clustering Algorithm.



Note: the figure shows the years in the study period grouped by a hierarchical clustering algorithm based on the linear combinations for supply paths to refineries, upstream substitute and complement paths (and their difference), and downstream substitute and complement paths (and their difference). There are two groups in panel A, three groups in panel B, four groups in panel C, and seven groups in panel D. In each panel, one of the groups is represented by the non-highlighted years, such that panel A has one highlight color (orange), panel B has two highlight colors (orange and blue), and so on.

later price increases in its causes/consequences, which stands in contrast to the historical view that the former tended to be supply-driven while the latter tended to be demand-driven.

If the algorithm is directed to split the years into three groups, it leaves the highlighted group from panel A unchanged, but splits out two periods (in blue) from the reference group: 1981-1982 and 1991 (Figure 2.11, panel B). As described above, both of these periods experienced substantial drops in the oil price. They also coincide with U.S. recessions and marked the beginning of stretches of general price decreases. The indicators suggest these were times when intermediate demand was shifting away from petroleum products, accompanied by the pass-through of decreasing prices evidenced by both the supply paths indicator and the upstream evolutionary indicators.

Adding another group, the algorithm retains the highlighted groups from panels A and B, but splits out four other periods (in green) from the reference group: 1975, 1993, 2002, and 2014-2016 (Figure 2.11, panel C). Unlike the years in the other groups, the years in the green group are marked by inflections in the oil price where it was first decreasing—or increasing at a slower rate—and then subsequently reversed course. I note that 1975 occurs earlier within this down-then-up pattern, 1993 and 2002 occur later within this pattern, and the interval 2014-2016 includes both a decrease and an increase (see additional discussion below).

Finally, moving from four groups to seven groups introduces several changes: (1) 1995 and 2000 are split from the reference group to form a new group (in red); (2) 1984, 1987/1989, 2005-2006, 2010-2011, and 2017 are split from the reference group to form a new group (in yellow); and (3) 1993, 2002, and 2015-2016 are split from their original group (in green) to form their own group (in purple; Figure 2.11, panel D). The red years and the yellow years both coincide with times of oil price increases. The former are distinguished by connectivity supply-side pass-through (as measured by the supply paths indicator) without an accompanying shift towards petroleum products.[34] In contrast, the latter are characterized by such a shift towards petroleum products, reflecting dynamics similar to those in the orange years (with the main difference between the yellow and orange years being the magnitudes of the relevant indicators).

The split of the purple years (1993, 2002, and 2015-2016) from the green group mirrors the differences noted above. Specifically, the years remaining in the green group—1975 and 2014—coincide with periods when the growth rate of the oil price was declining (just before inflecting), while those in the purple group—1993, 2002, and 2015-2016—coincide with periods when the growth rate of the oil price was increasing (just after inflecting).

[34] In fact, the downstream evolutionary indicators suggest that intermediate input use was shifting away from petroleum products in 2000-2001.

Considering the two-, three-, four- and seven-group configurations together suggests several take-aways. The first is that many oil price increases appear to have shared similar features (at least as reflected in the given network indicators), including the episodes in 1973-1974, in the late 1980s, and before and after the Great Recession. In the two-group configuration this is illustrated by the years highlighted in orange, while in the seven-group configuration it is further suggested by the years highlighted in yellow. For both of these groups, the indicators suggest that the corresponding price increases were at least partially driven by domestic intermediate demand. In contrast, although the years highlighted in red also coincide with times of oil price increases, they are distinguished by strong connectivity pass-through on the upstream side and lack of evidence of a shifting towards petroleum products. Although more research is needed, this latter finding suggests that the corresponding price increases may have been driven by supply-side factors and/or by dynamics exogenous to the U.S. economy.

The blue years, green years, and purple years (in the seven-group configuration) all coincide with periods of oil price decreases or decreases-then-increases, though they are classified in separate groups because the underlying dynamics appear to be different. In particular, the blue and purple years are similar in that some industries benefitted from the supply-side pass-through of decreasing prices via the connectivity aspect of the network. The main difference between them is that the upstream evolutionary indicators provide additional evidence for the blue years of such pass-through occurring due to industries' substitution possibilities. Somewhat counter-intuitively, the opposite pattern emerges in the green years, where—despite decreases in the oil price—the upstream evolutionary indicators suggest that some industries performed relatively worse due to their inability to substitute away from petroleum products. Although more research is needed, these results provide evidence that the trend in the oil price, on its own, is not sufficient to predict the potential heterogeneous effects experienced by industries.

Altogether, the groupings illustrated in Figure 2.11 suggest that the clustering algorithm is able to—based only on sets of linear combinations for each year—identify and categorize many of the oil price episodes over the past half century. In turn, these findings provide additional support both for the interpretations in the previous subsections and for the relevance of the indicators themselves.

2.5.5 Alternative Context: Information Technology

Lastly, I explore an application of the empirical approach to the information technology context, where instead of considering industries' connections to the petroleum refineries, oil/gas extraction, and automobile manufacturing industries, I consider their connections to the 6-digit-NAICS industry "data processing services" and the 3-digit-NAICS industry "data processing, internet publishing, and other information services." I shorten the study period to encompass the three decades between 1988 and 2018 (see below).

Although there are potentially many industries that could be relevant for this analysis, I choose these two industries as broadly capturing components of what we might think of as information technology. Specifically:

- The data processing services industry includes processing/preparation of reports from customer data, automated data entry, computer time rental, and optical scanning services.[35]
- The data processing, internet publishing, and other information services industry broadly encompasses businesses that supply, store, process, and/or provide access to information, ranging from news syndicates to Internet service providers to establishments that provide electronic mail services, bulletin boards, browsers, and search routines.[36]

As with the oil/gas extraction and automobile manufacturing industries, I conceptualize the data processing services industry as affecting manufacturing industries through the downstream (demand) channel. As such, I create an indicator by counting the number of directed paths from each industry i to the data processing services industry:

$$\text{demand\_paths\_data}_{it} = \text{num\_paths}(i, data|A_t)$$

where $data$ represents the data processing services industry, A_t is the 6-digit-NAICS direct requirements table for the year t, and $\text{num\_paths}(x, y|A_t)$ is the number of paths from x to y calculated using depth-first search over the directed network represented by A_t.[37]

I calculate the upstream and downstream substitute and complement indicators as before, with the modification that the pairwise substitutability/complementarity estimates from which I generate these indicators reflects industries' estimated input patterns with respect to the data processing and information services industry. For each industry i and year t, call these indicators $\text{down\_comp\_paths\_IT}_{it}$, $\text{down\_sub\_paths\_IT}_{it}$, $\text{up\_comp\_paths\_IT}_{it}$, and $\text{up\_sub\_paths\_IT}_{it}$.

Analogous to the oil context, I run the following regression:

[35] See details for industry 514200: https://www.census.gov/naics/resources/archives/sect51.html.

[36] See details for industry 514: https://www.census.gov/naics/resources/archives/sect51.html.

[37] As before, I use the Python library NetworkX to determine the number of paths, and only paths of length three or less are included in the total.

$$
\begin{aligned}
y_{it} = \alpha + Y_t + 2\text{digit}_i \cdot Y_t + \beta_{emp} \cdot \text{emp}_{it} + \\
&\beta_{direct} \cdot \log(\text{direct}_{h(i),t-2}) + \beta_{direct,t} \cdot \log(\text{direct}_{h(i),t-2}) \cdot Y_t + \\
&\beta_{data} \cdot \log(\text{demand\_paths\_data}_{i,g(t)}) + \beta_{data,t} \cdot \log(\text{demand\_paths\_data}_{i,g(t)}) \cdot Y_t + \\
&\beta_{dcomp} \cdot \log(\text{down\_comp\_paths\_IT}_{h(i),t-2}) + \beta_{dcomp,t} \cdot \log(\text{down\_comp\_paths\_IT}_{h(i),t-2}) \cdot Y_t + \\
&\beta_{dsub} \cdot \log(\text{down\_sub\_paths\_IT}_{h(i),t-2}) + \beta_{dsub,t} \cdot \log(\text{down\_sub\_paths\_IT}_{h(i),t-2}) \cdot Y_t + \\
&\beta_{ucomp} \cdot \log(\text{up\_comp\_paths\_IT}_{h(i),t-2}) + \beta_{ucomp,t} \cdot \log(\text{up\_comp\_paths\_IT}_{h(i),t-2}) \cdot Y_t + \\
&\beta_{usub} \cdot \log(\text{up\_sub\_paths\_IT}_{h(i),t-2}) + \beta_{usub,t} \cdot \log(\text{up\_sub\_paths\_IT}_{h(i),t-2}) \cdot Y_t + \epsilon_{it}
\end{aligned}
$$

where y_{it} is either the percentage change in the value of shipments or the percentage change in value added, α is a constant, Y_t is a set of year fixed-effects, 2digit_i is a set of 2-digit-NAICS fixed effects, emp_{it} is the industry-level change in employment,[38] direct_{it} is the fraction of industry i's inputs purchased from data processing in year t, $g(t)$ maps each year t to one of {1987, 1997} (whichever most closely precedes t), $h(i)$ maps each 6-digit-NAICS industry i to its 3-digit-NAICS counterpart, and the other regressors are generated as described above.

I calculate some summary statistics for each of these indicators (Table 2.6) as well as the pairwise correlations between the indicators (Table 2.7). As in the oil context, I take logs of all of the indicators. The table of correlations suggests that the indicators are capturing distinct pieces of information regarding industries' relationships to information technology.

Table 2.6: Summary Statistics for Indicators in Information Technology Context.

Variable	Mean	Std. Dev.	Min.	Max.
demand_paths_data	0.661	2.177	0.000	17.000
down_comp_paths_IT	0.075	0.096	0.000	0.509
down_sub_paths_IT	0.074	0.119	0.000	0.428
up_comp_paths_IT	0.115	0.102	0.000	0.389
up_sub_paths_IT	0.049	0.056	0.000	0.436

The results are presented below (Figures 2.12-2.16). As in the oil context, I plot against the left axis the change in value of shipments / change in value added for each year based on (1) the linear combination of the non-interacted and interacted coefficients for each indicator when (2) moving from the 25th percentile to the 75th percentile in the values for that indicator. Given that the 25th and 75th percentiles for demand_paths_data are zero, I use $\log(\text{pct}(\cdot, 90)/\min_{>0}(\cdot))$ as the multiplier

[38] As above, this is the change in employment as compared to the average over the previous two years.

Table 2.7: Pairwise Correlations of Indicators in Information Technology Context.

	(1)	(2)	(3)	(4)	(5)
(1) demand_paths_data	1.00				
(2) down_comp_paths_IT	0.06	1.00			
(3) down_sub_paths_IT	-0.01	-0.18	1.00		
(4) up_comp_paths_IT	-0.08	0.12	-0.01	1.00	
(5) up_sub_paths_IT	0.09	0.46	-0.01	0.01	1.00

for this indicator (where, as above, $\text{pct}(\cdot,\cdot)$ represents the percentile operator and $\min_{>0}(\cdot)$ finds the smallest value greater than zero).

Finally, for clarity, I again only show the linear combinations in each year that are statistically significant at the 95 percent level (standard errors are clustered at the industry level). For comparison purposes, I include in the appendix a set of figures that display linear combinations in each year that are statistically significant at the 80 percent level.

The coefficients on the non-interacted versions of the indicators are generally statistically significant and reflect that, over this time period: (1) manufacturing industries with a greater number of demand paths to data services tended to do better year-over-year than other industries; and (2) industries with a greater number of substitute (complement) paths on their upstream and downstream sides tended to do better (worse) year-over-year than other industries (Table 2.8). Taken together, these patterns suggest that some manufacturing industries benefitted during the study period in their role as suppliers to a growing sector of the economy (i.e., information technology taken broadly), while some experienced counteracting effects as IT goods and services became more expensive relative to other intermediate inputs during this time (see below).

Unlike the oil episodes context, there is not a single price that can be used to assess supply and demand as it relates to information technology. As an alternative, I look to the NASDAQ Composite Index, which I take to be a proxy of both (1) general demand for information technology and (2) the relative cost of information technology as an intermediate input from the perspective of manufacturing. In turn, for illustrative purposes, I plot against the right axis in each figure the change in the NASDAQ Composite Index (where the change is calculated against the average of the prior two years). Note that, as with the WTI spot price above, the NASDAQ values are shown for comparison purposes only and were not used in the analysis in any way.

Information Technology Results

Given that the digital revolution and the information technology industry as we know it today did

Table 2.8: Coefficients and Standard Errors for Non-Interacted Terms in IT Context.

	Value of shipments	Value added
# of demand paths to data services	0.00311***	0.00488**
	(0.00117)	(0.00200)
downstream complement paths	-0.00510	-0.00878
	(0.00446)	(0.00765)
downstream substitute paths	0.0193***	0.0235***
	(0.00364)	(0.00614)
upstream complement paths	-0.0208***	-0.0305***
	(0.00475)	(0.00857)
upstream substitute paths	0.0158***	0.0272***
	(0.00432)	(0.00762)
N	10,137	10,137
adj. R^2	0.622	0.448

Standard errors clustered by industry.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

not begin until the 1970s and 1980s, I shorten the study period to the 31 years between 1988 and 2018.[39] In turn, I am able to base the demand paths indicator on snapshots of the U.S. production network in 1987 and 1997, while I estimate the input substitutability/complementarity relationships using data for the years 1986-2016.

I note that one significant event for the information technology industry during this time was the "dotcom bubble," which emerged during the latter half of the 1990s and burst beginning in 2000. The impacts of this event—in terms of the effects on manufacturing industries—are reflected in different ways across the figures (see below).

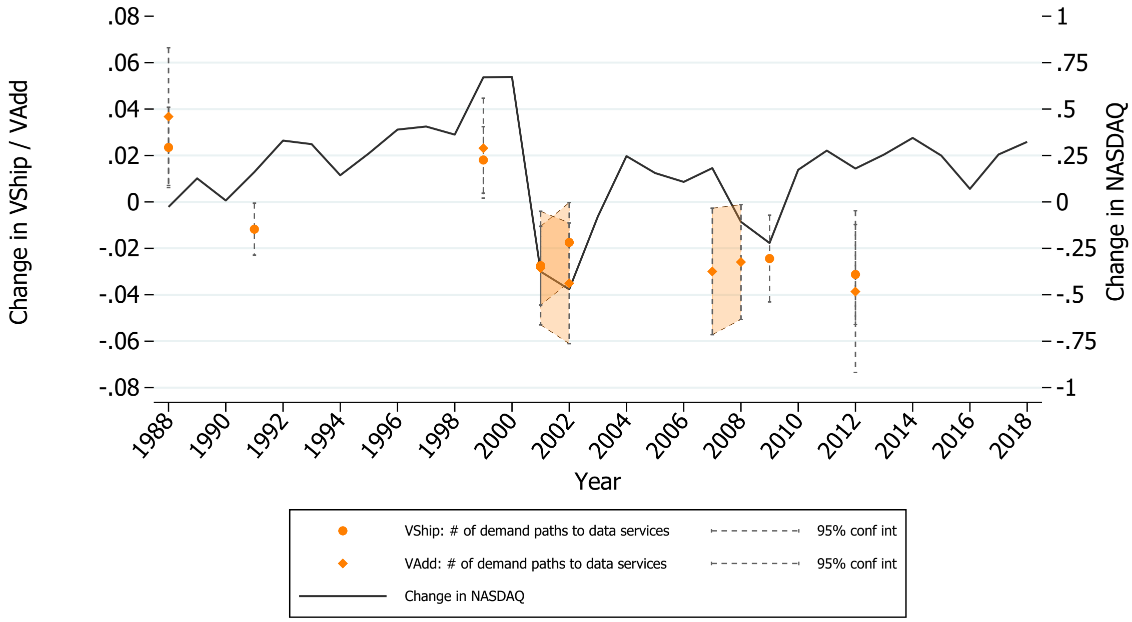
Similar to the indicators in the oil episodes context, I take demand paths to data services—through measuring the connections downstream of industries—as becoming relevant when there are changes to demand for such activities as channeled from downstream in the network. In other words, this indicator serves as a proxy for shock transmission via the demand-side channel.

Demand paths to data services becomes positive and statistically significant at the 95 percent level two times during the study period, while it becomes negative and statistically significant four times (Figure 2.12). These patterns generally align with the change in the NASDAQ index, which suggests that during periods of information technology growth (decline), manufacturing industries

[39] I note that the 3-digit-NAICS input-output tables contain non-zero values for the data processing, internet publishing, and other information services industry beginning only in the 1970s.

with a greater exposure to data processing services on their downstream sides did better (worse) than other industries.

Figure 2.12: IT Indicator Results: Demand Paths to Data Processing Services.



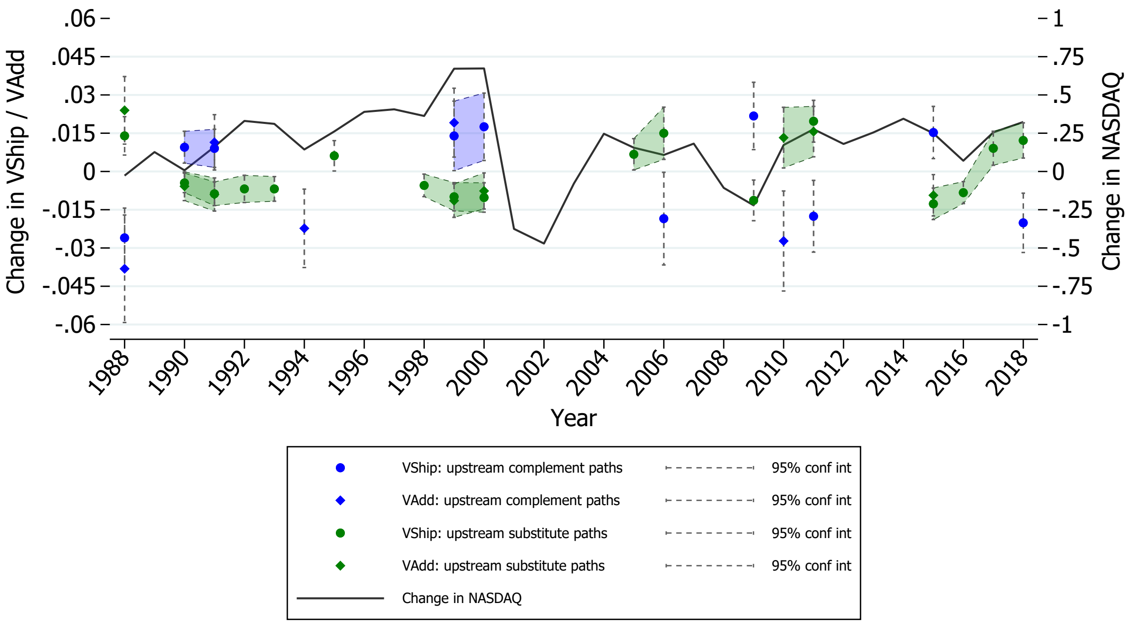
Note: the figure shows the change in value of shipments / value added based on linear combinations of $\beta_{data} + \beta_{data,t}$ for $t \in [1988, 2018]$ moving from the smallest value greater than zero to the 90th percentile. Only estimates significant at the 95 percent level are pictured. Successive points are connected by confidence areas.

I note that 2001 and 2007-2009 both correspond to periods of U.S. recession, which by themselves could explain the negative indicator values during those years. However, the results at the 80 percent significance level (in the appendix) provide some additional evidence that this indicator captures an IT-specific relationship beyond what can be explained by general macroeconomic trends. This is especially the case for the years before and after the dotcom bubble (1999-2003), though more research would be needed to understand these patterns fully.

Turning to the evolutionary indicators, we see that—unlike the oil context—the upstream indicators tend to become statistically significant more often than do the downstream indicators (Figures 2.13 through 2.16). This broadly suggests that, as far as their relationship with information technology is concerned, U.S. manufacturing industries were more affected via the supply-side channel than via the demand-side channel over this time.

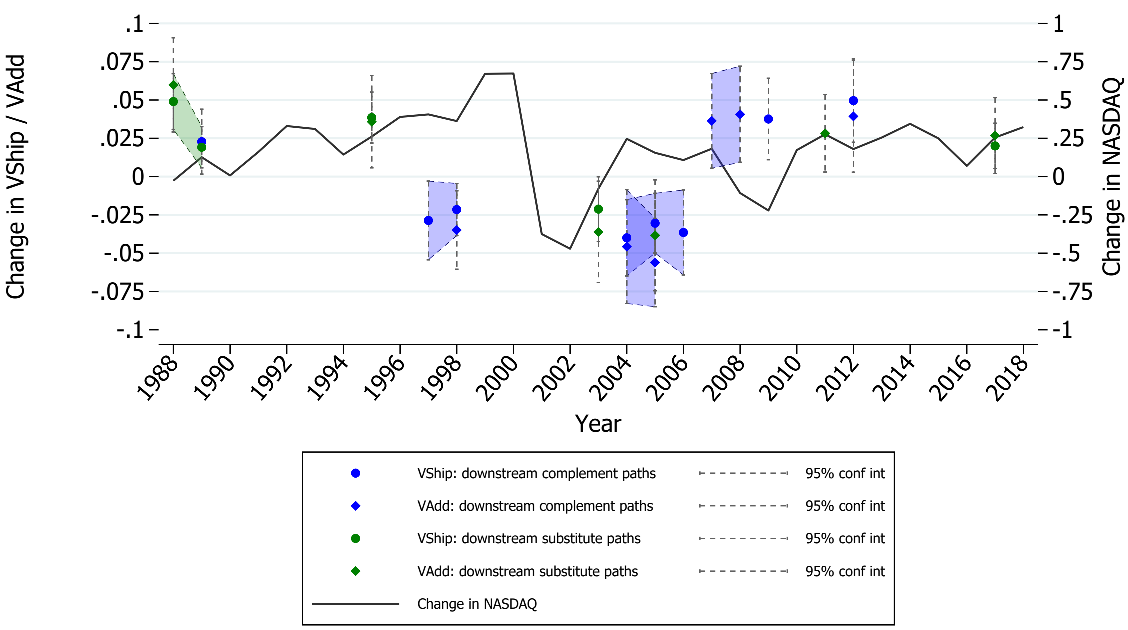
Specifically, the upstream evolutionary indicators most clearly suggest the pass-through of decreasing IT costs during three periods: the first few years of the 1990s, the last few years of the 1990s (including 2000), and the mid-2010s. The first of these corresponds to a temporary drop in

Figure 2.13: IT Indicator Results: Upstream Chains of Substitutability and Complementarity.



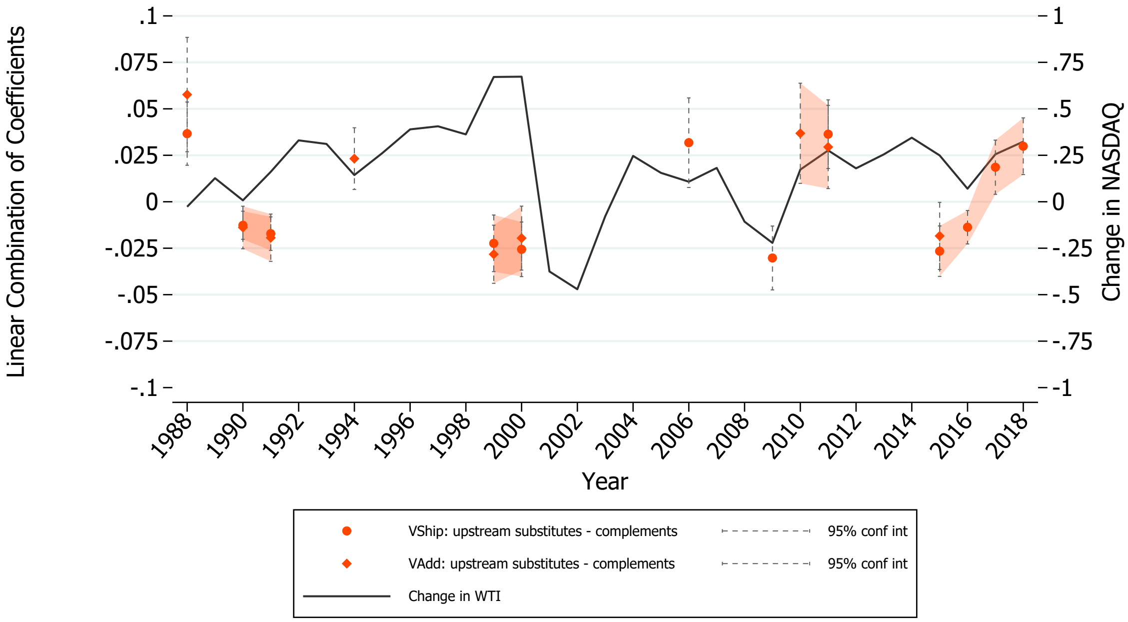
Note: the figure shows the change in value of shipments / value added based on linear combinations of $\beta_{ucomp} + \beta_{ucomp,t}$ and $\beta_{usub} + \beta_{usub,t}$ for $t \in [1988, 2018]$ moving from the 25th percentile to the 75th percentile. Only estimates significant at the 95 percent level are pictured. Successive points are connected by confidence areas.

Figure 2.14: IT Indicator Results: Downstream Chains of Substitutability and Complementarity.



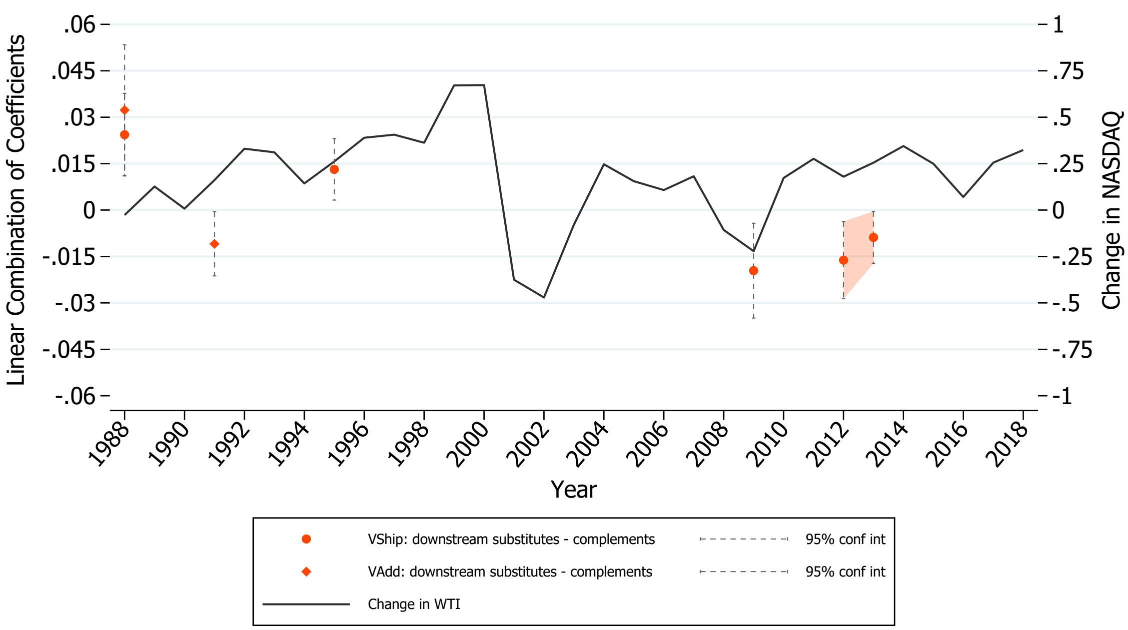
Note: the figure shows the change in value of shipments / value added based on linear combinations of $\beta_{dcomp} + \beta_{dcomp,t}$ and $\beta_{dsub} + \beta_{dsub,t}$ for $t \in [1988, 2018]$) moving from the 25th percentile to the 75th percentile. Only estimates significant at the 95 percent level are pictured. Successive points are connected by confidence areas.

Figure 2.15: Difference Between Upstream IT Substitutability/Complementarity Indicators.



Note: the figure shows linear combinations of $\beta_{usub} + \beta_{usub,t} - \beta_{ucomp} - \beta_{ucomp,t}$ for $t \in [1988, 2018]$. Only estimates significant at the 95 percent level are pictured. Successive points are connected by confidence areas.

Figure 2.16: Difference Between Downstream IT Substitutability/Complementarity Indicators.



Note: the figure shows linear combinations of $\beta_{dsub} + \beta_{dsub,t} - \beta_{dcomp} - \beta_{dcomp,t}$ for $t \in [1988, 2018]$. Only estimates significant at the 95 percent level are pictured. Successive points are connected by confidence areas.

the NASDAQ index, which also aligns with the U.S. recession in 1990-1991. The second occurs in the period leading up to the burst of the dotcom bubble. That the indicators become significant beginning in 1998 suggests that manufacturing industries may have already benefitted from reduced IT costs at this point—as channeled via the supply-side—even though the NASDAQ index did not peak until early 2000. The last of these corresponds to just before and during a period when the NASDAQ leveled (2015) and then temporarily declined (2016).

Conversely, the upstream evolutionary indicators suggest the pass-through of increasing (rather than decreasing) IT costs most prominently at three times: 2005-2006, 2010-2011, and 2017-2018. All three of these correspond to periods when the NASDAQ had recovered after previous drops: after the dot-com bubble (and 2001 recession), after the Great Recession, and after the 2015-2016 leveling/decline described in the previous paragraph.

Lastly, the downstream evolutionary indicators tend to suggest a shifting-away from data processing as an intermediate input during periods of IT growth (for example, in 1995, 1997-1998, and 2004-2006), while they indicate a shifting-towards during periods of decline (most prominently in 2007-2009). These patterns are consistent with supply-driven (rather than demand-driven) cost changes that affect industries via changes in intermediate demand.

Taken together, the evolutionary indicators suggest that manufacturing industries may have been affected not only through the supply-side channel, but that the associated IT cost changes—unlike many cases in the oil episodes context—may themselves have been supply-driven.

Overall, these indicators—based on a part of the production network separate from petroleum refineries, oil/gas extraction, and auto manufacturing—appear to contain their own set of local and distinct information. In this way, they suggest both (1) that these results (for IT) and those above (for oil) are reflecting different shocks as they propagate in the U.S. economic production network, and (2) that the approach in this chapter may be relevant in contexts beyond oil episodes.

2.6 Conclusion

In this chapter, I aim to investigate two primary questions. First, when an oil price episode occurs, how does an industry's place in the production network mediate the impacts it feels? Second, can we "see" oil price episodes through the lens of differentiated industry impacts, and if so, what does this tell us about the episodes themselves?

With regards to the first question, the results suggest that industry outcomes are impacted by both the connectivity and the evolutionary aspects of the production network in the presence of shocks. Said another way, it is the arrangement of the connections among industries—as well as how those connections change in response to shocks—that are relevant for industry outcomes during these times. The results further suggest that, in terms of economic magnitude, the latter of these is just as important, if not more important, than the former.

In addition, the connectivity and evolutionary indicators frequently become statistically significant at the same time, which suggests not only that the indicators are reflecting different types of information, but that multiple network dynamics are often at work simultaneously even in the context of a single shock. Additional research, perhaps through the use of various other combinations of indicators, could help to further disentangle these relationships.

With regards to the second question, the results provide evidence both (1) that oil price episodes impacted industries differentially and (2) that the episodes themselves were associated with heterogeneous causes and consequences. On this latter point, the results suggest a key role for demand-driven price increases throughout the 1970s, the 1980s, the 2000s, and the 2010s, which corroborates some of the recent work in the oil episodes literature that has attributed oil price increases to demand factors in addition to (or in place of) supply factors. At the same time, the results also suggest a strong role for the supply-side transmission of shocks during certain episodes, including the finding that industries were differentially impacted by the ability (or inability) of their upstream suppliers to shift away from petroleum products during some times.

If we take such supply-side transmission to be indicative not just of a mechanism by which industries were affected, but also as a sign of the underlying causes of the associated oil price changes, then the results further suggest (1) a substantial supply-side driver before and/or after the 1973-1974, 1979-1980, and 1990 episodes, and (2) a decrease in the relative importance of supply-related factors—as compared to demand-related factors—in causing oil price fluctuations over the last several decades.

Finally, as discussed in the introduction, both the approach and the findings of this chapter may be able to inform the retrospective evaluation of past policies and/or the development of future policies. For instance, similar to the information technology exercise in the current chapter, the indicators developed here could also be applied to other contexts. In addition, to the extent that oil price shocks provide a proxy for the potential effects of energy and environmental policies, the specific coefficient estimates from the oil episodes empirical analysis could be used to predict industry-level differentiation in response to the enactment of such policies.

Chapter 3

Green Jobs and Occupational Transitions: An Empirical Exploration

3.1 Introduction

Evolution towards a U.S. economy that exhibits greater energy efficiency, uses more renewable resources, and produces less pollution will require both structural and technological changes. One of the largest concerns surrounding such a transition—and surrounding laws and regulations that intend promote it—is the potential effect on employment.

Specifically, an important aspect of the movement towards a "green economy" is the potential to affect industries heterogeneously. Industries related to renewable energy, energy efficiency, environmental policy, and engineering of new technologies, among many others, may stand to benefit. On the other hand, we might expect industries that rely heavily on fossil fuels, and/or produce greater quantities of pollution (including carbon emissions), to be negatively affected—at least in the short term—by such a transition.

In both cases, occupations associated with these industries will also be impacted. Some workers in declining industries will transition out of their occupations or may become unemployed, while nascent or burgeoning industries will attract flows of workers into the occupations with which they are associated. Exploring these movements is the focus of this two-part study.

In the first part, I take the evolution of U.S. industries during the period 2001-2012 as given and train a series of machine learning models that relate this structure to historical occupation-to-occupation flows. Although there are many reasons why workers would transition from one occupation to another, I take regional-industrial factors (such as state-by-industry GDP and changes to GDP) as the primary drivers, and in so doing, attempt to capture the systematic relationship between these factors and occupational transitions. I test the models by examining their predictive power for a hold-out sample covering the years 2013-2016.

In the second part of the study, I consider the implications of several scenarios of economic change for a focus set of 84 occupations. I choose these particular occupations because they have been categorized in the previous literature as either "green" or "brown," depending on a combination of their titles, the tasks they perform, the industries in which they are employed, and/or their general prognosis for growth/decline in a transition to a green economy.

Specifically, I form this list of 84 occupations from a subset of (1) green occupations identified by the U.S. Department of Labor's Occupational Information Network (O*NET; Dierdorff et al., 2009, 2011) and (2) a set of green occupations and a set of brown occupations identified by Vona et al. (2018). The O*NET green occupations are separated into three categories ("green increased demand," "green enhanced skills," and "green new and emerging"), while each of the Vona et al. (2018) green occupations is associated with a "greenness" measure based on the tasks that the occupation performs.

I train the machine learning models to predict the flows between these 84 occupations and all of the other occupations with which they exchange workers (including unemployment),[40] where the latter group may include, but is not restricted to, the 84 focus occupations themselves. I use the trained models to predict the change in net inflow—that is, the difference between total flows of workers into and out of an occupation—for each of the focus occupations under each of the scenarios.

The predicted change in net inflow represents the overall anticipated fluctuation in demand for an occupation vis-à-vis the other occupations with which it exchanges workers. I use this metric (expressed in percentage terms) to rank the 84 occupations, which I interpret as an ordering that ranges from "most green" (positive change in net inflow) on one end to "least green" (negative change in net inflow) on the other. In this sense, I define greenness not by characteristics of the occupations themselves or by features of the industries in which they work, but by their predicted growth or decline under the various scenarios. In such a classification scheme, jobs that would otherwise be considered brown—or be thought of as outside of the green-brown spectrum—can be placed onto either the green end or the brown end of the spectrum based on their potential growth trajectory in the overall movement towards a green economy.

I choose the scenarios themselves to represent different ways in which the greening of the economy may affect industries. Specifically, of the five scenarios I consider, two are based on a list of "brown industries" identified by Vona et al. (2018).[41] These industries—which include much of manufac-

[40]Specifically, I train two models for a pair of occupations o and p (one model for transitions o-to-p and another model for transitions p-to-o) if there are at least two transitions per year, on average, between the two occupations (in one direction or the other, or both) during the study training period 2001-2012. See section 3.3.1 for additional details.

[41]The authors use this list of brown industries to identify the occupations they categorize as brown. See section

turing as well as parts of mining, oil/gas extraction, utilities, wholesale trade, transportation, and waste management/remediation—are considered brown based on their emissions of a combination of the six criteria air pollutants and CO2. I construct the two scenarios by applying—separately (and then averaging the resulting predictions) in the first scenario and simultaneously in the second—a ten percent decrease to the historical GDP growth of each of the industries in this list.[42]

In the other three scenarios, I instead draw on the results in Chapter 2 to estimate the potential impacts to manufacturing industries of an oil price shock as propagated through the U.S. economic production network. These three scenarios consider the hypothetical manufacturing industry impacts if such a shock were channeled (1) by upstream (i.e., supply-side) network dynamics only, (2) by downstream (i.e., demand-side) network dynamics only, or (3) by both simultaneously. Unlike the brown industries scenarios, these scenarios allow for differential changes in GDP growth for some industries as compared to others, which reflects the fact that production network dynamics can lead to heterogeneous effects even among a group of generally polluting industries.

I use the resulting transition predictions, and the rankings constructed from them, to explore three main questions:

- How do the occupational rankings differ by scenario, and what might this tell us about the labor implications of a transition to a green economy?
- How do the rankings based on percentage change in net inflow align with, or differ from, the categorizations produced by O*NET and Vona et al. (2018)?
- What patterns emerge when considering the predictions at the level of occupations pairs, and does this provide any insight into the overall rankings themselves?

I find that there is a statistically significant correlation among the rankings across all of the scenarios. Specifically, the two scenarios based on the brown industries list produce rankings that are highly correlated with one another (coefficient of 0.97), while they are also correlated with the rankings from the oil scenarios with coefficients in the range of 0.69-0.79.

Although these tests show that the rankings are not statistically independent, inspection at the level of individual occupations reveals that there are patterns of diversity underlying these high-level correlations. For example, some occupations tend to have similar rankings across all of the scenarios,

3.4.2 for additional details.

[42]Specifically, one of the predictors that I use to train the machine learning models captures state-by-industry-level growth in GDP over each block of time in the study period. To construct the scenarios, I decrease this predictor for the relevant industries by ten percent, which represents about a one-standard-deviation change.

while others have rankings in the oil regression scenarios that differ substantially from their rankings in the brown industries scenarios.

I explore these patterns using a hierarchical clustering algorithm, which splits the 84 occupations into groups based on differences in their rankings. I find that the clustering algorithm is able to use these differences to group occupations in a sensible way, and the results suggest that about half of the occupations in the study experience moderate to large changes in their rankings when moving from some scenarios to others. The patterns in these differences further suggest that the greening of the economy may have heterogeneous impacts for occupations based not just on the particular industries affected, but also on the manner in which resulting price changes are propagated through the U.S. production network (e.g., via upstream channels versus downstream channels).

With regards to the second question above, I find mixed evidence about the relationship between the rankings based on change in net inflow and the green and brown categorizations developed by Vona et al. (2018). Specifically, I first calculate the rank correlation between the rankings for each of the scenarios and the ordering generated by the authors' greenness score. The resulting coefficients range from 0.25-0.29, which provides some evidence that greenness score is positively related to the change in net inflow. At the same, none of the correlations are quite statistically significant at standard levels. I also compare the mean percentage change in net inflow for the green occupations with that of the brown occupations. Contrary to what we might expect, I find that the green occupations are predicted to do worse under the scenarios than are the brown occupations, though the differences are not statistically significant.

I perform similar tests comparing the means of the occupations in the O*NET "green increased demand" category—which, per the O*NET definitions, are expected to be in higher employment demand as a result of green economy activities and technologies—to the means of the occupations in the other two O*NET categories. The predictions suggest increased growth for the O*NET increased demand occupations vis-à-vis the occupations in the other O*NET categories, though the differences are again not statistically significant. In addition, there is some evidence that the increased demand occupations tend to do better in the upstream-only oil scenario as compared to the other two oil scenarios. This suggests that the occupations in this category may be more sensitive than the other O*NET occupations to the propagation of shocks from downstream to upstream in the production network, at least in the context of changes to the price of oil.

Lastly, I consider patterns of worker exchange at the level of occupation pairs. As mentioned above, I model how the 84 focus occupations exchange workers with a broad spectrum of other occupations, which may include, but are not limited to, the 84 focus occupations themselves.

In analyzing the predicted pairwise occupational flows, I find that certain of these other "exchange occupations" play a central role in the rankings, in the sense that there is a statistically significant relationship between the individual transition contributions of such occupations and the overall rankings that result. This is most true in the case of the "Unemployed" occupation, which plays a significant role across all of the scenarios. This suggests that the transition to a green economy may affect particular occupations not only through the exchange of workers with other occupations, but also by patterns of how workers are sent to, or brought out of, unemployment.

I also consider patterns of worker exchange among the Vona et al. (2018) groups of green and brown occupations. In the actual (non-predicted) transition data for 2001-2012, I find that a transition into a brown or green occupation is more likely to occur for a worker already in a brown or green occupation, respectively. The predictions suggest a potential dampening of such patterns, in the sense that green-green and brown-brown transitions appear to balance out while green-to-brown transitions are predicted to further accelerate. Along with the other results, this suggests that the evolution towards a green U.S. economy may not be accompanied by a homogeneous transfer of workers from some occupations to others, and in particular, by the transition of workers from "brown" to "green" jobs. Rather, taken together, the results suggest that a given set of industrial changes may have heterogeneous effects based on the back-and-forth trading of workers between focus and exchange occupations, where such transitions may benefit some workers while detrimentally affecting others.

Overall, the central policy-related goal of the chapter, which also represents its main contribution, is to provide a different (and perhaps more nuanced) view of the notion of green jobs. This goal is motivated by the observation that there are at least two interpretations of, or uses for, the categorization of certain occupations as green. One is to identify specific subsets of the economy involved in producing what are considered to be green products or in increasing the efficiency of (and/or reducing the waste associated with) production processes generally. A second, and related, use is to identify occupations for which demand will likely grow in the future. In a similar manner, brown occupations may be considered brown because of their relationships to particularly polluting activities and/or industries, but may also be categorized as such because we expect demand for these occupations to decline over time.

As illustrated by the O*NET and Vona et al. (2018) schemes, previous studies that categorize jobs as either green or brown have often viewed occupations as related to the products that they produce, the processes that they facilitate, and/or the industries in which they are employed. In this respect, the classification methodology of these studies is closely tied to the interpretation/use of green jobs as they relate to subsets of the economy. However, the results of these classifications are sometimes interpreted (in part) through the lens of changing occupational demand, which implicitly takes the evolution of the structure of the U.S. economy as given.

The current study aims to explicitly consider these background dynamics by placing occupations within a framework that considers industries' interrelationships. In doing so, it adds to the existing literature both (1) by taking predicted occupational demand (as captured by change in net inflow) as the primary metric that differentiates occupations and (2) by considering green jobs in the context of production networks, which allows non-green occupations and/or non-green industries to be both positively and negatively affected by the greening of the economy. As described above, the results suggest that this broader perspective at least partially blurs the distinction between green and brown occupational categorizations.

In doing this, a secondary contribution of the chapter is to empirically explore these dynamics using a machine learning approach. Although some previous research has focused on occupational transitions as the primary outcome of interest, these studies have generally leveraged econometric techniques and have considered occupational relatedness based on features of occupations themselves. In this way, the current chapter deviates from previous research both in using predictive techniques to analyze hypothetical occupational transitions and in leveraging information about the broader U.S. economic production network to do so.

Related Literature. The chapter's substantive focus and methodology builds on a number of papers across several strands of literature, including green jobs, industry/occupational relatedness and transitions, and the industrial and labor impacts of environmental and energy regulation. The conceptual framework for the project follows primarily from two additional sets of studies: one that models the macroeconomy as the result of industry- or firm-level connections in an economic production network, and another that considers how labor relates to skills, tasks, and technological change. I discuss some studies in each of these areas below.

Green Jobs. As mentioned above, studies that categorize jobs as either green or brown have often viewed occupations as related to the products that they produce, the specific intra-firm or intra-industry processes that they facilitate, and/or the industries in which they are employed.

For example, in addition to the O*NET and Vona et al. (2018) categorizations previously described (and which I elaborate on in section 3.4.2), the U.S. Bureau of Labor Statistics (BLS) has classified green jobs using two methods: the output approach and the process approach.[43] The former considers the goods or services that establishments produce, and if they are considered green,

[43]See: https://www.bls.gov/green/.

classifies the associated occupations as green. The latter approach instead looks at the processes of establishments, and if the processes are deemed to be environmentally friendly, takes the occupations associated with the greening processes to be green. Given this, the output approach is relevant only for some industries, while the process approach is relevant for all industries.

As described above, the current study differs from these types of systems in that it takes the predicted change in occupational-level net inflow as the primary measure of greenness, rather than characteristics of occupations or of industries themselves.

Beyond classification schemes, another set of research has drawn on the O*NET database, or on methodologies similar to those used by the U.S. BLS, to consider the topic of green jobs within a number of specific contexts.

For instance, Shutters, Muneepeerakul, and Lobo (2016) consider the difficulty with which urban economies can become green. To do so, they draw on a metric developed by Muneepeerakul et al. (2013) that measures how an urban area's specialization in one occupation either promotes or hinders its specialization in another.[44] By calculating this metric for every pair of occupations, Muneepeerakul et al. (2013) create a network—termed the "occupation space"—in which occupations form the nodes and the pairwise metric values form the edges. Shutters, Muneepeerakul, and Lobo (2016) use this occupation space to calculate a "Green Jobs Index" for each U.S. metropolitan area, which in a sense averages the network distance between an area's current occupation mix and a static green occupation mix based on the classifications in O*NET. The authors find substantial variation in this index across areas. They also note that movement towards a greener occupation mix is slow, and that movement away is nearly as likely during the years in their study period.

Bowen, Kuralbayeva, and Tipoe (2018) also use the O*NET database to examine green employment in the United States. They categorize occupations into "green" and "non-green," and then into five sub-categories. Using data from the U.S. BLS, they find that nearly 20 percent of workers are in occupations that fall into the green category, while another 44 percent work in occupations that could serve as transition points into green employment. The authors also investigate the similarity among occupations across the five sub-categories using several different metrics.

Finally, Burger et al. (2019) analyze the skills and education required in occupations related to the "circular economy" (an economy that focuses on minimizing waste and reducing negative production impacts). They draw on work by Van Oort et al. (2018) to categorize industries as "core,"

[44] I note that this research has some overlap with the literature on agglomeration economies, which considers the benefits that stem from the co-location of people and firms within geographic areas (see, for example, Glaeser (2010)). I do not discuss this literature explicitly here as the current study is more directly related to the notion of green jobs and occupational transitions broadly and not to their dynamics within particular geographic settings.

"enabling," or neither core nor enabling (these loosely correspond to green, green-supporting, and "other" categorizations in some of the previous literature). The authors then analyze the educational and skill requirements of occupations associated with circular economy industries, finding that the requirements are very diverse and not substantially different from the rest of the economy.

Altogether, and analogous to the classification schemes discussed above, these studies focus on the greenness of urban areas and/or of particular occupations as measured by features of occupations or industries themselves. Although some—such as Bowen, Kuralbayeva, and Tipoe (2018)—consider the transition possibilities among different occupations, these possibilities are predicated on similarities among occupations rather than on predictions of how worker transitions will change in response to the evolution of the economy. In this way, the current study provides a complement to these approaches in explicitly considering these latter dynamics.

Industry/Occupational Relatedness and Transitions. Another set of papers has considered the relationships among occupations more generally, as well as flows of workers among occupations and/or industries. I briefly describe several studies here.

Allen et al. (2012) develop and implement an algorithm to calculate the relatedness among 858 occupations in the O*NET database. The resulting product, termed the "career changers matrix," identifies for each occupation up to ten possible transition occupations that workers could immediately pursue. To create this matrix, the authors draw on information in O*NET, as well as their own categorization work, to assign five vectors to each occupation: knowledge, skills, work activities, work context, and job zone (education/work experience). To determine the relatedness between each pair of occupations, they first calculate the Euclidian distance between occupations' knowledge vectors, and then convert the distances into z-scores to standardize the values across all pairs of occupations. They repeat this process for each of the other categories, and then take a weighted average of the results. For each occupation, the top ten closest matches are those listed in the career changers matrix. The authors evaluate the matches both quantitatively and qualitatively.

Mealy, del Rio-Chanona, and Farmer (2018) use U.S. Current Population Survey (CPS) data to find the probability of transitions from one occupation to another. Based on work activities from O*NET, they devise a measure of the similarity between each pair of occupations, finding that the correlation between this measure and the transition frequency is 0.3 and highly statistically significant. Drawing on the methodology of Hidalgo et al. (2007), the authors also create two conceptual networks. The first connects a set of occupations based on the similarity of their associated work activities, while the second connects a set of work activities based on their likelihood to occur within the same job.

Nedelkoska, Diodato, and Neffke (2018) use CPS and O*NET data to construct a dataset of job transitions at the annual level among 368 occupations. They then build three conceptual networks that relate these occupations, where the links in each network are based on the similarity of the tasks, tools, or knowledge/skills/abilities associated with the occupations. They regress the flow of workers between each pair of occupations on these similarity measures, finding that they are predictive of occupational switching. Using Frey and Osborne's (2017) measure of job automatability, they also evaluate specific occupations in terms of their automation risk, transition potential, and earnings.

Neffke, Otto, and Weyh (2017) use German social security data for the years 1999-2008 to consider workers' industry transitions from one year to the next. They find that workers frequently cross industry boundaries (even at the most aggregated level) and that most worker transitions occur among a small number of industry pairs.

Lastly, Parrado, Caner, and Wolff (2007) use a worker-level dataset for the United States during 1968-1993 to examine labor mobility among an aggregated set of eight occupational groups and 11 industries. They find that occupational and industrial changes occurred more frequently in the later part of their study period, suggesting that macro-level economic turbulence in the 1980s was mirrored at the micro level.[45]

Altogether, the current chapter is directly related to this previous work in that it focuses on occupational transitions as the primary outcome of interest. However, as in the case of the green jobs literature above, the main difference is that many of these previous studies take the relatedness of occupations to be based on some combination of the knowledge, skills, work activities, and/or other characteristics associated with, and shared among, different occupation pairs and/or groups of occupations. In the current study, I instead take occupational transitions to be primarily driven by industry-level changes, though I implicitly allow for pair-level differences in the relatedness of occupations in the sense that I train two separate models for every occupation pair (one in each direction; see below).

Impacts of Environmental and Energy Regulation. This study is also related to a large literature that has considered the industrial and employment effects of environmental and energy regulation, which will likely play a driving role in the movement towards a green economy. I briefly describe several papers in this literature here.

Morgenstern, Pizer, and Shih (2002) use plant- and industry-level data to estimate the employment effect of increased environmental spending in four industries: pulp and paper mills, plastic

[45] Using a regression approach, these authors also find that occupational and industrial transitions were associated with lower earnings for men and with either higher or lower earnings for women depending on the sample and the regression.

manufacturing, petroleum refining, and iron and steel mills. On average across the industries, the authors find the employment effects to be economically and statistically insignificant. Considering each industry in isolation, they find no effect for pulp and paper mills and steel mills, though they do identify a positive effect for plastics and petroleum, which they attribute to the labor intensity of these industries' environmental activities and relatively inelastic demand for their products.

Using data on pairs of adjacent U.S. counties, Kahn and Mansur (2013) find that counties with higher electricity prices have lower levels of employment in electricity-intensive manufacturing industries. In addition, this effect increases as industry electricity intensity increases, with the largest impacts—out of the approximately 20, 3-digit-NAICS manufacturing industries that the authors consider—to primary metal manufacturing, paper manufacturing, and textile mills. They conclude that energy prices are a significant factor in determining geographic location for a handful of manufacturing industries but that such effects are modest for the typical manufacturing industry.

Aldy and Pizer (2014) also consider the employment (as well as the competitiveness) effects of electricity prices for manufacturing industries. Using detailed data for the period 1986-1994, they find that—for industries with energy intensity above 3-4 percent—an increase in electricity prices of five percent corresponds to decreases in employment of about 1-1.5 percent. The authors leverage these results to simulate the potential effect of a \$15 per ton CO2 price in the power sector, which they find leads to gross employment impacts of about 0.2 percent for manufacturing broadly and in the range of 1-2 percent for energy-intensive industries.

Walker (2013) uses data at the worker-firm level to estimate the earnings and employment effects of environmental regulation under the 1990 Clean Air Act Amendments. The author finds negative impacts to both earnings and employment for workers in newly regulated plants, and importantly, discovers that these costs are driven by sectoral downsizing and associated job transitions. Workers who remain at their original firms do not experience wage losses.

Curtis (2018) estimates the employment effects of the NOx Budget Trading Program, which was in effect from 2003 to 2008 and affected certain electricity generators and industrial plants. Using two employment datasets, the author finds decreases in employment for the most energy-intensive industries, caused primarily by a decrease in hiring rates.

Lastly, Kuminoff, Schoellman, and Timmins (2015) develop a model that considers the welfare effects associated with job losses and potential relocation due to environmental regulation. Applying the model to data on working households in Northern California, the authors find that sudden job loss for the average worker would result in a decline of about \$5,500 in annual earnings, approximately 70 percent of which is due to loss of job-specific human capital and 30 percent of which is due to forgone wages during unemployment. The authors also find that this earnings loss accounts for only three-quarters of a worker's overall welfare loss, the remainder of which is due to relocation (to a less desirable community) and/or a longer commute.

Overall, the current chapter is related to this research in that it aims to understand how changes in the economy—potentially induced by regulation—may affect employment outcomes. One of the themes across this research is that job changes and/or transitions into/out of unemployment may be important channels by which these policies have effects. In this chapter, I consider such dynamics by disaggregating the analysis to the occupation-pair level, which allows for heterogeneity in impacts across the focus occupations as well as the consideration of occupation-to-occupation transitions separately from transitions into and out of unemployment. This said, I do not take account (as some of the above studies do) of potential earnings impacts, which represents an opportunity for future research using the type of approach proposed here.

Economic Production Networks. The conceptual framework for the study (see section 3.2) and three of the scenarios draw on a literature that conceptualizes the economy as a network of firms or industries. These papers generally fall into two broad categories: those that are primarily theoretical and those that take theory to data.

Theoretical papers include early work by Long and Plosser (1983) as well as more recent studies by Gabaix (2011), Acemoglu et al. (2012), Acemoglu, Ozdaglar, and Tahbaz-Salehi (2017), Baqaee (2018), and Baqaee and Farhi (2018), among others. A common theme across these papers is an examination of how industry or firm interconnections can lead micro-level shocks to influence aggregate outcomes.

Empirical papers in this literature include Foerster, Sarte, and Watson (2011), Acemoglu, Akcigit, and Kerr (2016), Barrot and Sauvagnat (2016), Carvalho et al. (2021), Atalay (2017), and Di Giovanni, Levchenko, and Mejean (2018). These papers use firm- or industry-level datasets (1) to examine the impacts of various types of shocks (such as industry-level versus common shocks) as well as (2) to explore how supply and demand shocks propagate through production networks.

The current study draws on the concept of production networks to suggest how specific industries may be differentially affected by environmental or energy regulations that increase the price of petroleum products. Specifically, three of the scenarios I consider are based on the idea that a direct shock to the petroleum refineries industry will propagate through the production network via industry-to-industry linkages, which will result in heterogeneous impacts to other industries throughout the economy. In the second part of the study, I translate these impacts into predicted occupational movements via the machine learning models that I train.

Labor: Skills, Tasks, and Technological Change. Lastly, the conceptual framework in the current study is also broadly related to previous work that considers labor as it relates to skills, tasks, and technological change. I briefly review some studies below.

Lazear (2009) develops a model of human capital with two primary parts: (1) workers invest in a variety of general skills, each of which may be relevant at multiple firms; and (2) each firm weights the importance of specific skills differently. As with earlier models, the author's model predicts that workers who experience involuntary job changes will also experience lower earnings at their new positions. Unlike the traditional view, this loss is explained not by firm-specific human capital (with limited transferability to other firms) but by differences in how firms value various general skills and the resulting skill investment choices of workers.

Acemoglu and Autor (2011) develop a model that first categorizes workers as low-skilled, medium-skilled, or high-skilled, and then endogenously allocates worker skills to work tasks. Sets of work tasks, in turn, are used to produce unique final goods. Technological change can affect the productivity of different types of workers across all tasks and for specific tasks, and new technologies can replace workers to complete particular tasks. One of the implications of the authors' model is that factor-augmenting technical progress can increase the wages of some worker groups while decreasing the wages of others.

Timmermans and Boschma (2014) investigate the impacts of heterogeneous skill inflows on plant productivity in Denmark. Drawing on the results of Neffke and Henning (2013), they first determine the skill-relatedness of 7,750 pairs of industries based on worker flows between these industries in Sweden. The authors then classify the Danish labor inflows to each plant as similar, related, or unrelated to the plant's industry, based on the industry(ies) in which the incoming individuals previously worked. Using an econometric framework, they find that inflows of workers with related skills had a positive impact on plant-level labor productivity, while inflows of workers with similar skills had a negative impact. The authors conclude that it is not labor mobility or the inflow of skilled labor that produces positive impacts, but instead the particular type of skills associated with the inflows that makes a difference.

Finally, Nedelkoska, Diodato, and Neffke (2018) explore the transferability of knowledge, skills, and abilities (KSA) in the context of technological change. Drawing on the models of Acemoglu and Autor (2011) and Autor and Handel (2013), they conceptualize KSA as being applied to job tasks, which in turn produce goods and services. KSA may also be applied to the use of particular tools and technologies, which are used, in turn, to complete job tasks (which lead to the production of goods and services). One of the authors' key conclusions is that workers know more than what they do; in other words, the KSA that a worker has acquired is broader than what is needed for a particular set of tasks in any given job.

Altogether, the ideas in these studies underpin the conceptual basis for this chapter, in the sense that the transition of a worker from one occupation to another involves the re-application of certain knowledge or skills from the first occupational context to the second (and may also require the application of knowledge/skills that the worker already possessed but had not been previously using). I explicitly include these types of dynamics in the model that I describe in section 3.2. As mentioned above, I implicitly recognize the distinct relationship between each pair of occupations by training separate models for every pair. An alternative approach—which I leave for future work—is to train a single model across multiple occupation pairs, which could leverage predictors about occupational relatedness in addition to the economic factors I consider in this study.

The chapter proceeds as follows: in section 3.2, I present a model that relates changes in the economic production network to occupational transitions; in section 3.3, I describe the empirical approach, including the construction of the predictors and how I train the machine learning models; in section 3.4, I discuss the results of predicting occupational flows under the scenarios of economic change; and in section 3.5, I provide some concluding thoughts.

3.2 Model

As in Chapters 1 and 2, the economy is conceptualized as a production network of a set of industries I, where each industry uses the outputs of other industries to produce its own output. The process by which an industry $i \in I$ combines given inputs into an output is itself a network of production steps. Each production step is facilitated by some amount of human capital and physical capital. The former is embodied in the workers that perform the production step, while the latter includes any machines or, more broadly, "tools" used to complete the step. Unless a production step is entirely automated, we might use the term "task" to describe the step, which reflects the human part of the process.

Abstracting from an industry i's internal production steps, we can summarize its overall production process with the function $f_i(\cdot)$, which takes two sets of inputs: (1) the outputs of other industries (which may be tools, technologies, materials, and/or information) and (2) labor (sets of tasks). I assume that each industry produces its output in a lowest-cost manner to satisfy the intermediate demand of other industries as well as the final demand of consumers.

Take the set of occupations to be O, the set of tools/technologies to be T, and the set of work tasks/activities to be A. An occupation $o \in O$ is defined by two subsets: the tools and technologies it uses ($TT_o \subset T$) and the work activities it performs ($WA_o \subset A$). In turn, an industry i's production process, $f_i(\cdot)$, gives rise to a required set of occupations based on the tasks, tools, and technologies with which it is associated. In particular, an industry hires workers in each occupation to cover the needs of its production steps: the activities to be completed and the ability to use the tools and technologies instrumental to each task.

In this framework, the economy changes in multiple ways, including: shifts in consumers' final demand for products; changes in the quantities of inputs that industries source from other industries; and intra-industry changes in production processes. Such changes propagate throughout the economy, generating increased need for human capital in some locations of the production network while reducing the need in others. Specifically, as an industry i adjusts (1) its output to satisfy fluctuating demand, (2) its inputs to accommodate varying input prices, and/or (3) its internal production steps in response to the introduction of new technologies or processes, it will increase or decrease employment in each of the occupations with which it is associated. These dynamics, in turn, give rise to a network of worker redistributions. For the purposes of the current chapter, I abstract from industries' input use and internal processes to focus on changes in the structure of the production network as captured by industry output at the U.S. state level, which for industry i in state s I denote as Y_{is}.

Although each industry has its own unique production process, commonalities in the steps within those processes results in the employment of each occupation across a range of industries. Use I_o to denote the set of industries in which occupation o is employed. A worker in occupation o in industry i who loses or voluntarily leaves his/her job may pursue work in another industry in I_o, may transition to new work in occupation $p \neq o$ (which may or may not involve a change in industry), or may become unemployed. Similarly, an industry j hiring workers in occupation p may find those workers in another industry within I_p, may hire workers transitioning out of another occupation $q \neq p$, or may bring workers out of unemployment.

The transition of a worker from occupation o to occupation p is facilitated or hindered by a number of factors. Specifically, it is possible (and likely) that a worker in occupation o will not necessarily only be trained in the use of the tools and technologies in TT_o, nor only have the capability to perform the activities in WA_o. In this way, each worker generally knows more, and can do more, than he/she actually does in a particular job. In the absence of additional schooling, training, or experience, this gives rise to the possibility of a worker transitioning from occupation o to occupation p because the occupations o and p are similar or related based on TT_o, WA_o, TT_p, and

WA_p. With additional schooling, training, or experience, we imagine that the same worker would be able to transition from occupation o to occupation q, where occupations o and q are somehow less similar, or less related, than are occupations o and p. For the purposes of the model, I group similar workers—in terms of their education, experience, etc.—together, which yields a set of worker categories $w \in W$.

To capture the "transition flow count" for workers between occupations, define for origin occupation o, destination occupation p, origin industry i, destination industry j, state s, and worker category w, the following function:

$$T(o, p, i, j, s, w \mid I_o, I_p, TT_o, TT_p, WA_o, WA_p, Y_{is}, Y_{js}, X_w) \tag{3}$$

where o may or may not equal p, i may or may not equal j, X_w is a vector of factors (such as education) specific to worker group w, and the remainder of the variables are defined as above. This function represents the expected quantity of worker transitions—over a given period of time in state s for worker group w—from occupation o in industry i to occupation p in industry j. This value is dependent on all of the factors previously defined, including which occupations are employed in which industries, the relatedness of occupations, state-by-industry output, and the details of worker groups. I describe in the next section the simplifications I make to this structure for the purposes of the current chapter.

Taken all together, the model states the following: changes in the demand for industries' outputs lead industries to increase or decrease employment in each of the occupations with which they are associated, based on industries' production processes and the manner in which industries are connected in the production network. In turn, these dynamics lead to flows of workers among industries and occupations (and into and out of unemployment), based on the characteristics of different types of workers as well as on the relationships among industries/occupations themselves.

The components of this model can be visualized as part of one large network that combines industries, occupations, tasks, and tools together. In this network, industries are connected to: (1) the industries they source their inputs from; (2) the industries they sell their outputs to; and (3) the occupations in which they employ workers. As above, occupations are taken to be collections of work tasks, which require workers to use specific tools and technologies. In this way, occupations in this network are directly connected to—in addition to the industries that employ them—both the work tasks they perform and the tools and technologies they use.[46] The model posits that, over a given

[46] In a more detailed version of this model, as alluded to in the beginning of this section, industries could be disaggregated into their constituent production steps, which themselves form a network. In this case, tasks, tools,

period of time, the structure and evolution of these network relationships induces two connections between every pair of occupations: a transition count in each direction.[47] The aim of the machine learning models in the current study is to capture a reduced-form version of these dynamics.

I note that this framework is related to the conceptual model discussed in Nedelkoska et al. (2018) (which draws on the relationships proposed in Acemoglu and Autor (2011) and Autor and Handel (2013)). As described in the literature section above, Nedelkoska, Diodato, and Neffke (2018) take knowledge, skills, and abilities (KSA) as being applied to job tasks, which in turn produce output. KSA also allows workers to utilize tools and technologies, which are used to complete job tasks, which in turn produce output. Similar to this model, I consider work tasks as the necessary human capital associated with each production step, which may also require the application of particular tools or technologies. Unlike this model, I abstract from the KSA that underlies tasks, tools, and technologies, while also embedding the production process of each industry within the larger production network.

3.3 Empirics

As described in the introduction, the chapter's empirical approach is broken into two parts. In the first part, I take the evolution of industries as given and train a series of machine learning models that relate U.S. economic structure to historical occupation-to-occupation flows. In the second part of the study, I use the models to construct predicted flows between every pair of occupations under several scenarios of U.S. economic change. I then leverage the predictions to investigate a number of questions related to green jobs.

I use a machine learning approach because it allows for the flexible combination of variables in predicting occupational flows. Specifically, as I describe in detail in section 3.3.2 below, I construct two sets of predictor variables that vary over time based on regional GDP data: one that captures information at the state-industry level, and a second that reflects information at the state-occupation level. Some of these predictors—which I term "industry score," "occupation score," and "relative occupation score"—are designed to reflect relevant industry and occupation information at the state level. However, I also include a full set of industry-level GDP values (as averages, standard deviations, and changes; see below) as predictors, which broadly represent the status of a state's economic production network over a given period of time.

and technologies could also be associated with the production steps in which they are used. Similarly, occupations could be associated with specific production steps within an industry rather than with each industry as a whole.

[47]In this network, each occupation can be thought of as a single entity regardless of the industry in which it is employed, or as a different entity for each of the industries in which it is employed. As I describe in the next section, I take the former approach for the purposes of this chapter.

In this way, although I provide the machine learning algorithm with my own set of indicators, the algorithm is also able to—in a sense—create its own measures of what drives occupational transitions based on this broad snapshot of the economy. Given that I train a separate model for the transition from each focus occupation to every other occupation (and another model in the reverse direction), the algorithm is able to assemble these measures to capture the idiosyncratic relationships that arise between every occupation pair.

As mentioned in the introduction, previous research on occupational transitions has often taken the pairwise relatedness of occupations—as reflected, for example, by indicators that capture the overlap in required knowledge, skills, and/or abilities—as the driving force behind occupational transitions. Here, by training pair-level models given a broad set of GDP-based information, I implicitly acknowledge the heterogeneity among occupation pairs while also allowing for flexibility in which economic factors may be most relevant to occupational change.

Altogether, the predictions made by the models—given changes to state-level industrial activity—represent the occupational flows we might expect to see if such industrial evolution actually took place. The model described in the previous section holds that these employment changes are the direct result of such industrial fluctuations, in the sense that workers are hired to complete production steps within a production process, and each production process is executed to fulfill the sum of intermediate and final demand. Although this interpretation underlies my discussion of the results in section 3.4, the predictions—as well as my general conclusions—are still relevant even if the models are taken to reflect a correlative rather than a causal relationship.

Finally, as mentioned in the introduction, I test the models by examining their predictive power for a hold-out sample covering the years 2013-2016. Specifically, I compare the models' predictions against a baseline "no-information" guess, which is the prediction that there will be zero flow between an occupation pair in a given time period (as zero is the most common flow count in the dataset). I perform this comparison as a general check of the models' accuracy and find that they perform relatively well (see section 3.4.1 for more details).

In computing the predictions, I also apply a threshold where the change in flow between two occupations is only considered to be non-zero if it exceeds two measures of error associated with the model for that occupation pair. By doing so, I explicitly take account of the error inherent in the models' predictions while also recognizing that this error will differ across models (see section 3.4.2 for additional details).

3.3.1 Overview of Approach

I provide additional detail about the two parts of the analysis here before turning to an overview of the data and the construction of the predictors.

Training the Models

To implement the approach described above, I train two separate models for each pair of occupations o and p: one that predicts the flows from o to p, and one that predicts the flows from p to o. To do so, I leverage two types of variation: geographic and temporal. For the first, I use the U.S. state associated with each transition in the original data. For the second, I divide the sample period 2001-2016 into four "year groups": (1) 2001-2004, (2) 2005-2008, (3) 2009-2012, and (4) 2013-2016. I then aggregate the transition counts to create a total for each combination of occupation pair, state, and year group, which forms the variable that the models are trained to predict.

The predictors themselves are based on 3-digit-NAICS industry GDP data, which vary both by state and by year group (see below for details). In this way, I use variation in state-industry information over time to predict occupation-to-occupation transitions at the state level over time. I train the models using data from the first three year groups (2001-2012) and then assess the accuracy of the predictions using the fourth (2013-2016).

I note that by training and evaluating the models in this way, I assume there is no dependency between the year groups, in the sense that occupational transitions in one year group are taken to be independent of the transitions that occur in the following group(s). A valuable avenue for future research would be to incorporate such longitudinal dynamics explicitly into the models for each occupation pair.

I also make several simplifications to the model outlined in the previous section. First, I assume that the relatedness among occupations (as captured by TT_o, TT_p, WA_o, and WA_p in the model) is constant over the study period, and in turn, I do not include any predictors to capture these relationships. However, by training a separate model for each occupation pair, I do allow each pair o-p to have a specific dynamic that differs from every other pair. In future research, an alternative approach would be to train models for more than one pair at a time, which would likely require the inclusion of additional predictors that capture the relatedness of each pair.

Second, I focus on worker transitions without regard to the encompassing industries. In other words, in constructing the aggregate number of worker transitions from occupation o to occupation p in a state and year group, I include all such transitions whether an individual crosses an industry boundary or not. In the language of the model above, this amounts to summing $T(\cdot)$ (equation 3)—for each group of occupational transitions o-to-p—over all origin industries i and destination industries j. This said, a subset of the predictors (as discussed in more detail below) do account for the fact that occupations are employed at varying rates across industries.

Lastly, I also abstract from characteristics of workers themselves. As in the case of industries, this means that in the construction of the aggregate number of worker transitions from occupation o to occupation p in a state and year group, I include all such transitions regardless of any information I may have about individuals, such as education. In the language of the model, this again amounts to summing $T(\cdot)$ (equation 3)—for each group of occupational transitions o-to-p—over all worker groups w.

My motivation for abstracting from industries and from worker characteristics is to limit the number of dimensions I include in the analysis. Even at the level of aggregation I use, there are many occupation pairs with zero flow, and many occupation pairs that have non-zero flow in some but not all time periods. In turn, by aggregating across industries, demographics, and other factors, I facilitate the training of the machine learning models by providing a dataset with meaningful variation in occupational flows across occupation pairs, states, and year groups.

In addition, given that I train models at the level of occupation pairs, I implement a threshold to determine which pairs will be included in the analysis. Said another way, this threshold determines, for each occupation o, the possible occupations $p \neq o$ from which and to which occupation o will receive and send workers, respectively. To implement this, I limit the sample by including only those pairs with at least two transitions per year, on average, across the training period (2001-2012).

Specifically, I include an occupation pair o-p—and models are trained in both directions—if this threshold is met in at least one direction (e.g., if the flow count from o to p exceeds the threshold but the flow count from p to o does not, then the pair is retained and models for o-to-p and p-to-o are both trained). This retains nearly 97 percent of the original flows in the data but eliminates many occupation pairs that have a small number of transitions and/or that exhibit little variation. Note that I treat unemployment as an occupation, and in turn, I train models o-to-unemployment and unemployment-to-o if the associated flow counts meet the threshold.

Overall, with respect to the model above, these simplifications essentially (1) eliminate some dimensions associated with transition counts and (2) take some factors to be constant, yielding an aggregated $T(\cdot)$ function dependent on o, p, s, and industry-level output information for each period. In this way, the trained machine learning models (taken together across occupation pairs) can be thought of as an analog to this simplified $T(\cdot)$ function.

Predicting Occupational Transitions

To operationalize the second part of the study, I consider multiple methods for classifying occupations as green and brown that roughly correspond to viewing occupations at different levels of connectedness within the industry-occupation-task-tools network described at the end of the previous section. In particular, I consider occupational categorizations based on how occupations are:

- Related to their specific tasks and tools;
- Connected to the industries in which they are employed; and
- Related to other occupations themselves, in the sense that changes to the economy induce transitions of workers from some occupations to others.

To the extent that an occupation's tasks and tools remain constant, the first method yields a static classification of which occupations are most green and which are most brown. If the tasks performed by occupations—or the tools used to perform those tasks—change over time, then the classifications under this method would also shift (though such task/tool changes would, in a sense, also represent at least partial re-definitions of what each occupation is).

Similarly, to the extent that the set of industries that employ each occupation—and our assessment of which industries are green or brown—remains fixed, then the second method also produces static classifications of which occupations are green and which are brown. If industries' production processes change in such a way that they hire different sets of occupations over time (or perhaps different proportions of the occupations that they already employ), and/or if such changes altered the assessment of these industries as being either green or brown, then the occupational classifications under this method would also change accordingly.

The third method, by virtue of considering worker transitions, necessarily follows from some change in the industry-occupation-task-tools network (as a static network implies, in the framework used in this study, constant levels of employment in each industry). In this way, effects for particular occupations are evaluated in the context of specific network transformations.

As described in the introduction, I consider two sets of green jobs based primarily on the first two of these methods and one set of brown jobs based on the second method. Specifically, these are the green occupations identified within O*NET (based on the work of Dierdorff et al., 2009, 2011) and the green and brown occupations identified by Vona et al. (2018). Taken together, these amount to 277 unique green and brown occupations, 84 of which form the set of focus occupations for this study (see section 3.4.2 below).

I compare the O*NET and Vona et al. (2018) categorizations with the outcomes of the third approach, which I implement by leveraging the predictions of the trained models under various scenarios of economic change. I describe the construction of the scenarios, the resulting predictions, and the O*NET and Vona et al. (2018) categorizations in more detail in the context of the results.

3.3.2 Data Sources and Construction of Predictors

I use two primary data sources for creating the predictors and training the models:

- The U.S. Current Population Survey (CPS; as published through IPUMS[48]), which provides monthly information on workers' geographic locations, industries, employment status, and occupations; and

- The U.S. Bureau of Economic Analysis' (BEA) state-by-industry GDP data, which is aggregated at the 3-digit-NAICS level.

The CPS is structured such that a household is interviewed for four consecutive months, is left out of the sample for eight months, and then is interviewed again for four consecutive months. Given that I aggregate the transition counts to blocks of time that span four years, I include occupational transitions regardless of whether they occur during the first interview period, during the second interview period, or during the time span between the two periods. If an occupational change appears between two year groups, I count it as part of the total for the first group.

I use two key pieces of information within the CPS dataset to determine if an occupational transition occurred. The first is an individual's employment status within a month, such as whether the person was employed, unemployed, or not in the labor force. The second is an individual's occupation. For the combination of each individual and each month, I determine whether or not the person experienced an occupational transition in three steps: (1) I classify the person as employed if their employment status is "at work" or "has job, not at work last week" and unemployed if their status is "unemployed, experienced worker" or "unemployed, new worker"; (2) I compare this classification to the analogous classification for the next month in the sample; and (3) if the classification changes from employed to unemployed, unemployed to employed, or remains employed but is associated with an occupational change, then I mark the transition to be included in the total for the relevant occupation pair (where unemployment is itself treated as an occupation). This approach ignores transitions of workers into and out of the labor force, which I do to focus specifically

[48] IPUMS-CPS, University of Minnesota, www.ipums.org.

on those individuals who are working or looking for work and who are likely to be most affected by changes in the production network.

Finally, once I have a list of these transitions, I am able to construct the dependent variable given the aggregations described in the previous subsection. I note that one drawback of the CPS is that respondents only remain in the survey if they stay at their current address. An opportunity for future research would be to apply the types of methods outlined in this chapter to a dataset that includes such geographic movements.

I construct two sets of predictors using the U.S. BEA GDP data, which I describe in detail below.

State-Industry Predictors

To create these predictors, I first construct a state-by-industry "industry score" for each individual year in the data (ignoring the year groups for now), which represents the relative size of each industry in a state as compared to the average size of that industry across all states. Specifically, for each combination of state s, industry i, and year y, I calculate the ratio of the GDP in that state-industry-year against the mean GDP for that industry across all states in that year:

$$\text{ind\_score}_{siy} = GDP_{siy} \Big/ \left[\left(\sum_{a \in S} GDP_{aiy} \right) / |S| \right]$$

where S is the set of states and $|S|$ is the number of states. With this completed, I have industry scores—alongside the original GDP values—for each combination of state, industry, and year.

I then consider the patterns in GDP and industry scores within each year group to construct three predictors based on GDP and three predictors based on industry score. Specifically, for each combination of state s, industry i, and year group g, these are: (1) GDP average (gdp_avg_{sig}) and industry score average ($\text{ind\_score\_avg}_{sig}$) of the four observations within each state-industry-year group; (2) GDP standard deviation (gdp_sd_{sig}) and industry score standard deviation ($\text{ind\_score\_sd}_{sig}$) of the four observations within each state-industry-year group; and (3) GDP change over time (gdp_change_{sig}) and industry score change over time ($\text{ind\_score\_change}_{sig}$) across the four observations within each state-industry-year group.

The last two of these are based on a simple linear regression of the GDP or industry score values on the year and a constant. For GDP, the specification is $GDP = a + b \cdot year$, where $year \in [2001, 2004]$ for year group one, $year \in [2005, 2008]$ for year group two, etc. The predictor is then calculated as $\text{gdp\_change}_{sig} = b/\text{gdp\_avg}_{sig}$. For industry score, the specification is $\text{ind\_score} = c + d \cdot year$, where $year$ again takes the values 2001-2004 for year group one, 2005-2008 for year group two, and so on.

The predictor is calculated as $\text{ind\_score\_change}_{sig} = d/\text{ind\_score\_avg}_{sig}$.

State-Occupation Predictors

I also construct state-level scores for each occupation that intend to reflect a state's specialization in an occupation at a given time. To do this, I first load the full CPS dataset and keep only the occupation and industry variables (discarding any observations where the occupation or industry is not specified). For each occupation, I then calculate the percentage of observations for that occupation where the respondent is employed in each industry. Call these percentages occ_ind_{oi}, where o is an occupation, i is an industry, and $\sum_{i \in I} \text{occ\_ind}_{oi} = 1$. This provides an estimate of the relative frequency of an occupation's employment across industries (and serves as a rough analog to the variable I_o in the model above).

Then, for each state s, industry i, and year y (again ignoring year groups for now), I calculate the fraction of the state's GDP in that industry in that year out of its total GDP in that year:

$$\text{GDP\_pct}_{siy} = GDP_{siy} \Bigg/ \left(\sum_{a \in I} GDP_{say} \right)$$

where I is the set of industries. Finally, I construct an occupation o's score in state s in year y as a weighted average of the state-by-industry GDP percentage values, where the weights are the occupation-industry percentages as calculated from the CPS data:

$$\text{occ\_score}_{osy} = \sum_{i \in I} \text{occ\_ind}_{oi} \cdot \text{GDP\_pct}_{siy}$$

where I is again the set of industries. The result is a value between zero and one (inclusive) that indicates the relative specialization of a state in a particular occupation in a particular year. For instance, if a state's entire GDP in a certain year comes from industry a, and occupation o is only employed in industry a, then the occupational score for occupation o in that state in that year is one. Conversely, if a state does not produce any output in the industry(ies) associated with a particular occupation, then the occupational score for that occupation in that state will be zero.

This occupational score is based on the relative industry GDP values within each state. To capture how an occupation-state's score compares to the scores in other states, I also compute the ratio of each occupation-state score against the average of the scores across all states, which I term the "relative occupation score":

$$\text{rel\_occ\_score}_{osy} = \text{occ\_score}_{osy} \bigg/ \left[(\sum_{a \in S} \text{occ\_score}_{oay})/|S| \right]$$

where S is the set of states and $|S|$ is the number of states. With this, I have occupation scores and relative occupation scores for each combination of state, occupation, and year.

Then, as with the state-industry predictors described above, I calculate the average, standard deviation, and four-year change for each combination of occupation o, state s, and year group g. Specifically, these are: (1) occupation score average ($\text{occ\_score\_avg}_{osg}$) and relative occupation score average ($\text{rel\_occ\_score\_avg}_{osg}$); (2) occupation score standard deviation ($\text{occ\_score\_sd}_{osg}$) and relative occupation score standard deviation ($\text{rel\_occ\_score\_sd}_{osg}$); and (3) occupation score change over time ($\text{occ\_score\_change}_{osg}$) and relative occupation score change over time ($\text{rel\_occ\_score\_change}_{osg}$).

As before, the last two of these are constructed from a linear regression of the occupation score and relative occupation score values on the year and a constant. For occupation score, the specification is $\text{occ\_score} = a + b \cdot year$ and the predictor is calculated as $\text{occ\_score\_change}_{osg} = b/\text{occ\_score\_avg}_{osg}$. For relative occupation score, the specification is $\text{rel\_occ\_score} = c + d \cdot year$ and the predictor is calculated as $\text{rel\_occ\_score\_change}_{osg} = d/\text{rel\_occ\_score\_avg}_{osg}$.

3.3.3 Predictor Summary Statistics

I present summary statistics for the predictors below (Tables 3.1 and 3.2). I note that the change variables (gdp_change, ind_score_change, occ_score_change, and rel_occ_score_change) are normally distributed with means near zero, while the remainder of the variables are highly right-skewed.

Table 3.1: Summary Statistics for State-Industry Predictors.

Variable	Mean	Std. Dev.	Min.	Max.
gdp_avg	4,233.791	11,948.070	0.000	403,272.900
gdp_sd	414.833	1,241.361	0.000	42,856.250
gdp_change	0.057	0.113	-1.200	1.200
ind_score_avg	1.000	1.691	0.000	29.030
ind_score_sd	0.065	0.154	0.000	3.728
ind_score_change	0.001	0.102	-1.200	1.200

Note: the minimum and maximum values of -1.2 and 1.2, respectively, for gdp_change and ind_score_change are the result of year groups that have a single non-zero value at either the beginning or the end of the four-year period. The mean of ind_score_avg is one by construction (see section 3.3.2 for additional details).

Table 3.2: Summary Statistics for State-Occupation Predictors.

Variable	Mean	Std. Dev.	Min.	Max.
occ_score_avg	0.045	0.034	0.000	0.410
occ_score_sd	0.001	0.002	0.000	0.043
occ_score_change	0.000	0.032	-0.382	0.447
rel_occ_score_avg	1.000	0.365	0.003	10.868
rel_occ_score_sd	0.028	0.042	0.000	1.694
rel_occ_score_change	0.000	0.027	-0.324	0.390

Note: the mean of rel_occ_score_avg is one by construction (see section 3.3.2 for additional details).

3.3.4 Fitting Models to Historical Occupational Transitions

I create the final dataset by merging the flow count for a particular origin-destination occupation pair o-p, state s, and year group g (y_{opsg}) with the predictors described above.

For the predictors at the state-industry level (GDP and industry score averages, standard deviations, and changes), the merge is done by the combination of state s and year group g. Therefore, a flow count y_{opsg} is matched with the state-industry predictors for all industries that correspond to state s in year group g.

For the predictors at the state-occupation level (occupation score and relative occupation score averages, standard deviations, and changes), the merge is done twice: once by origin occupation o, state s, and year group g, and once by destination occupation p, state s, and year group g. Therefore, a flow count y_{opsg} is assigned one set of state-occupation predictors that correspond to the triple o-s-g and another set of state-occupation predictors that correspond to the triple p-s-g.

With this completed, I then create a separate dataset for each occupation pair in the sample. For instance, for a given origin occupation o and destination occupation p, the observations in the "o-p dataset" are the transitions from occupation o to occupation p across all states s and year groups g.[49] I divide each o-p dataset into a training set (composed of year groups one, two, and three) and a hold-out set (composed of year group four). I train the eXtreme Gradient Boosting (XGBoost) machine learning algorithm on each o-p dataset separately, saving the resulting model. I use the XGBoost algorithm both for its flexibility (in the form of many hyperparameters) and for its incorporation of regularization terms to avoid over-fitting. I tune the hyperparameters using grid search and allow for a different combination to be chosen for each pair of occupations.

At the end of this process, the models together are able to predict the worker flows into and out of any occupation o given the necessary GDP measures. In turn, I generate such predictions using

[49]Each of these datasets contains 204 observations: one for each combination of state (50 U.S. states and the District of Columbia) and year group (2001-2004, 2005-2008, 2009-2012, and 2013-2016).

the predictors for the hold-out set. I discuss the accuracy of those results in more detail below.

As described above, by using these separate datasets, I allow the algorithm to determine the most important industry predictors related to transitions between each pair of occupations. Specifically, the GDP-based industry and occupational scores I construct are just one way to reflect the economic structure of a state at a particular time. By including, in addition, a full set of industry GDP covariates, I allow the algorithm to create its own predictive measures for each pair of origin and destination occupations.

3.4 Results

Before turning to the scenarios and the green/brown occupational categorizations in detail, I first examine the predictions of the models for the hold-out set.

3.4.1 Predictions for the Hold-Out Set

I assess the machine learning algorithm's predictions for the hold-out sample using two metrics: root-mean-squared error (RMSE) and mean absolute error (MAE). I calculate the metrics using the full, floating point (non-integer) predictions from the models as well as the predictions rounded to the nearest integer. In this subsection, I use the metrics to evaluate the models' performance as a whole, though I consider model-specific performance (i.e., at the level of occupations and individual occupation pairs) when predicting transitions under the various scenarios (see below).

For the purposes here, I consider as a baseline the metrics under the "no-information" guess, which corresponds to the values of the RMSE and MAE when all predictions are taken to be zero (the most common flow count in the data). These values serve as a benchmark against which to compare the models' performance.

The results show that the algorithm's predictions are approximately 41-43 percent better than the corresponding no-information RMSE baseline, depending on whether the unrounded or rounded predictions are used (Table 3.3). The results show that the algorithm's predictions are 6 percent (unrounded predictions) or 22 percent (rounded predictions) better than the corresponding MAE baseline.

Table 3.3: RMSE and MAE Results for the Predictions for the Hold-Out Set.

Prediction	RMSE	MAE
No Information	1.94	0.58
Full Predictions	1.10	0.55
Rounded Predictions	1.14	0.46

Taken together, these metrics suggest that the models' predictions are more accurate than the no-information guess of zero worker transitions between each occupation pair, especially when using the RMSE metric. In the results below, I take these same metrics into account at the level of both occupations and occupation pairs.

3.4.2 Categorizing Green and Brown Jobs

In section 3.3 above, I outline three ways of categorizing occupations as green and brown that roughly correspond to viewing occupations at different levels of connectedness within a network that relates industries, occupations, tasks, and tools. These methods view occupations as: (1) related to their specific tasks and tools; (2) connected to the industries in which they are employed; and (3) related to other occupations themselves, in the respect that changes to the economy induce transitions of workers from some occupations to others.

In the following subsections, I consider each of these approaches (using the results of the models as necessary) and explore what this may tell us about green jobs.

Categorizing Occupations by Tasks/Tools and by Industries

I begin by considering two categorizations of green jobs and one categorization of brown jobs based on characteristics of occupations themselves and on characteristics of the industries in which occupations are employed.

The first categorization of green jobs comes from the U.S. Department of Labor's O*NET database, which provides detailed information on the work activities, tools, and technologies associated with occupations. O*NET identifies 204 green occupations across three categories: "green increased demand" (64 occupations); "green enhanced skills" (62 occupations); and "green new and emerging" (78 occupations). These occupation lists were created by Dierdorff et al. (2009, 2011) in several steps. First, more than 60 reports were reviewed to identify job titles associated with green economy activities. Next, these job titles were sorted and distilled into a set of green occupations. Finally, these occupations were assigned to one of the three categories based on the degree of overlap with the occupations that existed in the O*NET database at the time: occupations with a direct match were categorized as "green increased demand," occupations with a close match (but with task or title differences) were categorized as "green enhanced skills," and occupations involving significantly different work were categorized as "green new and emerging."

The O*NET documentation describes how to interpret these green occupational categories. Occupations labeled as "green increased demand" are expected to be in higher employment demand as a result of green economy activities and technologies, though this impact is not anticipated to involve changes in the work or worker requirements of the associated occupations. "Green enhanced skills" occupations are expected to be transformed in terms of their work and worker requirements, though they may or may not experience an increase in employment demand. Finally, "green new and emerging" occupations are entirely new—or significantly changed—occupations associated with unique work and worker requirements driven by the green economy.

I note that each of these has an analog in the model framework described in section 3.2: increased demand for an occupation corresponds to increased demand for the output of the industries in which it is employed; changes in occupational requirements correspond to changes in industries' production processes; and new occupations correspond to new combinations of tasks and/or tools within such processes. Unlike the model, however, the O*NET categorizations are based on job titles associated with green economy activities rather than on explicit scenarios of economic change, and in this way, all three O*NET categories (including increased demand) are most closely aligned with categorization schemes that consider occupations as they relate to their tasks/tools and industries.

The other two categorizations I consider come from Vona et al. (2018), who identify 111 green occupations and 85 brown occupations.[50] Their list of green occupations is based on a subset of the O*NET "green enhanced skills" and "green new and emerging" occupation lists described above. Specifically, in the O*NET database, each of the occupations in these categories is associated with two sets of occupation-specific tasks (i.e., tasks unique to the occupation): green specific tasks and non-green specific tasks. For each occupation, Vona et al. (2018) calculate the fraction of the occupation's specific tasks that are green, yielding a measure that they call "greenness." Occupations with a greenness score greater than zero are considered to be green occupations.[51] As Vona et al. (2018) describe, "defining the greenness of an occupation based on the number of green specific tasks affords a more accurate distinction of green and non-green jobs compared to the O*NET classification, which uses a binary classification to identify green jobs" (p. 719).

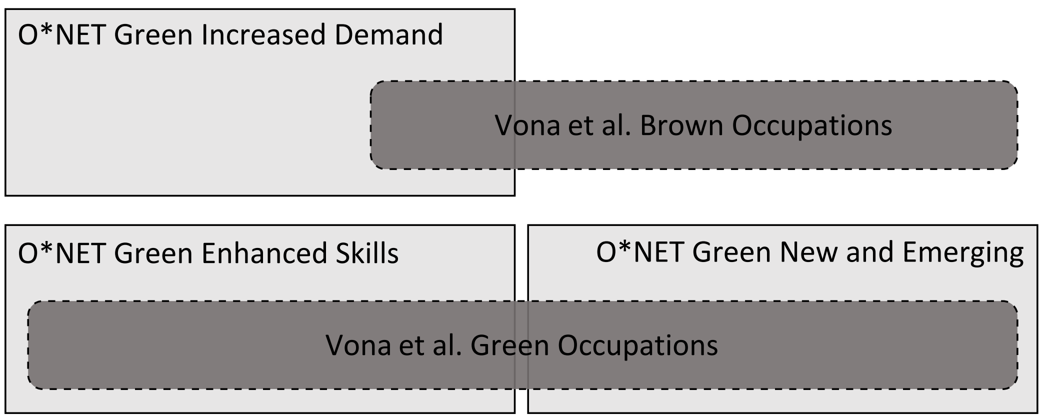
The authors construct their list of brown occupations by first identifying a set of 62 brown industries, which they define as those 4-digit-NAICS industries that emit at least three criteria air pollutants (CO, VOC, NOx, SO2, PM10, PM2.5, and lead) and/or CO2 in the top five percent

[50] In addition to these categorizations, Vona et al. (2018) also develop a data-driven methodology to identify workplace skills most relevant to the green economy. Specifically, using information published in the O*NET database, they identify "Green General Skills" across four high-level categories: engineering and technical; operation management; monitoring; and science. The authors then conduct an empirical analysis of the effect of a regional switch to Clean Air Act non-attainment status on changes in the demand for green skills. Although they do not find an overall effect on employment, such changes do appear to lead to significant, but modest, increases in the demand for green skills.

[51] In discussing this measure, the authors loosely group occupations (for illustrative purposes) into a high-greenness category (greenness of one), a mid-greenness category (greenness in the range [0.3,0.5]) and a low-greenness category (greenness lower than 0.3). See Table 2 and the surrounding discussion in Vona et al. (2018) for additional details.

of all industries as measured by emissions data published by the U.S. Environmental Protection Agency. Brown occupations are then defined as those employed in brown industries at least seven times more than the average employment in brown industries across all occupations. Notably, two of these brown occupations[52] appear on Vona et al.'s (2018) list of green occupations (where they are tied with two others for the lowest greenness score of 0.05). Ten other occupations on Vona et al.'s (2018) brown occupations list appear on the O*NET list of green increased demand occupations. See Figure 3.1 for an illustration of how the O*NET and Vona et al. (2018) categorizations relate to one another.

Figure 3.1: Illustration of O*NET and Vona et al. (2018) Occupational Categories.



Note: this figure illustrates the relationships between the three O*NET occupational categories (green increased demand, green enhanced skills, and green new and emerging) and the green/brown occupational categorizations created by Vona et al. (2018). The sizes of the boxes are for illustrative purposes and only approximate the actual number of occupations in each set. Although the Vona et al. brown occupation list and the Vona et al. green occupation list are shown as non-overlapping, there are two occupations that appear in both. See the main text and footnote 52 for additional details.

Overall, these lists and categorizations provide several benchmarks against which to compare the results of the current study: (1) 64 O*NET green increased demand occupations, 10 of which are labeled as brown occupations by Vona et al.; (2) 75 additional occupations labeled as brown by Vona et al.; and (3) 140 O*NET green enhanced skills / green new and emerging occupations, 111 of which make up Vona et al.'s list of green occupations with an associated greenness score.

Taken together, these form a total of 277 unique occupations, 84 of which (1) have an exact match in the transitions dataset[53] and (2) exchange workers with at least two other occupations. I focus on the outcomes for these 84 occupations and how they compare to the various categorizations

[52] These are "Service Unit Operators, Oil, Gas, and Mining" (occupation code 47-5013.00) and "Separating, Filtering, Clarifying, Precipitating, and Still Machine Setters, Operators, and Tenders" (occupation code 51-9012.00).

[53] Per the discussion above, the occupations in the transitions dataset are those that appear in the CPS data during the period 2001-2012 when considering only occupation pairs with a total flow count in at least one direction (across all states and years) of at least 24.

outlined here. I do not include any occupations in the O*NET and Vona et al. (2018) lists that match only a more aggregated occupation code in the transitions data. In addition, I only include occupations that exchange workers with at least two other occupations (given the sample restrictions described in section 3.3.1) to ensure that the predicted change in net inflow for each occupation is driven by more than a single model.

Green/Brown Occupational Flows during the Training Data Years. As a precursor to examining the predictions under the scenarios, I consider flows among green and brown jobs using the static classifications from Vona et al. (2018) described above. For the purposes of this exploratory analysis, I use the CPS data for the years 2001-2012 (the same dataset used to train the machine learning models). During this period, there were a total of 1,155,762 transitions, about 18 percent and four percent of which had green and brown occupations as their destination, respectively (Table 3.4).

Table 3.4: Green and Brown Outflows/Inflows as a Percentage of Total Occupational Flows.

	Green	Brown
Outflows ÷ Total Flows	18.42%	4.06%
Inflows ÷ Total Flows	17.88%	4.02%

Looking separately at transitions originating from green and brown occupations, we see that green-green and brown-brown flows are more common than brown-green and green-brown flows (Table 3.5). This pattern is especially pronounced when accounting for the relative inflows to each type of destination occupation. Specifically, brown-green flows and green-brown flows are approximately what we would expect given the total inflows to green and brown occupations, respectively. In contrast, brown-brown flows are nearly six times higher than what the brown inflow rate would predict, while the analogous ratio for green-green flows is about 1.6.

Table 3.5: Statistics on Type of Destination by Type of Origin.

	Green Origin	Brown Origin
Total Origin Flows	150,426	33,144
% of Flows to Brown	4.27%	23.90%
% of Flows to Green	28.56%	19.19%
% of Flows to Brown − % of All Flows to Brown	0.24%	19.87%
% of Flows to Green − % of All Flows to Green	10.67%	1.30%
% of Flows to Brown ÷ % of All Flows to Brown	1.06	5.94
% of Flows to Green ÷ % of All Flows to Green	1.60	1.07

Taken together, this suggests that the brown and green occupations defined by Vona et al. (2018) represent two distinct sets of workers, in the sense that a transition into a brown or green occupation is more likely to occur for a worker already in a brown or green occupation, respectively. I turn to this again in the context of the model predictions below.

Categorizing Occupations by Transitions

An alternative approach is to categorize occupations based on worker transitions, which can be conceptualized as the result of changes to the broader networked economy. I describe in this subsection the scenarios of economic change that I consider, how I measure their potential effects on occupations, and what the resulting patterns may tell us about the occupations themselves.

Scenarios of Economic Change. For the purposes of the current study, I generate five economic change scenarios. These scenarios are constructed by averaging the transition predictions that result from one or more sets of modifications to industries' economic growth. Specifically, I create a scenario—call it c—in two stages. In the first stage, I (1) apply a set of modifications, x, to industries' growth, (2) re-predict the transitions corresponding to these modifications, and (3) move to the next set of modifications. In the second stage—after this has been done for all "modification sets" corresponding to a scenario c—I average the transition results. For clarity in the discussion below, I refer to a scenario across all of its modification sets as c, while I refer to an individual modification set within a scenario as (cx).

Two of the scenarios have a single modification set, while the remaining three have multiple modification sets. In the latter case, I choose the modification sets for each of those scenarios so that they share certain characteristics, such as being driven by a given hypothesized shock and/or being transmitted via a particular mechanism. By averaging over these sets, I aim for the scenarios to differ most prominently in these characteristics and not in other idiosyncratic factors (see below).

The industry modifications take the form of adjustments to the values of the predictor gdp_change, which as described in section 3.3, is a metric that captures the average GDP change for each industry in each state across the years in a particular year group. I describe below how I determine, for each modification set within a scenario, which industries will experience a GDP change and by how much. Use $\Delta_{i(cx)}$ to denote the modification for industry i in scenario c's modification set x.

For each 3-digit-NAICS industry i and modification set (cx), I apply the change $\Delta_{i(cx)}$ equally across all states s and all year groups g, such that:

$$\text{gdp\_change}_{sig(cx)} = \text{gdp\_change}_{sig} + \Delta_{i(cx)}$$

where $\text{gdp\_change}_{sig(cx)}$ and gdp_change_{sig} represent the new and original (unmodified) values, respectively, for the predictor gdp_change for state s, industry i, year group g, and modification set (cx).[54] I leave the values for the remainder of the state-industry predictors, and all of the values for the state-occupation predictors, as they originally were.

In essence, these adjustments represent a partial equilibrium change to the economy, where average GDP values (predictor gdp_avg), GDP variation (predictor gdp_sd), and state-occupation factors remain unaffected but where the trend over time (within each year group) is rotated either upwards or downwards. I apply the change for each industry equally across all year groups and across all states so that I can later aggregate (over states) and average (over year groups) to generate transition estimates that are less influenced by the idiosyncratic variation in particular geographies at particular times.

I describe below how I compute these averages and the associated metrics I use in analyzing the results. Here, I turn to describing how I determine the change values $\Delta_{i(cx)}$ for each industry i and modification set (cx).

The first two scenarios are based on Vona et al.'s (2018) list of brown industries at the 4-digit-NAICS level. Given that the predictors for the machine learning models (including gdp_change) are based on data at the 3-digit-NAICS level, I match each of these 4-digit-NAICS brown industries with its closest 3-digit-NAICS counterpart in the U.S BEA GDP data.[55] This yields a list of 23 brown industries at the 3-digit-NAICS level (Table 3.6). Most of these industries fall into manufacturing, with several others across mining, oil/gas extraction, utilities, wholesale trade, transportation, and waste management/remediation.

I then create two scenarios: one where I apply a ten percent decrease to each of these industries separately (and then average the transition results), and another where I apply a ten percent decrease to all industries simultaneously. In other words, the first scenario is associated with 23 modification sets, where each modification set corresponds to a ten percent decrease in the GDP growth of one of the 23 brown industries. The second scenario is associated with a single modification set, where the ten percent changes are applied to all 23 industries at once. I choose a growth decrease of ten percent as it is approximately equal to one standard deviation for the predictor gdp_change.

For the third, fourth, and fifth scenarios, I draw on Chapter 2 to predict the hypothetical manufacturing impacts of historical oil price fluctuations as if they had occurred during the study period. I take these predictions as a proxy for the potential effects of a future policy that increases

[54] Note that I only apply this change if $\text{gdp\_change}_{sig} > 0$, as my intention is to modify the trend of an existing industry and not to introduce a trend where an industry was not originally present in a particular geography.

[55] Note that two of these industries, "Utilities" (NAICS code 22) and "Wholesale trade" (NAICS code 42), appear in the BEA data at the 2-digit level only.

Table 3.6: Percent Changes Based on Brown Industries List.

U.S. BEA 3-digit-NAICS Industry Name	U.S. BEA 3-digit-NAICS Code	Vona et al. 4-digit-NAICS Code	% Change
Oil and gas extraction	211	2111	-10%
Mining (except oil and gas)	212	2121-2123	-10%
Support activities for mining	213	2131	-10%
Utilities	22	2211-2213	-10%
Food and beverage and tobacco products manufacturing	311-312	3111-3115, 3119, 3121, 3122	-10%
Textile mills and textile product mills	313-314	3132, 3133, 3141	-10%
Apparel, leather, and allied product manufacturing	315-316	3161	-10%
Wood product manufacturing	321	3211, 3212, 3219	-10%
Paper manufacturing	322	3221, 3222	-10%
Petroleum and coal products manufacturing	324	3241	-10%
Chemical manufacturing	325	3251-3256, 3259	-10%
Plastics and rubber products manufacturing	326	3262	-10%
Nonmetallic mineral product manufacturing	327	3271-3274, 3279	-10%
Primary metal manufacturing	331	3311-3315	-10%
Fabricated metal product manufacturing	332	3321, 3322, 3328, 3329	-10%
Electrical equipment, appliance, and component manufacturing	335	3359	-10%
Motor vehicles, bodies and trailers, and parts manufacturing	3361-3363	3361	-10%
Other transportation equipment manufacturing	3364-3369	3365, 3366, 3369	-10%
Furniture and related product manufacturing	337	3371	-10%
Wholesale trade	42	4247	-10%
Pipeline transportation	486	4861, 4862, 4869	-10%
Other transportation and support activities	487-488, 492	4881, 4882	-10%
Waste management and remediation services	562	5622, 5629	-10%

Note: this table shows the Vona et al. (2018) 4-digit-NAICS brown industries and the 3-digit-NAICS BEA industries to which they are mapped. I construct two scenarios using these industries: one where I apply a ten percent decrease to the GDP change of each industry separately (and then average the transition results), and another where I apply a ten percent decrease to the GDP change of all industries simultaneously.

the cost of petroleum products for U.S. manufacturing. These scenarios also provide a complement to those based on the brown industries list, as they include many of the same industries while allowing for differential effects across industries.

To generate the values for these scenarios, I leverage the oil episodes regression in Chapter 2, which has its dependent variable the percentage change in value of shipments (or value added) for 6-digit-NAICS manufacturing industries and as its independent variables a series of fixed effects, percentage change in employment and energy usage, direct dependence on petroleum products, and seven industry-level indicators capturing network dynamics. These indicators appear in the regression both on their own and interacted with a year fixed-effect.[56]

For the purposes here, I modify this regression specification in two ways. First, I add an interacted binary variable, call it U, to the terms for upstream paths to refineries, upstream substitute paths, and upstream complement paths. As described in Chapter 2, all three of these indicators capture the effect of increasing or decreasing oil prices as they propagate from upstream to downstream in the production network. Specifically, the first of these indicators reflects the "connectivity" aspects of the network (which industries are connected to which others) while the latter two reflect the "evolutionary" aspects of the network (how connections change in response to shocks).

I add a separate interacted binary variable, call it D, to the terms for downstream substitute paths and downstream complement paths. These two indicators capture the demand-side effect of substitution away from or towards petroleum products during oil price episodes, which can have opposite effects for different industries depending on how their customers—and those customers' customers, and so on—use petroleum products as a substitute or complement to industries' outputs.

The three scenarios based on this regression correspond to predicted manufacturing industry changes in the presence of (1) upstream (i.e., supply-side) shock propagation only, (2) downstream (i.e., demand-side) shock propagation only, and (3) both types of propagation simultaneously.

To construct the associated modification sets, I first predict the change in value of shipments for 6-digit-NAICS manufacturing industries when both binary variables U and D are set to one. I take a weighted average (based on value of shipments) to aggregate these predicted changes to the 3-digit-NAICS level.

I then repeat this process three times: once with U set to zero, once with D set to zero, and once with both set to zero (aggregating to the 3-digit-NAICS level each time). Each of these essentially removes the marginal effect of, respectively, the supply-side network dynamics, the demand-side network dynamics, or both. By taking the difference between the "baseline" prediction (when both

[56] Please see the appendix for additional details regarding the construction of the oil scenarios.

U and D are one) and each of these other predictions, I form an estimated percentage change for each of the 15, 3-digit-NAICS manufacturing industries in the sample—and for each year in the Chapter 2 study period (1968-2018)—corresponding to the supply-side effect, the demand-side effect, and the simultaneous supply-side/demand-side effect of oil price dynamics in each year.

I construct the modification sets for each of the scenarios by choosing sets of percentages from years that—according to the results in Chapter 2—reflect the dynamic of interest. Specifically, for the supply-side scenario, I construct four modification sets, each corresponding to the percentage changes for the 15 manufacturing industries in one of 1974, 1975, 1979, and 1988. I choose these years as the Chapter 2 results suggest that they involved substantial supply-side pass-through of increasing oil prices.

For the demand-side scenario, I construct another four modification sets, each corresponding to the percentage changes for the 15 manufacturing industries in one of 1975, 1985, 1990, and 1997. I choose these years as the Chapter 2 results suggest that they involved a strong shifting away from petroleum products on industries' downstream sides.

For the final scenario, I construct one modification set, which corresponds to the year that appears in both of the previous lists: 1975. For this year, the Chapter 2 results show both the supply-side propagation of increasing oil prices as well as the simultaneous shifting away from petroleum products on the demand side. Unlike above, I consider the upstream and downstream dynamics simultaneously (i.e., I use the percentages generated when both U and D are set to zero, rather than when only one or the other is set to zero).

Finally, to better match the percentages constructed from the brown industries list, I shift and then scale the percentages within each modification set so that they range from -10 percent to zero.[57] In addition, as "petroleum and coal products manufacturing" (3-digit-NAICS industry 324) does not appear in the Chapter 2 sample,[58] I add it to each modification set with a change of -10 percent. I do this both to match the values for the first two scenarios (Table 3.6) and also to represent the precipitating "shock" on which the change values for the other industries are based. I show the percentage values for the four modification sets for the upstream-only scenario below (Table 3.7) and present the values for the other two scenarios in the appendix.

Predicting Transitions. I leverage the trained models to predict the change in worker transitions

[57]Specifically, I make these changes in two steps: (1) I shift all of the percentages so that the largest value becomes zero; and (2) I divide the resulting values by the smallest negative value (i.e., the value of highest magnitude) and then multiply by -0.1. This results in values in the range [-0.1,0.0] while still retaining the original ratios among the values.

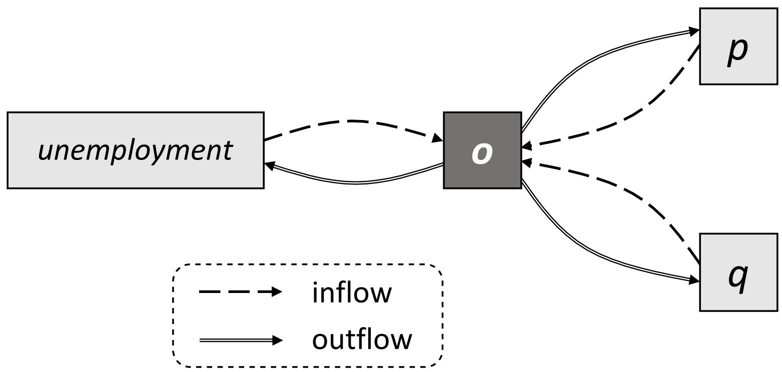
[58]I exclude "petroleum and coal products manufacturing" as an outcome industry in Chapter 2 because I instead focus on the effects of a shock to this industry as it impacts other manufacturing industries via the production network. See Chapter 2 for additional details.

Table 3.7: Percent Changes Based on Upstream-Only Indicators in Oil Regression Results.

U.S. BEA 3-digit-NAICS Industry		% Change			
Name	Code	1974	1975	1979	1988
Food and beverage and tobacco products manufacturing	311-312	-5.98%	-3.92%	-5.38%	-6.22%
Textile mills and textile product mills	313-314	-9.55%	-8.71%	-9.29%	-10.00%
Apparel, leather, and allied product manufacturing	315-316	-9.28%	-9.67%	-8.67%	-8.60%
Wood product manufacturing	321	-6.67%	-4.54%	-4.10%	-6.24%
Paper manufacturing	322	-8.41%	-6.24%	-7.43%	-8.38%
Printing and related support activities	323	-7.53%	-4.09%	-5.85%	-9.32%
Petroleum and coal products manufacturing	324	-10.00%	-10.00%	-10.00%	-10.00%
Chemical manufacturing	325	-6.21%	-5.57%	-4.62%	-6.91%
Plastics and rubber products manufacturing	326	-10.00%	-10.00%	-10.00%	-9.49%
Nonmetallic mineral product manufacturing	327	-7.49%	-4.78%	-5.71%	-7.42%
Primary metal manufacturing	331	-6.82%	-4.6%	-5.78%	-6.93%
Fabricated metal product manufacturing	332	-6.21%	-4.66%	-4.53%	-5.41%
Machinery manufacturing	333	-5.58%	-2.79%	-2.95%	-5.67%
Computer and electronic product manufacturing	334	0.00%	0.00%	0.00%	0.00%
Electrical equipment, appliance, and component manufacturing	335	-8.46%	-7.52%	-8.06%	-8.37%
Furniture and related product manufacturing	337	-5.85%	-6.05%	-6.07%	-5.29%
Miscellaneous manufacturing	339	-9.14%	-7.92%	-8.36%	-7.89%

Note: this table shows the percentage changes for the modification sets corresponding to the upstream-only oil scenario. The values have been shifted and scaled within each modification set so that they fall between -10 percent and zero.

Figure 3.2: Illustration of Occupation-Level Inflows and Outflows.



Note: this figure illustrates the flows of workers between a hypothetical occupation o and three other occupations: p, q, and unemployment. Each of the lines represents a trained machine learning model. Combining the predictions of the three models represented by the dashed lines produces an estimate of the total inflow to o. Combining the predictions of the three models signified by the double solid lines produces an estimate of the total outflow from o. Subtracting the total outflow from the total inflow yields the net inflow into o.

for the 84 focus occupations given the modified gdp_change predictor under each scenario.

Specifically, as described in section 3.3.1, I train models for the transitions o-to-p and p-to-o if the flow counts between o and p in the training data exceed a pre-defined threshold. I similarly train models for o-to-unemployment and unemployment-to-o based on this same threshold. In turn, collecting the models together for an occupation o allows for the prediction of the total inflow to the occupation (i.e., the flow from other occupations p and from unemployment to o) as well as the total outflow from the occupation (i.e., the flow from o to other occupations p and to unemployment).

As an example, consider an occupation o that sends/receives workers from two other occupations, p and q, as well as unemployment (Figure 3.2). Combining the predictions of the three models represented by the dashed lines produces an estimate of the total inflow to o. Similarly, combining the predictions of the three models signified by the double solid lines produces an estimate of the total outflow from o. Subtracting the total outflow from the total inflow yields the net inflow into o: a measure of how employment in o is growing (if net inflow is positive) or declining (if net inflow is negative) vis-à-vis the other occupations in the sample.[59]

Similar to how Vona et al. (2018) take the share of green specific tasks associated with an occupation to be a measure of the greenness of the occupation, I take the predicted change in

[59] I note that there is some similarity between the measure of occupational net inflow described here and the flow approach described by Davis, Faberman, and Haltiwanger (2006). However, whereas Davis, Faberman, and Haltiwanger consider job creation and destruction at the level of the economy and of individual industries—as well as worker flows into and out of unemployment—I consider the flows into and out of specific occupations (beyond unemployment) and use a predictive approach to model such flows.

occupational net inflow—under scenarios that relate to greening of the economy—to be a measure of the greenness of occupations.

To calculate this measure, I first use the trained models to predict, for each occupation o, the flow from o to each other occupation p given the unadjusted (original) values of the predictor gdp_change. I do this at the level of the observations in the transitions dataset, which are identified by the combination of occupation transition o-to-p, state s, and year group g. Call these predicted flows $\hat{y}_{opsg}$. I similarly use the trained models to predict the flow from each other occupation p to o using the unadjusted values, which yields the values $\hat{y}_{posg}$. I then repeat this process using the adjusted values for gdp_change under each modification set (cx). Call the resulting values $\hat{y}_{opsg(cx)}$ and $\hat{y}_{posg(cx)}$.

To calculate net inflow, I first sum these predictions across all states to generate a flow total at the level of occupation transition o-to-p and year group g. Similar to above, call these aggregates $\hat{y}_{opg}$ and $\hat{y}_{pog}$ when using the unadjusted gdp_change values and $\hat{y}_{opg(cx)}$ and $\hat{y}_{pog(cx)}$ when using the adjusted gdp_change values.

Given that each model exhibits error in its predictions, I only consider there to be non-zero difference between the unadjusted and adjusted versions of the predicted values (for each occupation pair, year group, and modification set) if the difference exceeds both the RMSE and the MAE for the model associated with the pair. The notion is that a predicted change less than either the RMSE or MAE could be due to model error rather than a true change in the prediction given the adjustments to the growth rates of the industries.

Specifically, for the model that predicts flows from occupation o to occupation p, call the RMSE of the model's predictions for the hold-out set (i.e., the predictions for year group four across all states) rmse_{op}. Call the MAE of the model's predictions for the hold-out set mae_{op}. I calculate the difference in the prediction for transition o-to-p in year group g and modification set cx as:

$$\text{diff}_{opg(cx)} = \begin{cases} \hat{y}_{opg(cx)} - \hat{y}_{opg} & \text{if} \quad \text{abs}[\hat{y}_{opg(cx)} - \hat{y}_{opg}] > \max[\text{rmse}_{op}, \text{mae}_{op}] \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

In essence, this determines for an occupation o whether each inflow and each outflow—for each combination of year group and modification set—will be treated as having zero or non-zero change. Using Figure 3.2 as an example, this amounts to deciding whether the change in flow corresponding to each of the six arrows will be zero or non-zero over each time period and under each modification set (where such change is measured against the baseline prediction using the original, non-modified gdp_change values).

With this determined, I am then able to calculate the change in net inflow between two occupations o and p (from o's perspective) with the difference $\text{diff}_{pog(cx)} - \text{diff}_{opg(cx)}$. This represents how much more—versus the baseline—occupation o is receiving workers from occupation p than it is sending to occupation p for a given time period and under a given modification set. More specifically, a positive value signifies that occupation o is receiving more workers from p (on net) than it was in the baseline, while a negative value reflects that occupation o is sending more workers to p (on net) than it was in the baseline.

I leverage these values to create two metrics on which the results are based. The first is a measure, for an occupation o, of each other occupation p's individual contribution to o's overall change in net inflow. The second is the overall change in net inflow itself, which I calculate in percentage terms so that the metric can be compared across occupations.

Specifically, I create the first metric for an occupation o in three steps: (1) calculate the fraction of the net inflow change based on p (p-to-o minus o-to-p) out of o's total change in net inflow; (2) scale this fraction by the largest such fraction for o across all of the occupations with which o exchanges workers; and (3) average this scaled measure across the year groups and modification sets within each scenario. These steps are shown in equations 5-7:

$$\text{net\_diff\_frac}_{opg(cx)} = \frac{\text{diff}_{pog(cx)} - \text{diff}_{opg(cx)}}{\text{abs}\left[\sum_{a \in P}\left(\text{diff}_{aog(cx)} - \text{diff}_{oag(cx)}\right)\right]} \tag{5}$$

$$\text{net\_diff\_scaled}_{opg(cx)} = \frac{\text{net\_diff\_frac}_{opg(cx)}}{\max_{a \in P}[\text{abs}(\text{net\_diff\_frac}_{oag(cx)})]} \tag{6}$$

$$\text{net\_diff\_scaled}_{opc} = \left(\sum_{x}\sum_{g}\text{net\_diff\_scaled}_{opg(cx)}\right)\Big/(4 \cdot |X|) \tag{7}$$

where P is the set of occupations with which o exchanges workers and $|X|$ is the number of modification sets within scenario c.

I create the second metric in two steps: (1) calculate the percentage change in net inflow across all of the occupations with which o exchanges workers, using as the denominator the predicted net inflow based on the unadjusted gdp_change values; and (2) average these percentages across the year groups and modification sets within each scenario. These steps are shown in equations 8 and 9:

$$\text{net\_pct}_{og(cx)} = \frac{\sum_{a \in P} \Big(\text{diff}_{aog(cx)} - \text{diff}_{oag(cx)} \Big)}{\text{abs}\Big[\sum_{a \in P} \Big(\hat{y}_{aog} - \hat{y}_{oag} \Big) \Big]} \tag{8}$$

$$\text{net\_pct}_{oc} = \Bigg(\sum_{x} \sum_{g} \text{net\_pct}_{og(cx)} \Bigg) \Big/ (4 \cdot |X|) \tag{9}$$

where P is the set of occupations with which o exchanges workers and $|X|$ is the number of modification sets within scenario c.

I note that in calculating equation 8, I apply a threshold at the occupation level similar to the one I use at the occupation-pair level (equation 4). Specifically, considering all of the models related to occupation o (i.e., models where o is either the origin or the destination), call the RMSE of the models' predictions for the hold-out set rmse_o. Call the MAE of the models' predictions for the hold-out set mae_o. Analogous to above, I consider the percentage change in net inflow for o to be non-zero within a year group and modification set only if the numerator and the denominator in equation 8 are both larger than $\max[\text{rmse}_o, \text{mae}_o]$. This avoids a situation where the numerator and/or denominator in equation 8 are non-zero only due to model error, resulting in a non-zero value for $\text{net\_pct}_{og(cx)}$ that is either very small (in the former case) or potentially very large (in the latter case).

As described above, by calculating the metrics separately for each year group and modification set (equations 5, 6, and 8) and then averaging across them (equations 7 and 9), I aim to capture differences in patterns of transitions due to the main characteristics of the scenarios themselves rather than due to differences across the year groups—such as the historical GDP values for the non-affected industries—or due to other idiosyncrasies of the particular modification sets.

Altogether, the values produced by equation 9 induce a ranking—for each scenario—over the set of 84 focus occupations, which I discuss below in more detail. Using Figure 3.2 as an example, equation 9 produces the predicted percentage change in net inflow for focus occupation o (e.g., +2 percent) under a given scenario and across all of the occupations with which o exchanges workers (p, q, and unemployment together).

The values produced by equation 7 instead create an ordering for each occupation o of the contributions to o's net inflow change in a scenario attributable to each other occupation. In Figure 3.2, this corresponds to the individual contributions of occupations p, q, and unemployment to o's overall change, where these individual contributions may be positive or negative but where they "average out" to the overall effect (e.g., combine to form +2 percent).

Results of Approach. Using these metrics, I consider the transitions associated with the scenarios in three ways: (1) the overall rankings of the occupations, including how they differ by scenario; (2) the relationships between the rankings and the categorizations provided by O*NET and Vona et al. (2018); and (3) patterns in worker exchange among occupation pairs.

Given the percentage change in net inflow values (equation 9 above), I am able to rank the 84 focus occupations from "most green" (rank 1) to "least green" (rank 84) for each of the five scenarios (see the appendix for a full list of these rankings). As a first method to examine the orderings, I calculate the Spearman rank correlation coefficient for each scenario pair (Table 3.8). All of the correlations are significant at the one percent level.

Table 3.8: Rank Correlations Among Scenario Occupational Orderings.

Scenario	(1)	(2)	(3)	(4)	(5)
(1) Brown industries (individual changes)	1.00				
(2) Brown industries (simultaneous changes)	0.97	1.00			
(3) Regression industries (upstream only)	0.74	0.79	1.00		
(4) Regression industries (downstream only)	0.69	0.73	0.92	1.00	
(5) Regression industries (both up and down)	0.69	0.72	0.93	0.96	1.00

Note: the values in this table are the Spearman rank correlation coefficients when comparing the rankings across the five scenarios. All of the correlations are significant at the one percent level.

The table reveals that the two rankings based on the brown industries list are highly correlated, which suggests that applying the differences to the affected industries individually (and then averaging) does not produce a significantly different occupational ordering than when applying the differences simultaneously. The oil regression scenarios are correlated with the brown industries scenarios with coefficients in the range of 0.69-0.79. The upstream-only oil regression scenario is somewhat more correlated with the brown industries scenarios than are the other two oil scenarios.

Altogether, the correlations broadly suggest that the rankings across the scenarios are not statistically independent. At the same time, visual inspection of the rankings reveals that there are some patterns underlying these high-level findings. For instance, while some occupations tend to have similar rankings across all of the scenarios, others have rankings in the oil regression scenarios that differ substantially from their rankings in the brown industries scenarios.

In the latter case, these differences (1) sometimes appear noticeably in the upstream-only scenario but not in the downstream-only scenario, (2) sometimes in the downstream-only scenario but not in the upstream-only scenario, and (3) sometimes in both. Although all three of these suggest that some occupations may be relatively more affected—as compared to the brown industries scenarios—by the propagation of a petroleum products price increase through the U.S. production network, the first two further suggest that the manner in which the shock is propagated is also relevant.

To explore this more systematically, I use hierarchical clustering to split the 84 occupations into groups.[60] This method works by first assigning each occupation to its own group, and then iteratively combining the two closest groups (based on differences in the rankings; see below) until only a single group remains. Working backwards, the algorithm is then able to specify which occupations belong to which groups when moving from one group to two groups, two groups to three groups, and so on.

There are various techniques, also known as stopping rules, to determine when to conclude this splitting process to yield the optimal number of groups. I consider both the Calinski-Harabasz and the Duda-Hart stopping-rules. The first of these is global in the sense that it considers information about all groups (regardless of where the algorithm is in the splitting process), while the second is local in that it considers the cluster structure of the next group to be split. Based on these rules, I decide to explore how the occupations are partitioned when there are ten groups.[61]

As mentioned above, the closeness of the groups is based on differences in the rankings, with the goal of differentiating occupations that have substantially shifted rankings in some scenarios versus others. To accomplish this, I provide the clustering algorithm with six differences, all of which are calculated as absolute values. The first three of these are differences between the average of the rankings in the brown industries scenarios and the rankings in each of the regression scenarios. The next two are differences between the ranking in the combination upstream/downstream oil scenario and the rankings in each of the upstream-only and downstream-only oil scenarios. The last is the difference between the rankings in the upstream-only and the downstream-only scenarios. Taken together, these values allow the algorithm to consider how the rankings for each occupation change between the brown industries scenarios and the oil regression scenarios, as well as how the rankings change among the oil scenarios themselves.

After inspecting the ten groups, I assign them names to aid in interpretability (see the appendix for a full table of which occupations are placed into which groups). The first of these groups—which I call "Similar"—contains half of the occupations. As its name suggests, the occupations in this group have generally similar rankings across all scenarios, with differences in rank between the brown industries scenarios and the oil regression scenarios of about 4-5 on average (Table 3.9).

For example, the occupations "Construction Workers, Other" and "Computer Control Program-

[60]Specifically, I use the hierarchical agglomerative linkage clustering algorithm as implemented in Stata. I apply Ward's method to perform the clustering, which by default in Stata uses squared Euclidian distance.

[61]Starting with four groups, the Calinski-Harabasz approach produces pseudo-F values that generally increase with each additional group, up to a local maximum with 11 groups. Based on the Je(2)/Je(1) index and the associated pseudo-T^2 value, the Duda-Hart rule suggests that the occupations be split into three, four, six, eight, ten, or larger groups (12 or more). As a balance between the two stopping rules—and to aid interpretability—I choose to explore the clustering using ten groups.

mers and Operators" have relatively high rank across the five scenarios, ranging from #14 to #19 and from #5 to #12, respectively. Both occupations are categorized by the algorithm as "Similar." At the other end of the greenness spectrum are occupations like "Paper Goods Machine Setters, Operators, and Tenders" (rankings ranging from #69 to #77) and "Mining Machine Operators" (rankings ranging from #80 to #84), which are also both categorized as "Similar."

Table 3.9: Clustering Algorithm Group Names and Summary of Differences in Rankings.

		Average Magnitude of Change:		
Group	Count	Up	Down	Combination
Similar	42	4	5	4
Differences (smallest)	4	11	8	8
Differences (smaller)	16	16	19	18
Differences (larger)	4	37	37	37
Differences (largest)	2	54	54	54
Upstream/Combination	10	9	4	10
Downstream/Combination	3	4	28	10
Civil Engineers	1	3	22	53
Packaging and Filling Machine Operators and Tenders	1	36	2	13
Other Installation, Maintenance, and Repair Workers	1	53	76	78

Note: this table lists the ten groups produced by the hierarchical clustering algorithm. For each group, the table shows the number of focus occupations placed into that group, as well as the magnitudes of the ranking changes seen for members of the group within each of the oil regression scenarios (as compared to the average of their rankings in the brown industries scenarios).

Another 20 occupations are placed into two groups that I label "Differences (smallest)" and "Differences (smaller)." These are occupations that have different rankings in the oil regression scenarios as compared to the brown industries scenarios, though the differences are modest (averaging about 8-11 for the first group and 16-19 for the second) and are consistent across the three regression scenarios (Table 3.9). Two examples in the "Differences (smaller)" group are "Construction Managers" and "Chemists and Materials Scientists." The first of these is ranked #49 and #52 in scenarios 1-2 and #27, #29, and #26 in scenarios 3-5. The second is ranked #27 in scenarios 1-2 and #40, #48, and #50 in scenarios 3-5. In both cases, the rankings in the brown industries scenarios are closer to one another than they are to the rankings in the regression scenarios.

The algorithm identifies two additional groups that share this type of pattern, though the ranking changes are of greater magnitude. I call these groups "Differences (larger)" and "Differences (largest)," and they contain four and two occupations, respectively (Table 3.9). An example in the first group is "Material Moving Workers, Other" (rankings #67-#67-#22-#23-#24), and an example in the second group is "Chemical Engineers" (rankings #69-#58-#5-#13-#14). As above, the rankings in the regression scenarios for these occupations tend to be similar, though they differ substantially from the rankings in the brown industries scenarios.

For some of the occupations in these "Differences" groups, a likely contributing factor of the shifting patterns in their rankings is that the brown industries scenarios include a ten percent decrease for several non-manufacturing industries—including wholesale trade, mining, and waste management/remediation services—that the oil regression scenarios do not. For example, consider the occupation "Dredge, Excavating, and Loading Machine Operators," which is ranked higher in the oil regression scenarios (#32, #34, and #36 in scenarios 3-5) than it is in the brown industries scenarios (#56 and #53 in scenarios 1-2). This occupation trades workers with six other occupations (and unemployment), including several related to construction, construction equipment, and construction management. In turn, impacts to manufacturing—which may have knock-on effects for construction-related activities—without concurrent impacts to mining may be one reason why the predicted changes in net inflow for this occupation are relatively larger (i.e., more favorable for the occupation) under the oil regression scenarios.

At the same time, the results suggest that such dynamics are likely more nuanced than the simple inclusion or exclusion of specific industries. Instead, it appears to be the degree to which industries are affected—and more specifically, differentially affected compared to other industries—that can lead the rankings to vary by scenario. This is evidenced by the groups I call "Upstream/Combination" and "Downstream/Combination." The occupations in these groups are those that experience changes in their rankings in the oil regression scenarios as compared to the brown industries scenarios, though the changes are more pronounced in certain regression scenarios than others.

Specifically, the ten occupations in the "Upstream/Combination" group have average ranking changes of 9 and 10 in the upstream-only and combination scenarios, respectively, while they have an average ranking change of 4 in the downstream-only scenario (Table 3.9). Similarly, the three occupations in the "Downstream/Combination" group have average ranking changes in the downstream-only and combination scenarios of 28 and 10, respectively, compared to an average ranking change of 4 in the upstream-only scenario. That there is some but not complete consistency in the rankings moving from the brown industries scenarios to the oil regression scenarios (for the occupations in these groups) suggests that the manner in which shocks are propagated in the economic production network—such as supply-side pass-through versus the back-propagation of demand—could be a relevant factor leading to differential outcomes for some occupations.

Finally, the last three groups contain a single occupation each, which are identified by the algorithm as having ranking patterns that differ from all of the other focus occupations. These three occupations are "Civil Engineers" (rankings #83-#84-#81-#62-#31), "Packaging and Filling Machine Operators and Tenders" (rankings #15-#18-#52-#18-#29), and "Other Installation, Maintenance, and Repair Workers" (rankings #3-#5-#57-#80-#82). The first two of these exhibit ranking patterns similar to, respectively, the occupations in the "Downstream/Combination" and "Upstream/Combination" groups, though the magnitudes of the changes are generally larger (Table 3.9). The "Other Installation, Maintenance, and Repair Workers" occupation has rankings that are, in some sense, a combination of patterns seen in the "Differences (largest)" and "Downstream/Combination" groups, with large ranking changes across all three of the oil regression scenarios that are especially large in the downstream-only and combination scenarios. I leave for future research more detailed exploration of why these occupations experience such ranking changes.

Altogether, the rank correlations provide evidence that the scenario orderings—specifically, those based on the oil regressions as compared to those based on the brown industries list—are not statistically independent. At the same time, the rankings do differ substantially for some occupations, and a clustering algorithm is able to use these differences to group occupations in a sensible way. The results suggest that about half of the occupations in the study experience moderate to large changes in their rankings when moving from some scenarios to others. The patterns in these differences further suggest that the greening of the economy may have heterogeneous impacts for occupations based not just on which industries are affected, but also on the manner in which resulting price changes are propagated through the U.S. production network.

More research would be needed to better understand how and why these (and potentially other) scenarios produce different rankings for some occupations. For now, I take these scenarios and their associated rankings as given and turn instead to considering how the rankings correspond to the O*NET and Vona et al. (2018) categorizations.

To do this, I first calculate the Spearman rank correlation between the ordering created by the Vona et al. (2018) greenness score and the orderings created by the five scenarios (for the 27 occupations in the study sample that have a greenness score; Table 3.10). These correlation coefficients range from 0.25 to 0.29, which provides some evidence that the greenness score of green occupations has a relationship to predicted change in net inflow, at least under the scenarios considered here. At the same, none of the correlations are quite statistically significant at standard levels.

These correlations consider only the greenness scores of the green occupations and, in turn, do not consider the potential differences in outcomes between the green and brown occupations. To explore this in more detail, I calculate the predicted mean percentage change in net inflow for the Vona et al. (2018) green and brown occupations in each scenario (Table 3.11).

Across all of the scenarios, green occupations are predicted to do worse than brown occupations,

Table 3.10: Rank Correlations Between Scenario Orderings and Greenness Ordering.

Scenario	Correlation	p-value
Brown industries (individual changes)	0.29	0.14
Brown industries (simultaneous changes)	0.27	0.17
Regression industries (upstream only)	0.27	0.18
Regression industries (downstream only)	0.25	0.22
Regression industries (both up and down)	0.29	0.14

Note: the values in this table are the Spearman rank correlation coefficients between the ordering created by the Vona et al. (2018) greenness score and the orderings created by the five scenarios (for the 27 occupations in the study sample that have a greenness score). None of the correlations are statistically significant at standard levels.

in the sense that they have negative change in net inflow on average while the brown occupations have positive change in net inflow on average. However, one-sided tests of the differences of these means are not statistically significant at any standard level.

Table 3.11: Means: Green versus Brown.

	Mean		p-value	
Scenario	Green	Brown	Green > Brown	Green < Brown
Brown industries (individual changes)	-0.01	0.00	0.77	0.23
Brown industries (simultaneous changes)	-0.17	0.09	0.81	0.19
Regression industries (upstream only)	-0.03	0.08	0.69	0.31
Regression industries (downstream only)	-0.02	0.01	0.58	0.42
Regression industries (both up and down)	-0.01	0.01	0.54	0.46

Note: this table displays the predicted mean percentage change in net inflow for the green and brown occupations categorized by Vona et al. (2018). The p-values correspond to Welch's t-test for samples with unequal variances.

Overall, these results suggest—as measured by change in net inflow—that (1) there may be a relationship between an occupation's greenness score and its predicted outcomes but (2) that green occupations overall are not predicted to grow more, on average, than are brown occupations. In future research, the consideration of a greater number of occupations (i.e., a larger sample that includes more of the Vona et al. green and brown occupations) as well as a greater diversity of scenarios could help to clarify such patterns.

I construct similar means that compare the O*NET increased demand occupations to the other two O*NET categories (green enhanced skills and green new and emerging; Table 3.12). I consider the increased demand occupations separate from occupations in the other two O*NET groups because the former—per the O*NET definitions—are those that we would expect to have greater change in net inflow given the greening of the economy.

The patterns here are, in some sense, the opposite of above, where the increased demand occupations are predicted to have higher change in net inflow, on average, than are the other two groups. I note that the differences are more pronounced in the first three scenarios, though none are statistically significant at any standard level.

Table 3.12: Means: Increased Demand versus Enhanced Skills / New and Emerging.

	Mean		p-value	
Scenario	ID	ES/NE	ID > ES/NE	ID < ES/NE
Brown industries (individual changes)	0.00	-0.01	0.23	0.77
Brown industries (simultaneous changes)	0.08	-0.17	0.21	0.79
Regression industries (upstream only)	0.07	-0.03	0.34	0.66
Regression industries (downstream only)	-0.01	-0.02	0.47	0.53
Regression industries (both up and down)	0.02	-0.01	0.43	0.57

Note: this table displays the predicted mean percentage change in net inflow for the occupations in the O*NET "green increased demand" category ("ID") as compared to the corresponding mean for the occupations in the other two O*NET categories ("ES/NE"). The p-values correspond to Welch's t-test for samples with unequal variances.

Relatedly, although more research would be needed to understand why increased demand occupations are predicted to do better in the upstream-only scenario as compared to the other two oil scenarios, one potential factor is the combination of occupations included in each O*NET category. Specifically, relative to the enhanced skills and new and emerging occupations—which include many managers, engineers, technicians, technologists, and scientists—the O*NET increased demand occupations tend to be those involved with various types of production, equipment, and machinery. These latter occupations are more likely to be employed in industries that produce outputs that will be used alongside petroleum products. In turn, substitution away from petroleum products on the downstream side (and towards other industries) may be one reason that the change in net inflow for increased demand occupations is predicted to be lower in the two scenarios involving such dynamics.

Finally, using the contribution metric constructed above (equation 7), I consider (1) which occupations are predicted to have the largest contributions—positive, negative, and in absolute value—to other occupations' change in net inflow, (2) how patterns in the contribution measure correlate with the overall occupational rankings, and (3) how worker transitions are predicted to change between green and brown occupations specifically.

For clarity in this discussion, I continue to refer to the 84 occupations that have been ranked above as focus occupations. From the perspective of these occupations, there is a set of occupations with which they exchange workers; call these "exchange" occupations. Note that some focus occupations may also appear as exchange occupations from the perspective of other focus occupations.

I first consider which exchange occupations have the largest positive values, the largest negative values, and the largest absolute values for the scaled contribution measure. Across all of the scenarios, the values of this metric are normally distributed with means of approximately zero (Table 3.13).

Table 3.13: Summary Statistics for Contribution Measure by Scenario.

Scenario	Mean	Std. Dev.	Min.	Max.
Brown industries (individual changes)	0.00	0.03	-0.22	0.19
Brown industries (simultaneous changes)	0.00	0.27	-1.00	1.00
Regression industries (upstream only)	0.00	0.25	-1.00	1.00
Regression industries (downstream only)	0.00	0.22	-1.00	1.00
Regression industries (both up and down)	0.00	0.23	-1.00	1.00

Note: these are summary statistics for the contribution measure net_diff_scaled as defined in equation 7. All values are rounded to the hundredths place.

For each scenario, I calculate the mean contribution for each exchange occupation across all of the focus occupations with which it exchanges workers. For the purposes of this analysis, I consider only those exchange occupations that trade workers with at least five focus occupations so that the averages represent more than just a handful of pairwise relationships.

I present lists of the top positive contributors and top negative contributors in the appendix. Exchange occupations with positive values for the contribution measure are those that are predicted to increase their exchange of workers in the direction of the focus occupations. This does not imply that the exchange occupations would themselves be predicted to decline overall (as they likely have pairwise relationships with occupations other than the focus occupations, which I do not consider), but it does suggest that these occupations, on average, are predicted serve as a source of workers in the scenarios. Some of these occupations are related to production, manufacturing, and machinery, including "Machinists," "Heavy Vehicle and Mobile Equipment Service Technicians and Mechanics," and "First-Line Supervisors of Mechanics, Installers, and Repairers." Others are more general, including "Carpenters," "Designers," and "Customer Service Representatives."

In contrast, exchange occupations with negative values for the contribution measure are those that are predicted to increase their trade of workers in the direction of the exchange occupations themselves. As above, this does not imply that these exchange occupations would be predicted to grow overall, but it does suggest that these occupations, on average, are predicted to serve as a sink of workers across the scenarios. This list includes a number of more technical and specialized occupations—such as "Electrical and Electronics Engineers" and "Industrial and Refractory Machinery Mechanics"—while it also includes many general occupations, such as "Waiters and Waitresses," "Cashiers," and "Grounds Maintenance Workers."

Although it is beyond the scope of this chapter, an opportunity for future research would be to consider the welfare and equity implications of the transitions associated with these lists. For instance, patterns in wage differentials across occupation pairs (i.e., the average change in wage for each exchange occupation across all of the focus occupations with which it exchanges workers) could help to illuminate the worker-level impacts of transitions associated with various sources and sinks.

I also consider exchange occupations that have the largest mean contribution in absolute value terms (see the appendix for a full table of these values). These are the occupations that, regardless of the direction in which they tend to exchange workers (i.e., towards or away from the focus occupations), are predicted to have the largest impacts on percentage change in net inflow for the focus occupations. As with the prior two lists, these include a variety of occupations, some more specialized (e.g., “Heating, Air Conditioning, and Refrigeration Mechanics and Installers”) and many more general (e.g., “Construction Laborers” and “Secretaries and Administrative Assistants”).

Notably, the second entry on this list is the “Unemployed” occupation, which reveals that the transition to, and/or the transition out of, unemployment is a major contributor to the predicted change in net inflow for many focus occupations.[62] I return to this in more detail below.

Another factor to consider at the pairwise level is whether the occupational rankings are systematically related to the back-and-forth transition of workers with certain exchange occupations. Specifically, although the ranking for each focus occupation is based on the predicted change in net inflow vis-à-vis at least two, and often many, exchange occupations, there may be some exchange occupations that stand out from others in that (1) they trade workers with many focus occupations and (2) their contribution measure is correlated with the rankings of those focus occupations. Where this type of relationship appears, it suggests that the associated exchange occupations may serve as “intermediaries” between other occupations, in the sense that they are predicted receive workers from declining occupations (i.e., those with negative change in net inflow) while they are predicted to send workers to growing occupations (i.e., those with positive change in net inflow).

To explore these dynamics, I separately regress the occupational rankings in each scenario against each exchange occupation’s scaled contribution measure. In essence, these regressions ask whether the scaled contribution measure of an exchange occupation is significantly related to the rankings of the focus occupations with which it exchanges workers.

The regression results show that the scaled contribution measure of each individual exchange occupation is often not related to the rankings of the focus occupations with which it trades workers.

[62] I note that although “Unemployed” appears on the list of top negative contributors, it is listed 21st. Combined with the results for the absolute value of the contribution measure, this suggests that the flows between the focus occupations and unemployment tend to occur in both directions, even though (on average across the scenarios) increased flow towards unemployment is the somewhat more prevalent pattern.

In fact, of the more than 250 exchange occupations that I consider, about 15 (on average) have a statistically significant relationship at the ten percent level in each scenario (see the appendix for a list of exchange occupations with at least one statistically significant relationship across the scenarios). Part of this may be due to the fact that some exchange occupations are associated with fewer focus occupations, though there are many exchange occupations that appear frequently in the data that also do not exhibit this type of relationship.

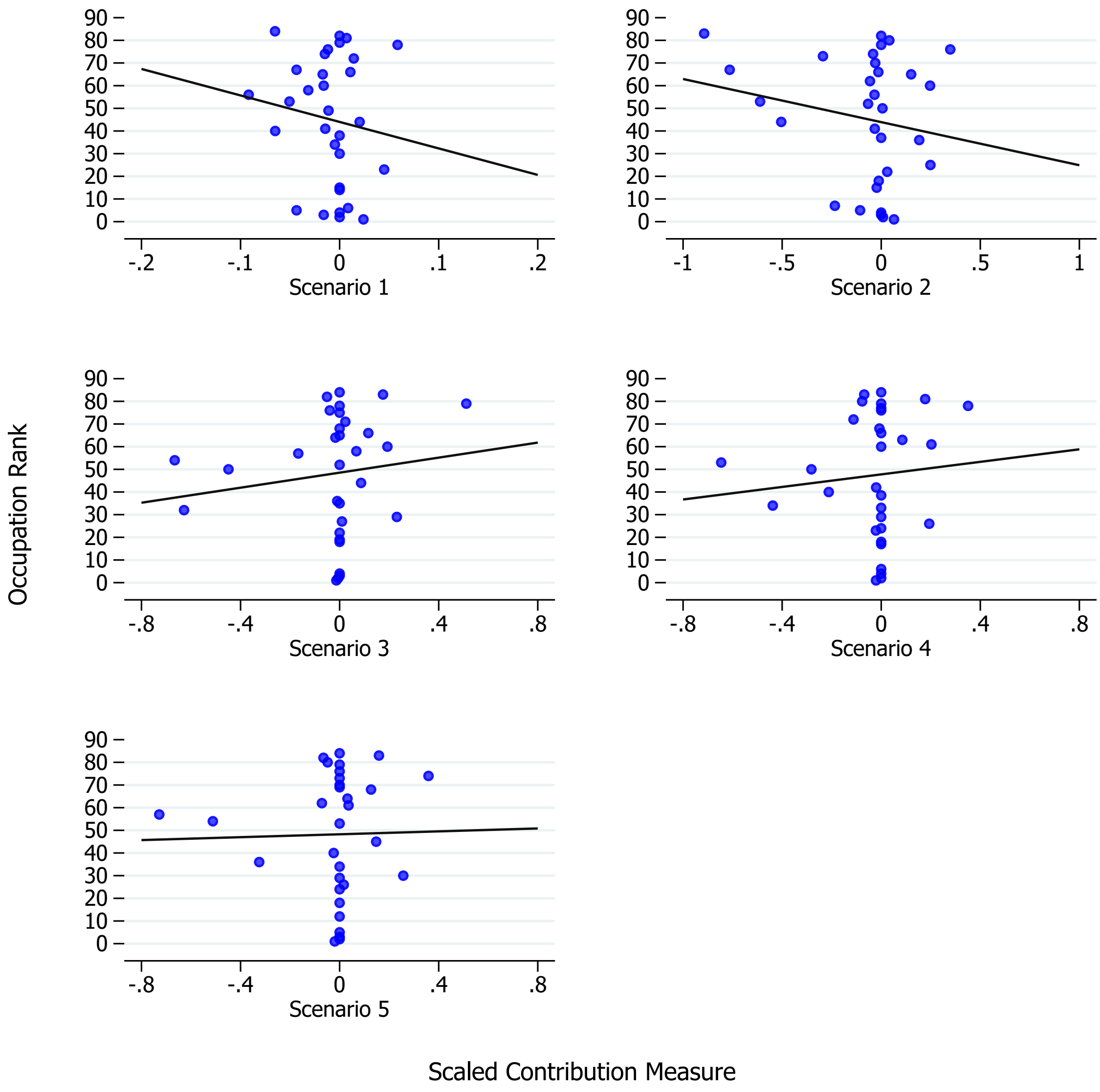
As an example, take "Driver/Sales Workers and Truck Drivers," which acts as an exchange occupation for 31 focus occupations in the study. The regression results show that the scaled contribution measure for this occupation is not systematically related to the rankings of the occupations with which it exchanges workers. This is evidenced by the plots in Figure 3.3, which illustrate that positive values for the contribution measure are associated with both high and low rankings, while negative values for the measure are also associated with both high and low rankings.

When a significant relationship does exist for an exchange occupation, it sometimes appears across all of the scenarios. As an example, take the occupation "Retail Salespersons," which is an exchange occupation for 20 focus occupations, including "Construction Managers," "Maintenance and Repair Workers, General," and "First-Line Supervisors of Production and Operating Workers." In its role as an exchange occupation, "Retail Salespersons" has a statistically significant relationship (at either the one or five percent level) in both of the brown industries scenarios and in all three of the oil regression scenarios. Figure 3.4 illustrates that a positive value for this exchange occupation's contribution measure (i.e., greater predicted inflow, on net, from this exchange occupation) tends to correlate with a higher rank, while a negative value for its contribution measure (i.e., greater predicted outflow, on net, to this exchange occupation) tends to correlate with lower rank.

There are other exchange occupations that exhibit a significant relationship with their focus occupations across some but not all of the scenarios (for additional details, refer to the full list of significant coefficients in the appendix). In some cases, this suggests that transmission mechanisms could be relevant in predicting an intermediary-type role for certain exchange occupations, though as with the occupational rankings (and associated groupings) more generally, additional research is needed to better understand such dynamics.

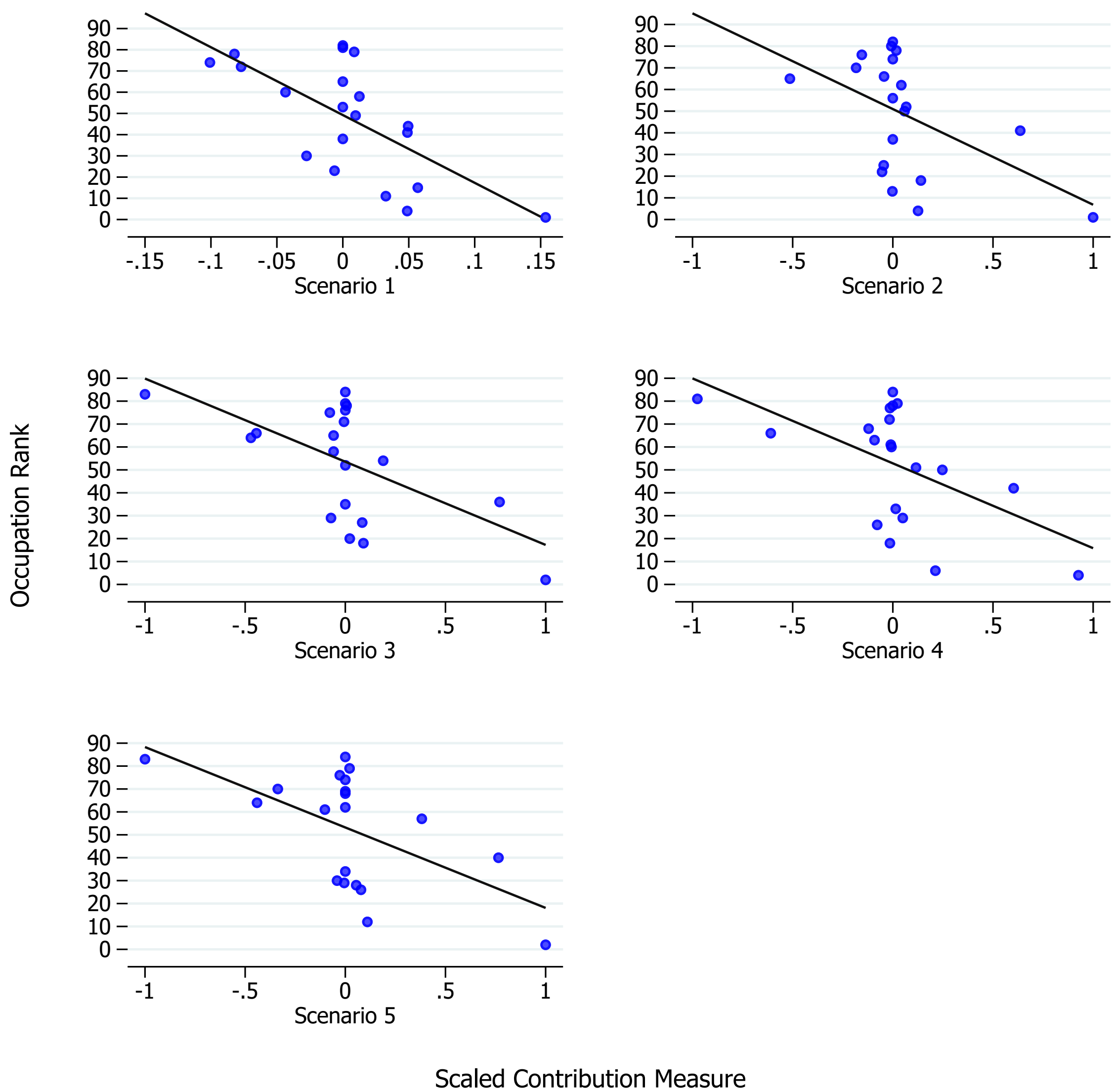
There is one exchange occupation that trades workers with more focus occupations than any other, and which also exhibits a strong relationship with the rankings: unemployment. As illustrated by Figure 3.5, the scaled contribution measure for the "Unemployed" occupation is closely related to the overall occupational rankings across all of the scenarios. As with "Retail Salespersons," when unemployment is predicted to send workers (on net) to another occupation, that occupation tends to

Figure 3.3: Contribution Measure: "Driver/Sales Workers and Truck Drivers"



Note: this figure illustrates the relationship between the scaled contribution measure for the exchange occupation "Driver/Sales Workers and Truck Drivers" and the overall occupational rankings. The lines in each plot are fitted through the points in the associated scenario. None of the relationships are statistically significant at standard levels.

Figure 3.4: Contribution Measure: "Retail Salespersons"



Note: this figure illustrates the relationship between the scaled contribution measure for the exchange occupation "Retail Salespersons" and the overall occupational rankings. The lines in each plot are fitted through the points in the associated scenario. The relationships in scenarios one, three, and five are statistically significant at the one percent level, and the relationships in scenarios two and four are statistically significant at the five percent level.

be ranked high. When unemployment is predicted to take workers (on net) from an occupation, that occupation tends to be ranked low. This suggests that—at least in these scenarios of a transition to a green economy—the exchange of workers into and out of unemployment is predicted to be one of the key factors determining which occupations are growing and which are declining.

I note again that the negative slopes in the figures for "Retail Salespersons" and "Unemployed" do not imply that these exchange occupations would themselves be predicted to be growing or declining overall (i.e., if these occupations were themselves studied as focus occupations, they could potentially have either positive or negative predicted change in net inflow[63]). Rather, these negative relationships suggest that—for the associated focus occupations—a high rank often involves "receiving more workers" from the associated exchange occupation while a low rank often involves "sending more workers" to the associated exchange occupation.

In fact, that many exchange occupations (excluding those on the top positive/negative contributors lists) have both positive and negative contribution measures across the scenarios suggests that exchange occupations are generally not predicted to serve as either sources of workers or sinks of workers, but both. Said another way, although some focus occupations have a higher ranking because they are predicted to draw a greater number of workers (on net) from a particular exchange occupation, there are simultaneously other focus occupations—even within the same scenario—that have a lower ranking because they are predicted to lose a greater number of workers (on net) to that same exchange occupation. From the perspective of a particular worker, this movement could be considered beneficial, neutral, or detrimental depending on the worker's preferences, the wages of the occupations in question, and other related factors. As mentioned above, one avenue for future work is a detailed exploration of the equity implications of such patterns.

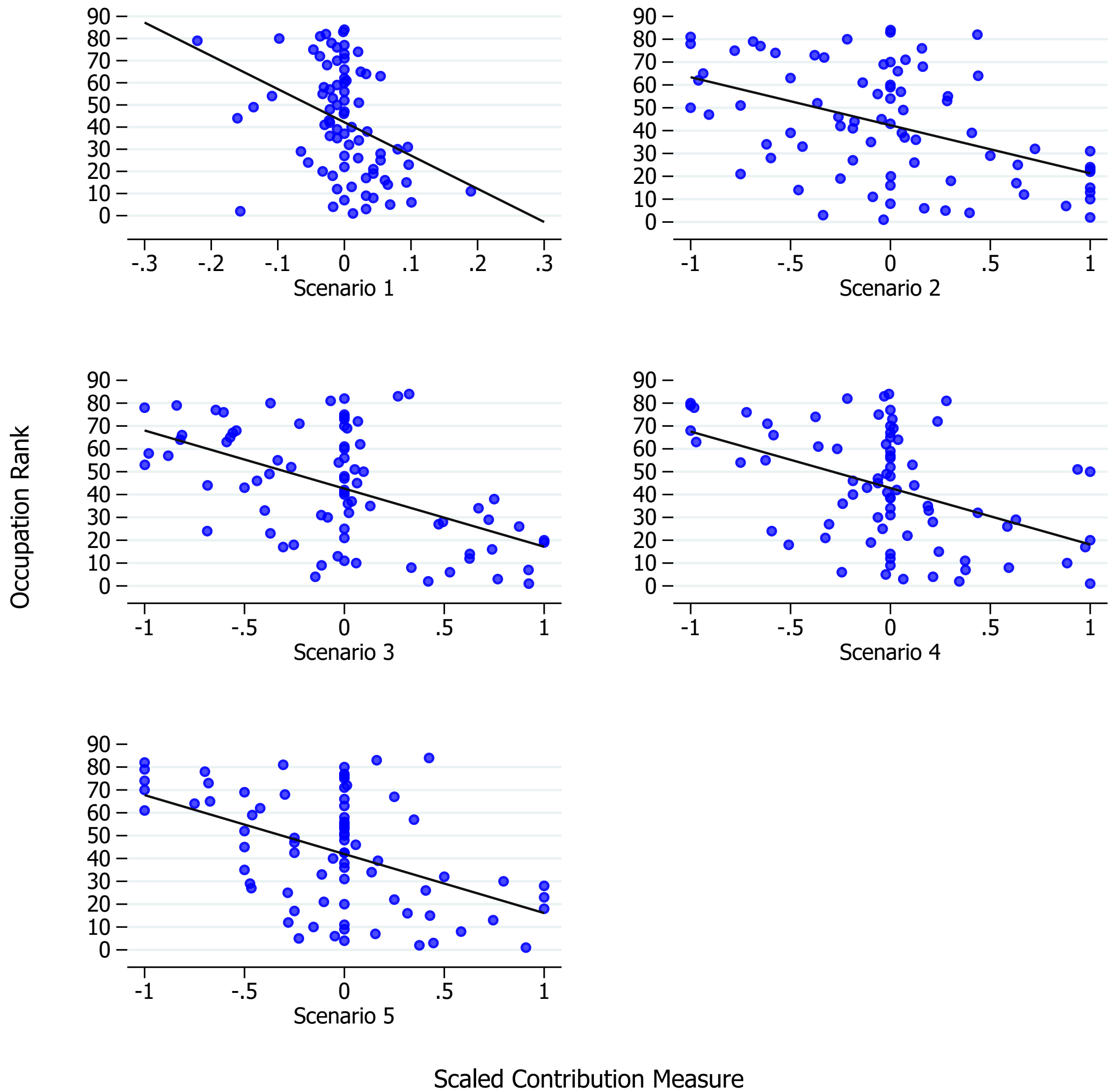
Finally, beyond considering individual exchange occupations one at a time, another approach is to consider exchange occupations as they would be categorized by Vona et al. (2018).

I do this first by regressing the rankings against the scaled contribution measures of (1) green exchange occupations as a group and (2) brown exchange occupations as a group. As with the occupations "Retail Salespersons" and "Unemployed," the results show that the rankings and the contribution measures are negatively correlated across all scenarios, with relationships that are statistically significant at the ten percent level in all but one case (see the appendix for a figure with these results).

As above, this suggests that back-and-forth worker trades with green and brown occupations

[63]In addition, that the points in the plots are not clearly to the left or to the right in any of the scenarios suggests that these occupations may be ranked towards the middle.

Figure 3.5: Contribution Measure: "Unemployed"



Note: this figure illustrates the relationship between the scaled contribution measure for the exchange occupation "Unemployed" and the overall occupational rankings. The lines in each plot are fitted through the points in the associated scenario. The relationships in all scenarios are statistically significant at the one percent level.

tend to be correlated with the overall occupational rankings. However, given that the results again reveal both positive and negative contribution measures across the scenarios, they do not provide insight into whether green or brown exchange occupations are, on average, predicted to be increasing or decreasing flows away from or towards the focus occupations.

To explore this, I calculate the means for the scaled contribution measure for green and brown exchange occupations under each scenario. I first consider the means of the scaled contribution measure when looking across all of the focus occupations, and I find that the values for the green exchange occupations are positive while the values for the brown exchange occupations are negative (Table 3.14). These differences are significant or near-significant at the 10 percent level. This suggests that, for the 84 focus occupations as a whole, green exchange occupations are predicted to send workers in the direction of the focus occupations while brown exchange occupations are predicted to take workers away from the focus occupations.

Table 3.14: Means of Green/Brown Contribution Measure for All Focus Occupations.

	Mean		p-value	
Scenario	Green	Brown	Green > Brown	Green < Brown
Brown industries (individual changes)	0.004	-0.001	0.09	0.91
Brown industries (simultaneous changes)	0.022	-0.014	0.12	0.88
Regression industries (upstream only)	0.039	-0.002	0.08	0.92
Regression industries (downstream only)	0.024	-0.005	0.12	0.88
Regression industries (both)	0.031	-0.002	0.11	0.89

Note: this table displays the mean contribution measure for green and brown exchange occupations when considering all focus occupations. The p-values correspond to Welch's t-test for samples with unequal variances.

If I look instead at these means when computed for green and brown focus occupations separately, somewhat different patterns emerge. Specifically, considering only green focus occupations, the contribution measure for green exchange occupations tends to be relatively small, while the measure for brown exchange occupations is negative and somewhat larger (Table 3.15). As above, these differences are statistically significant or nearly significant at the ten percent level. This suggests that green focus occupations are predicted to send workers towards brown exchange occupations.

When considering only brown focus occupations, the opposite relationship appears, where the contribution measure of brown exchange occupations is near zero while the measure for green exchange occupations is positive and larger (Table 3.16). The difference in means is only statistically significant at the ten percent level in scenarios one and three, though the pattern is consistent across all of the scenarios. These results suggest that brown focus occupations are predicted, on average, to receive workers from green exchange occupations.

Table 3.15: Means of Green/Brown Contribution Measure for Green Focus Occupations.

	Mean		p-value	
Scenario	Green	Brown	Green > Brown	Green < Brown
Brown industries (individual changes)	0.001	-0.007	0.08	0.92
Brown industries (simultaneous changes)	-0.006	-0.055	0.13	0.87
Regression industries (upstream only)	-0.006	-0.076	0.04	0.96
Regression industries (downstream only)	-0.010	-0.050	0.10	0.90
Regression industries (both)	-0.013	-0.055	0.12	0.88

Note: this table displays the mean contribution measure for green and brown exchange occupations when considering only green focus occupations. The p-values correspond to Welch's t-test for samples with unequal variances.

Table 3.16: Means of Green/Brown Contribution Measure for Brown Focus Occupations.

	Mean		p-value	
Scenario	Green	Brown	Green > Brown	Green < Brown
Brown industries (individual changes)	0.009	-0.001	0.07	0.93
Brown industries (simultaneous changes)	0.044	0.008	0.30	0.70
Regression industries (upstream only)	0.085	0.012	0.09	0.91
Regression industries (downstream only)	0.047	0.006	0.21	0.79
Regression industries (both)	0.065	0.009	0.16	0.84

Note: this table displays the mean contribution measure for green and brown exchange occupations when considering only brown focus occupations. The p-values correspond to Welch's t-test for samples with unequal variances.

Taken together, these findings provide some insight into the pattern seen above: that, in terms of change in net inflow, green occupations are predicted to decline while brown occupations are predicted to grow. (Noting that, as described previously, the differences in these growth rates are not statistically significant.) The results here suggest that, for green focus occupations, this pattern is at least partially related to the transition of workers towards brown exchange occupations. From the perspective of brown focus occupations, the pattern is partially a result of the predicted flow of workers from green exchange occupations.

These dynamics also provide perspective on another pattern noted earlier that, in the CPS data for 2001-2012, a transition into a green or brown occupation was more likely to occur for a worker already in a green or brown occupation, respectively. The results here predict that while green-green and brown-brown transitions may balance out, green-to-brown transitions may accelerate. This provides some additional evidence that the greening of the economy may not necessarily be associated with a homogeneous transfer of workers from brown to green occupations.

3.5 Conclusion

In this chapter, I take an empirical approach to exploring the relationship between occupational transitions and green jobs. Specifically, I build a series of models that relate the evolution of U.S. industries during the period 2001-2012 to historical occupation-to-occupation flows. I then leverage these models to explore the potential implications of five scenarios of economic change for a set of 84 focus occupations using predicted percentage change in net inflow as a central outcome measure.

I find that rankings based on this measure are generally correlated across the scenarios, though inspection at the level of individual occupations reveals that there are patterns of diversity underlying these higher-level trends. In particular, the groupings created by a hierarchical clustering algorithm suggest that the greening of the economy could have heterogeneous impacts for occupations based not just on the particular industries affected, but also on the manner in which resulting price changes are propagated through the U.S. production network (e.g., via upstream versus downstream channels).

Contrary to what we might expect, I find that the green occupations identified by Vona et al. (2018) are predicted to do worse under the scenarios than are the brown occupations, though there is enough variation in outcomes that the differences are not statistically significant at any standard level. At the same time, considering green occupations on their own, I find that the greenness score of an occupation is positively correlated with its rankings across the scenarios.

The models predict increased growth for the O*NET increased demand occupations vis-à-vis the occupations in the other O*NET categories, though the differences are again not statistically significant. As with the Vona et al. (2018) categorizations, additional research, including a larger set of focus occupations, could help to further illuminate how and why these occupational categorizations—based on occupations' titles, tasks/tools, and/or industries—may or may not differ from the rankings produced by the approach in this chapter.

The results at the level of individual occupation pairs suggest that some exchange occupations could play more of a prominent role than others in affecting the change in net inflow for the focus occupations. This is most true of the "Unemployed" occupation, which exchanges workers with almost every focus occupation in the sample and which is predicted to have the second-largest contribution (in absolute value) of any exchange occupation across the scenarios. Along with a number of other exchange occupations, its contribution measure is correlated in a statistically significant way with the overall occupational rankings in all of the scenarios that I consider.

These patterns for unemployment, as well as the patterns for many exchange occupations generally, suggest that the evolution towards a green U.S. economy may not be accompanied by a homogeneous transfer of workers from some occupations to others. Rather, in combination with the overall occupational rankings themselves, these results suggest that a given set of industrial changes may have heterogeneous effects based on the back-and-forth trading of workers between focus and exchange occupations, where such transitions may benefit some workers while detrimentally affecting others. In turn, a key avenue for future research is to better understand these dynamics from an equity and welfare standpoint.

References for Chapter 1

Acemoglu, Daron, Vasco M. Carvalho, Asuman Ozdaglar, and Alireza Tahbaz-Salehi. 2012. The Network Origins of Aggregate Fluctuations. Econometrica 80(5): 1977-2016.

Acemoglu, Daron, Ufuk Akcigit, and William Kerr. 2016. Networks and the Macroeconomy: An Empirical Exploration. NBER Macroeconomics Annual 30(1): 273-335.

Acemoglu, Daron, Asuman Ozdaglar, and Alireza Tahbaz-Salehi. 2017. Microeconomic Origins of Macroeconomic Tail Risks. American Economic Review 107(1): 54-108.

Atalay, Enghin. 2017. How Important are Sectoral Shocks? American Economic Journal: Macroeconomics 9(4): 254-280.

Barbosa, Igor Barros, Marco Cristani, Barbara Caputo, Aleksander Rognhaugen, and Theoharis Theoharis. 2018. Looking Beyond Appearances: Synthetic Training Data for Deep CNNs in Re-Identification. Computer Vision and Image Understanding 167: 50-62.

Barrot, Jean-Noel and Julien Sauvagnat. 2016. Input Specificity and The Propagation of Idiosyncratic Shocks in Production Networks. The Quarterly Journal of Economics 131(3): 1543-1592.

Baqaee, David Rezza. 2018. Cascading Failures in Production Networks. Econometrica 86(5): 1819-1838.

Baqaee, David Rezza and Emmanuel Farhi. 2018. Unbundling Macroeconomics Via Heterogeneous Agents and Input-Output Networks. NBER Working Paper Series No. 24684.

Baqaee, David Rezza and Emmanuel Farhi. 2019. The Macroeconomic Impact of Microeconomic Shocks: Beyond Hulten's Theorem. Econometrica 87(4): 1155-1203.

Berg, Alan M., Stefan T. Mol, Gabor Kismihok, and Niall Sclater. 2016. The Role of a Reference Synthetic Data Generator within the Field of Learning Analytics. Journal of Learning Analytics 3(1): 107-128.

Boehm, Christoph E., Aaron Flaaen, and Nitya Pandalai-Nayar. 2019. Input Linkages and the Transmission of Shocks: Firm-Level Evidence from the 2011 Tōhoku Earthquake. Review of Economics and Statistics 101(1): 60-75.

Brockway, Paul E., Matthew K. Heun, Joao Santos, and John R. Barrett. 2017. Energy-Extended CES Aggregate Production: Current Aspects of Their Specification and Econometric Estimation. Energies 10(2).

Brummitt, Charles D., Andres Gomez-Lievvano, Ricardo Hausmann, and Matthew H. Bonds. 2020. Machine-Learned Patterns Suggest that Diversification Drives Economic Development. Journal of the Royal Society Interface 17(162).

Carvalho, Vasco M. 2014. From Micro to Macro via Production Networks. Journal of Economic Perspectives 28(4): 23-48.

Carvalho, Vasco M. and Tahbaz-Salehi, Alireza. 2019. Production Networks: A Primer. Annual Review of Economics 11: 635-663.

Carvalho, Vasco M., Makoto Nirei, Yukiko U. Saito, and Alireza Tahbaz-Salehi. 2021. Supply Chain Disruptions: Evidence from the Great East Japan Earthquake. The Quarterly Journal of Economics 136(2): 1255-1321.

Dissou, Yazid, Lilia Karnizova, and Qian Sun. 2015. Industry-Level Econometric Estimates of Energy-Capital-Labor Substitution with a Nested CES Production Function. Atlantic Economic Journal 43(1): 107-121.

Dutting, Paul, Zhe Feng, Harikrishna Narasimhan, David C. Parkes, and Sai Srivatsa Ravindranath. 2019. Optimal Auctions through Deep Learning. In International Conference on Machine Learning (pp. 1706-1715). PMLR.

Feenstra, Robert C., Philip Luck, Maurice Obstfeld, and Katheryn N. Russ. 2018. In Search of the Armington Elasticity. Review of Economics and Statistics 100(1): 135-150.

Foerster, Andrew T., Pierre-Daniel G. Sarte, and Mark W. Watson. 2011. Sectoral versus Aggregate Shocks: A Structural Factor Analysis of Industrial Production. Journal of Political Economy 119(1): 1-38.

Gabaix, Xavier. 2011. The Granular Origins of Aggregate Fluctuations. Econometrica 79(3): 733-772.

Long, John B. and Charles I. Plosser. 1983. Real Business Cycles. Journal of Political Economy 91(1): 39-69.

MaCurdy, Thomas E. and John H. Pencavel. 1986. Testing between Competing Models of Wage and Employment Determination in Unionized Markets. Journal of Political Economy 94(3): S3-S39.

Mayer, Nikolaus, Eddy Ilg, Philipp Fischer, Caner Hazirbas, Daniel Cremers, Alexey Dosovitskiy, and Thomas Brox. 2018. What Makes Good Synthetic Training Data for Learning Disparity and Optical Flow Estimation? International Journal of Computer Vision 126: 942-960.

McNerney, James. 2009. Network Properties of Economic Input-Output Networks. International Institute for Applied Systems Analysis Interim Report IR-09-003.

Moate, Chris P., Stephen D. Hayward, Jonathan S. Ellis, Lee Russell, Ralph O. Timmerman, Richard O. Lane, and Thomas J. Strain. 2018. Vehicle Detection in Infrared Imagery Using Neural Networks with Synthetic Training Data. In International Conference Image Analysis and Recognition (pp. 453-461). Springer.

Mundra, Kusum and R. Robert Russell. 2004. Dual Elasticities of Substitution. Working Paper.

Papageorgiou, Chris, Marianne Saam, and Patrick Schulte. 2017. Substitution Between Clean and Dirty Energy Inputs: A Macroeconomic Perspective. Review of Economics and Statistics 99(2): 281-290.

Schmidt, Michael and Hod Lipson. 2009. Distilling Free-Form Natural Laws from Experimental Data. Science 324(5923): 81-85.

Stern, David I. 2011. Elasticities of Substitution and Complementarity. Journal of Productivity Analysis 36(1): 79-89.

References for Chapter 2

Acemoglu, Daron, Vasco M. Carvalho, Asuman Ozdaglar, and Alireza Tahbaz-Salehi. 2012. The Network Origins of Aggregate Fluctuations. Econometrica 80(5): 1977-2016.

Acemoglu, Daron, Asuman Ozdaglar, and Alireza Tahbaz-Salehi. 2017. Microeconomic Origins of Macroeconomic Tail Risks. American Economic Review 107(1): 54-108.

Acemoglu, Daron, Ufuk Akcigit, and William Kerr. 2016. Networks and the Macroeconomy: An Empirical Exploration. NBER Macroeconomics Annual 30(1): 273-335.

Aldy, Joseph E. and William A. Pizer. 2015. The Competitiveness Impacts of Climate Change Mitigation Policies. Journal of the Association of Environmental and Resource Economists 2(4): 565-595.

Antràs, Pol, Davin Chor, Thibault Fally, and Russell Hillberry. 2012. Measuring the Upstreamness of Production and Trade Flows. American Economic Review: Papers and Proceedings 102(3): 412-416.

Atalay, Enghin. 2017. How Important are Sectoral Shocks? American Economic Journal: Macroeconomics 9(4): 254-280.

Barrot, Jean-Noel and Julien Sauvagnat. 2016. Input Specificity and The Propagation of Idiosyncratic Shocks in Production Networks. The Quarterly Journal of Economics 131(3): 1543-1592.

Barsky, Robert B. and Lutz Kilian. 2002. Do We Really Know that Oil Caused the Great Stagflation? A Monetary Alternative. NBER Macroeconomics Annual 2001, 16: 137-183.

Baqaee, David Rezza. 2018. Cascading Failures in Production Networks. Econometrica 86(5): 1819-1838.

Baqaee, David Rezza and Emmanuel Farhi. 2018. Unbundling Macroeconomics Via Heterogeneous Agents and Input-Output Networks. NBER Working Paper Series No. 24684.

Baumeister, Christiane and Lutz Kilian. 2016. Forty Years of Oil Price Fluctuations: Why the Price of Oil May Still Surprise Us. Journal of Economic Perspectives 30(1): 139-160.

Carvalho, Vasco M., Makoto Nirei, Yukiko U. Saito, and Alireza Tahbaz-Salehi. 2021. Supply Chain Disruptions: Evidence from the Great East Japan Earthquake. The Quarterly Journal of Economics 136(2): 1255-1321.

Curtis, Mark E. 2018. Who Loses under Cap-and-Trade Programs? The Labor Market Effects of the NOx Budget Trading Program. The Review of Economics and Statistics 100(1): 151–166.

Davis, Steven J. and John Haltiwanger. 2001. Sectoral job creation and destruction responses to oil price changes. Journal of Monetary Economics 48: 465-512.

Di Giovanni, Julian, Andrei A. Levchenko, and Isabelle Mejean. 2018. The Micro Origins of International Business-Cycle Comovement. American Economic Review 108(1): 82-108.

Deschenes, Olivier, Michael Greenstone, and Joseph S. Shapiro. 2017. Defensive Investments and the Demand for Air Quality: Evidence from the NOx Budget Program. American Economic Review 107(10): 2958-2989.

Foerster, Andrew T., Pierre-Daniel G. Sarte, and Mark W. Watson. 2011. Sectoral versus Aggregate Shocks: A Structural Factor Analysis of Industrial Production. Journal of Political Economy 119(1): 1-38.

Fowlie, Meredith L., Mar Reguant, and Stephen P. Ryan. 2016. Measuring Leakage Risk. Report for the California Air Resources Board.

Gabaix, Xavier. 2011. The Granular Origins of Aggregate Fluctuations. Econometrica 79(3): 733-772.

Ganapati, Sharat, Joseph S. Shapiro, and Reed Walker. 2018. The Incidence of Carbon Taxes in U.S. Manufacturing: Lessons from Energy Cost Pass-Through. NBER Working Paper Series No. 22281.

Hamilton, James D. 1983. Oil and the Macroeconomy since World War II. Journal of Political Economy 91(2): 228-248.

Hamilton, James D. 1985. Historical Causes of Postwar Oil Shocks and Recessions. The Energy Journal 6(1): 97-116.

Kahn, Matthew E. and Erin T. Mansur. 2013. Do Local Energy Prices and Regulation Affect the Geographic Concentration of Employment? Journal of Public Economics 101: 105-114.

Kilian, Lutz. 2008a. The Economic Effects of Energy Price Shocks. Journal of Economic Literature 46(4): 871-909.

Kilian, Lutz. 2008b. Exogenous Oil Supply Shocks: How Big Are They and How Much Do They Matter for the U.S. Economy? Review of Economics and Statistics 90(2): 216-40.

Kilian, Lutz, and Thomas K. Lee. 2014. Quantifying the Speculative Component in the Real Price of Oil: The Role of Global Oil Inventories. Journal of International Money and Finance 42: 71-87.

Kilian, Lutz, and Daniel P. Murphy. 2014. The Role of Inventories and Speculative Trading in the Global Market for Crude Oil. Journal of Applied Econometrics 29(3): 454-78.

Langer, Ashley and Nathan H. Miller. 2013. Automakers' Short-Run Responses to Changing Gasoline Prices. Review of Economics and Statistics 95(4): 1198-1211.

Lee, Kiseok and Shawn Ni. 2002. On the dynamic effects of oil price shocks: a study using industry level data. Journal of Monetary Economics 49: 823-852.

Long, John B. and Charles I. Plosser. 1983. Real Business Cycles. Journal of Political Economy 91(1): 39-69.

Muehlegger, Erich and Richard L. Sweeney. 2017. Pass-Through of Input Cost Shocks Under Imperfect Competition: Evidence from the U.S. Fracking Boom. NBER Working Paper Series No. 24025.

Stolper, Samuel. 2016. Who Bears the Burden of Energy Taxes? The Critical Role of Pass-Through. Harvard Environmental Economics Program Discussion Paper 2016-70.

References for Chapter 3

Acemoglu, Daron and David Autor. 2011. Skills, Tasks and Technologies: Implications for Employment and Earnings. Handbook of Labor Economics 4b: 1043-1171.

Acemoglu, Daron, Vasco M. Carvalho, Asuman Ozdaglar, and Alireza Tahbaz-Salehi. 2012. The Network Origins of Aggregate Fluctuations. Econometrica 80(5): 1977-2016.

Acemoglu, Daron, Asuman Ozdaglar, and Alireza Tahbaz-Salehi. 2017. Microeconomic Origins of Macroeconomic Tail Risks. American Economic Review 107(1): 54-108.

Acemoglu, Daron, Ufuk Akcigit, and William Kerr. 2016. Networks and the Macroeconomy: An Empirical Exploration. NBER Macroeconomics Annual 30(1): 273-335.

Aldy, Joseph E. and William A. Pizer. 2014. The Employment and Competitiveness Impacts of Power-Sector Regulations. In C. Coglianese, A. M. Finkel, & C. Carrigan (Ed.), Does Regulation Kill Jobs? (pp. 70-88). University of Pennsylvania Press.

Allen, Matthew T., Gordon Waugh, Megan Shaw, Suzanne Tsacoumis, David Rivkin, Phil Lewis, Mark Brendle, Denise Craven, Christy Gregory, and Dan Connell. 2012. The Development and Evaluation of a New O*NET Related Occupations Matrix. National Center for O*NET Development.

Atalay, Enghin. 2017. How Important are Sectoral Shocks? American Economic Journal: Macroeconomics 9(4): 254-280.

Autor, David H. and Michael J. Handel. 2013. Putting Tasks to the Test: Human Capital, Job Tasks, and Wages. Journal of Labor Economics 31(S1): S59-S96.

Baqaee, David Rezza. 2018. Cascading Failures in Production Networks. Econometrica 86(5): 1819-1838.

Baqaee, David Rezza and Emmanuel Farhi. 2018. Unbundling Macroeconomics Via Heterogeneous Agents and Input-Output Networks. NBER Working Paper Series No. 24684.

Barrot, Jean-Noel and Julien Sauvagnat. 2016. Input Specificity and The Propagation of Idiosyncratic Shocks in Production Networks. The Quarterly Journal of Economics 131(3): 1543-1592.

Bowen, Alex, Karlygash Kuralbayeva, and Eileen L. Tipoe. 2018. Characterising Green Employment: The Impacts of “Greening” on Workforce Composition. Energy Economics 72: 263-275.

Burger, Martijn, Spyridon Stavropoulos, Shyaam Ramkumar, Joke Dufourmont, and Frank van Oort. 2019. The Heterogeneous Skill-Base of Circular Economy Employment. Research Policy 48: 248-261.

Carvalho, Vasco M., Makoto Nirei, Yukiko U. Saito, and Alireza Tahbaz-Salehi. 2021. Supply Chain Disruptions: Evidence from the Great East Japan Earthquake. The Quarterly Journal of Economics 136(2): 1255-1321.

Curtis, Mark E. 2018. Who Loses under Cap-and-Trade Programs? The Labor Market Effects of the NOx Budget Trading Program. The Review of Economics and Statistics 100(1): 151–166.

Davis, Steven J., R. Jason Faberman, and John Haltiwanger. 2006. The Flow Approach to Labor Markets: New Data Sources and Micro–Macro Links. Journal of Economic Perspectives 20(3): 3-26.

Dierdorff, Erich C., Jennifer J. Norton, Donald W. Drewes, Christina M. Kroustalis, David Rivkin, and Phil Lewis. 2009. Greening of the World of Work: Implications for O*NET-SOC and New and Emerging Occupations. The National Center for O*NET Development.

Dierdorff, Erich C., Jennifer J. Norton, Christina M. Gregory, David Rivkin, and Phil Lewis. 2011. Greening of the World of Work: Revisiting Occupational Consequences. The National Center for O*NET Development.

Di Giovanni, Julian, Andrei A. Levchenko, and Isabelle Mejean. 2018. The Micro Origins of International Business-Cycle Comovement. American Economic Review 108(1): 82-108.

Foerster, Andrew T., Pierre-Daniel G. Sarte, and Mark W. Watson. 2011. Sectoral versus Aggregate Shocks: A Structural Factor Analysis of Industrial Production. Journal of Political Economy 119(1): 1-38.

Frey, Carl Benedikt and Michael A. Osborne. 2017. The Future of Employment: How Susceptible are Jobs to Computerisation? Technological Forecasting and Social Change 114: 254-280.

Gabaix, Xavier. 2011. The Granular Origins of Aggregate Fluctuations. Econometrica 79(3): 733-772.

Glaeser, Edward L. (Ed.). 2010. Agglomeration Economics. University of Chicago Press.

Hidalgo, Cesar A., Bailey Klinger, A.-L. Barabasi, and Ricardo Hausmann. 2007. The Product Space Conditions the Development of Nations. Science 317: 482-487.

Kahn, Matthew E. and Erin T. Mansur. 2013. Do local energy prices and regulation affect the geographic concentration of employment? Journal of Public Economics 101: 105-114.

Kuminoff, Nicolai V., Todd Schoellman, and Christopher Timmins. 2015. Environmental Regulations and the Welfare Effects of Job Layoffs in the United States: A Spatial Approach. Review of Environmental Economics and Policy 9(2): 198-218.

Lazear, Edward P. 2009. Firm-Specific Human Capital: A Skill-Weights Approach. Journal of Political Economy 117(5): 914-940.

Long, John B. and Charles I. Plosser. 1983. Real Business Cycles. Journal of Political Economy 91(1): 39-69.

Mealy, Penny, Maria del Rio-Chanona, and J. Doyne Farmer. 2018. What You Do at Work Matters: New Lenses on Labour. Available at SSRN: http://dx.doi.org/10.2139/ssrn.3143064.

Morgenstern, Richard D., William A. Pizer, and Jhih-Shyang Shih. 2002. Jobs Versus the Environment: An Industry-Level perspective. Journal of Environmental Economics and Management 43(3): 412-436.

Muneepeerakul, Rachata, Jose Lobo, Shade T. Shutters, Andres Gomez-Lievano, and Murad R. Qubbaj. 2013. Urban Economies and Occupation Space: Can They Get "There" from "Here"? PLoS ONE 8(9): 1-8.

Nedelkoska, Ljubica, Dario Diodato, and Frank Neffke. 2018. Is Our Human Capital General Enough to Withstand the Current Wave of Technological Change? Harvard Center for International Development Fellow and Graduate Student Working Paper No. 93.

Neffke, Frank and Martin Henning. 2013. Skill Relatedness and Firm Diversification. Strategic Management Journal 34(3): 297-316.

Neffke, Frank, Anne Otto, and Antje Weyh. 2017. Inter-Industry Labor Flows. Journal of Economic Behavior and Organization 142: 275-292.

Parrado, Eric, Asena Caner, and Edward N. Wolff. 2007. Occupational and Industrial Mobility in the United States. Labour Economics 14(3): 435–455.

Shutters, Shade T., Rachata Muneepeerakul, and Jose Lobo. 2016. How hard is it for urban economies to become 'green'? Environment and Planning B: Planning and Design 43: 198-209.

Timmermans, Bram and Ron Boschma. 2014. The Effect of Intra- and Inter-Regional Labour Mobility on Plant Performance in Denmark: The Significance of Related Labour Inflows. Journal of Economic Geography 14: 289-311.

Van Oort, Frank, Shyaam Ramkumar, Spyridon Stavropoulos, Martijn Burger, Joke Dufourmont, and Mark Thissen. 2018. Putting the Circular Economy to Work: Measuring Gross Employment Effects in the Netherlands, 1996-2015. Working Paper. Erasmus University Rotterdam.

Vona, Francesco, Giovanni Marin, Davide Consoli, and David Popp. 2018. Environmental Regulation and Green Skills: An Empirical Exploration. Journal of the Association of Environmental and Resource Economists 5(4): 713-753.

Walker, Reed W. 2013. The Transitional Costs of Sectoral Reallocation: Evidence from the Clean Air Act and the Workforce. The Quarterly Journal of Economics 128(4): 1787-1835.

A Appendix for Chapter 1

A.1 Leontief Input-Output Analysis

This section briefly reviews the general method of input-output analysis developed by Wassily Leontief. Suppose the economy is categorized into n industries, each producing its own product. Industry i generates total output x_i (in dollars), and the final demand for industry i's product—that is, the amount sold to consumers, the government, and other end users—is d_i (in dollars). Industry i's product may also be sold to other industries in the economy. Define z_{ij} as the amount (in dollars) that industry i sells to industry j. From this, we can express x_i as follows:

$$x_i = z_{i1} + z_{i2} + ... + z_{in} + d_i = \sum_{j=1}^{n} z_{ij} + d_i$$

We can compactly express the equations x_i for all industries by using matrix notation. To do this, we organize the production and demand values into the column vectors x and d, respectively, and collect the values z_{ij} into a matrix Z. In addition, define k as a column vector of ones. Altogether, we have the following:

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \qquad d = \begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_n \end{bmatrix}, \qquad Z = \begin{bmatrix} z_{11} & z_{12} & \cdots & z_{1n} \\ z_{21} & z_{22} & \cdots & z_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ z_{n1} & z_{n2} & \cdots & z_{nn} \end{bmatrix}, \qquad k = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$$

We can then write:

$$x = Zk + d$$

As Miller and Blair (2009)[64] describe, input-output analysis takes the ratio of inputs for an industry to be fixed. To see this, define $a_{ij} = z_{ij}/x_j$. This is a unitless ratio (dollars divided by dollars) representing the amount industry j spends on industry i's output when producing one dollar's worth of its own output. Rearranging, we have $z_{ij} = a_{ij}x_j$. Holding a_{ij} fixed, this implies that as industry j's level of production (x_j) increases, the amount it demands from industry i (z_{ij}) increases proportionally. Given that all z_{ij} are defined in this manner, this also implies that—holding the a_{ij} fixed—the ratio of inputs for a particular industry will increase or decrease in fixed proportions. If

[64] Miller, Ronald E. and Peter D. Blair. 2009. Input-Output Analysis: Foundations and Extensions (2nd Edition). Cambridge University Press.

we collect the values a_{ij} into a matrix A, we can replace Zk in the above matrix equation with Ax, yielding the following:

$$x = Ax + d$$

The matrix A is also known as the direct requirements matrix. Under the assumption that the a_{ij} are given and fixed, we can solve this matrix equation for x, which provides the amount of production required to satisfy some given final demand d. We have:

$$\begin{aligned} x = Ax + d \quad &\iff \quad x - Ax = d \\ &\iff \quad (I - A)x = d \\ &\iff \quad x = (I - A)^{-1}d \end{aligned}$$

where I is the $n \times n$ identity matrix. We see that solving for x requires computation of the "Leontief inverse," $(I - A)^{-1}$ (and the inverse must exist). The matrix $(I - A)^{-1}$ is also known as the total requirements matrix (Miller and Blair, 2009).

A.2 Deriving the Leontief Inverse in the Analytical Model

As in section 1.2 of Chapter 1, the direct requirements matrix A is:

$$A \quad = \begin{bmatrix} f_{11} & (\frac{p_1}{p_2})f_{12} & (\frac{p_1}{p_3})f_{13} \\ (\frac{p_2}{p_1})f_{21} & f_{22} & (\frac{p_2}{p_3})f_{23} \\ (\frac{p_3}{p_1})f_{31} & (\frac{p_3}{p_2})f_{32} & f_{33} \end{bmatrix}$$

where industry i demands quantity f_{ji} of industry j's output to produce one unit of its own output (under given prices), and p_i is industry i's output price (in dollars).

In this section, we find the Leontief inverse, given by $(I - A)^{-1}$, where I is the 3×3 identity matrix. First we have:

$$I - A = \begin{bmatrix} 1 - f_{11} & -(\frac{p_1}{p_2})f_{12} & -(\frac{p_1}{p_3})f_{13} \\ -(\frac{p_2}{p_1})f_{21} & 1 - f_{22} & (\frac{p_2}{p_3})f_{23} \\ -(\frac{p_3}{p_1})f_{31} & -(\frac{p_3}{p_2})f_{32} & 1 - f_{33} \end{bmatrix}$$

Note that when there are no industry self-dependencies (that is, when $f_{11} = f_{22} = f_{33} = 0$), the matrix $(I - A)$ will have ones on the main diagonal. The inverse of $(I - A)$ is given by:

$$(I - A)^{-1} = \frac{1}{det(I - A)} \cdot [adjoint(I - A)]$$

where $det(I - A)$ is the determinant of $(I - A)$ and $adjoint(I - A)$ is the adjoint of $(I - A)$.

The adjoint is itself constructed as the transpose of the matrix of cofactors. Working through the algebra and matrix math, we have:

$$\begin{aligned} det(I - A) = &(1 - f_{11})(1 - f_{22})(1 - f_{33}) - f_{12}f_{21}(1 - f_{33}) \\ &- f_{23}f_{32}(1 - f_{11}) - f_{13}f_{31}(1 - f_{22}) - f_{12}f_{23}f_{31} - f_{32}f_{21}f_{13} \end{aligned}$$

In the case that there are no industry self-dependencies, we have:

$$det(I - A) = 1 - f_{12}f_{21} - f_{23}f_{32} - f_{13}f_{31} - f_{12}f_{23}f_{31} - f_{32}f_{21}f_{13}$$

Note that if there are no cycles in the graph of the production function—such as industry 1 depends on industry 2, which depends on industry 1; or industry 1 depends on industry 2, which depends on industry 3, which depends on industry 1—then the determinant equals one and the inverse $(I - A)^{-1}$ will exist. The adjoint matrix is:

$$adjoint(I-A) = \begin{bmatrix} (1 - f_{22})(1 - f_{33}) - f_{23}f_{32} & (\frac{p_1}{p_2})[f_{13}f_{32} + f_{12}(1 - f_{33}) & (\frac{p_1}{p_3})[f_{12}f_{23} + f_{13}(1 - f_{22}) \\ (\frac{p_2}{p_1})[f_{23}f_{31} + f_{21}(1 - f_{33}) & (1 - f_{11})(1 - f_{33}) - f_{13}f_{31} & (\frac{p_2}{p_3})[f_{13}f_{21} + f_{23}(1 - f_{11})] \\ (\frac{p_3}{p_1})[f_{21}f_{32} + f_{31}(1 - f_{22})] & (\frac{p_3}{p_2})[f_{12}f_{31} + f_{32}(1 - f_{11})] & (1 - f_{11})(1 - f_{22}) - f_{12}f_{21} \end{bmatrix}$$

In the case that there are no industry self-dependencies, we have:

$$adjoint(I - A) = \begin{bmatrix} 1 - f_{23}f_{32} & (\frac{p_1}{p_2})[f_{13}f_{32} + f_{12}] & (\frac{p_1}{p_3})[f_{12}f_{23} + f_{13}] \\ (\frac{p_2}{p_1})[f_{23}f_{31} + f_{21}] & 1 - f_{13}f_{31} & (\frac{p_2}{p_3})[f_{13}f_{21} + f_{23}] \\ (\frac{p_3}{p_1})[f_{21}f_{32} + f_{31}] & (\frac{p_3}{p_2})[f_{12}f_{31} + f_{32}] & 1 - f_{12}f_{21} \end{bmatrix}$$

Finally, as in standard input-output analysis, the vector of industry output (in dollars) is:

$$x = (I - A)^{-1} \cdot d = \frac{1}{det(I - A)} \cdot adjoint(I - A) \cdot d$$

where d is the vector of final demand (in dollars). See section 1.2 of Chapter 1 for the results.

A.3 Constant Returns to Scale of Production Processes

I present here an informal discussion of how constant returns to scale for all production nodes implies constant returns to scale for the overall production process.

First, constant returns to scale for a single production node α implies that $f_\alpha(kx_1, kx_2, ..., kx_n) = k \cdot f_\alpha(x_1, x_2, ..., x_n)$ for all $(x_1, x_2, ..., x_n)$ and $k > 0$. We can see that constant returns to scale for all production nodes implies constant returns to scale for the entire production process by considering three types of production nodes.

The first are production nodes that use only "primary" inputs. These are constant-returns-to-scale directly by the definition above.

The second are production nodes that use some combination of primary inputs and the outputs of other production nodes, where those latter nodes use only primary inputs. As an example, consider a production node β that uses l primary inputs and m inputs that are sourced from other production nodes: $f_\beta(x_1, x_2, ..., x_l, f_1(\cdot), f_2(\cdot), ..., f_m(\cdot))$. As the functions f_1 through f_m depend only on primary inputs, they exhibit constant returns to scale per above. In turn, if f_β has constant returns to scale in relation to the inputs it uses directly, then it also has constant returns to scale with regards to the primary inputs.

Lastly, there are production nodes that either: (1) use no primary inputs directly; or (2) use some primary inputs as well as the output of other production nodes, but those latter nodes may have non-primary inputs. As an example of the first case, take the production node γ, which produces its output using the outputs of n other production nodes with production functions f_i, $i \in \{1, 2, ..., n\}$: $f_\gamma(f_1(\cdot), f_2(\cdot), ..., f_n(\cdot))$. If the production functions $f_i(\cdot)$ have constant returns to scale with regards to the primary inputs per above, then multiplying the primary inputs by k will yield $f_\gamma(k \cdot f_1(\cdot), k \cdot f_2(\cdot), ..., k \cdot f_n(\cdot))$. In turn, if f_γ has constant returns to scale in relation to the inputs it directly uses, then it also has constant returns to scale with regards to the primary inputs.

Beginning from the start of the production process—where production nodes only use primary inputs—and moving forward through the process leads us to conclude that the overall production process has constant returns to scale if all of its production nodes have constant returns to scale.

A.4 Additional Details about the Computational Model

Step 1: Economic Configuration and Production Processes

Economic Configuration. The economic configuration is a direct result of which industries use which others as suppliers. As described in the main text, I assign suppliers for each industry—which represents customer assignment for the other industries—according to the steps outlined in Algorithm 1. The synthetic datasets that I produce for the chapter are based on the algorithm when using the parameter values shown in Table A.1.

Table A.1: Parameter Values for Industry Supplier/Customer Assignment Algorithm.

Parameter	Value
$shape\_param$	1
$threshold\_range_{(1)\text{ to }(n)}$	(1): 0; (2): 2; (3): 6; (4): 15
$threshold\_val_{(1)\text{ to }(n)}$	(1): 0.9; (2): 0.7; (3): 0.1; (4) 0.01
avg_inputs	5
var_inputs	1
min_inputs	2
max_inputs	8
$candidate\_sep$	1
$candidate\_var$	2
$num\_industries$	14

The algorithm essentially has two parts. The first determines each industry's centrality, while the second performs the actual supplier/customer assignment.

In the first part, I take a draw for each industry i from a Pareto Type II distribution (also known as the Lomax distribution). This value represents the centrality of industry i in terms of the relative number of customers it will supply. I use this distribution because it is a heavy-tailed distribution with support beginning at zero (i.e., it is a shifted Pareto distribution), which, over multiple draws, will result in a larger number of industries that are less central and a smaller number of industries that are more central. I use this centrality draw to set a threshold value for each industry (to be used below): the higher the centrality, the lower the threshold.

I then assign suppliers to each industry i. To determine i's total number of suppliers, I take a draw from a normal distribution. As described in the main text, I always assign the industry *shocked* to be one of i's suppliers. I then take a random draw from a normal distribution centered around $(i - candidate\_sep)$, where $candidate\_sep$ is a parameter provided to the algorithm. This random draw represents a candidate supplier for i. I center the distribution around $(i - candidate\_sep)$ so that i tends to receive its inputs from industries that came before it, which results in a network structure with some upstream, some midstream, and some downstream industries. If the candidate is not yet one of i's suppliers, I take a draw from a uniform distribution and compare it to the candidate's previously determined threshold value. If the threshold is exceeded, the candidate is assigned to be one of i's suppliers. I repeat this process until all of i's suppliers have been assigned, and then move on to assigning suppliers for the next industry in the economy.

To summarize, the larger an industry's draw from the Pareto distribution, the lower its threshold, and in turn, the more likely it is that the industry will be assigned as a supplier when randomly chosen during another industry's assignment process. As mentioned in the main text, this algorithm attempts to produce a network structure with a normal distribution in the number of suppliers and a skewed distribution in the number of customers.

Production Processes. As described in the main text, I use two types of production nodes when constructing industry production processes. The first is a Cobb-Douglas formulation:

$$f_{CD}(x_1, x_2) = A \cdot x_1^{a_1} \cdot x_2^{a_2}$$

where $a_1 + a_2 = 1$ so that the production node exhibits constant returns to scale. The second is a Leontief formulation:

$$f_L(x_1, x_2) = B \cdot \min[b_1 \cdot x_1, b_2 \cdot x_2]$$

where $b_1 > 0$ and $b_2 > 0$.

For the results in the chapter, I take: (1) A and B to be randomly drawn from the uniform distribution over [2.0, 5.0] (in increments of tenths); (2) a_1 to be randomly drawn from the uniform distribution over [0.40, 0.60] (in increments of hundredths) and $a_2 = 1 - a_1$; and (3) b_1 and b_2 to be randomly drawn from the uniform distribution over [0.1,10.0] (in increments of tenths). This provides reasonable variation in the specific forms of production nodes within each production process, while limiting the differences in the marginal products of x_1 and x_2 within each node. I note that it may be possible to estimate some of these distributions from empirical data, though it would likely require observations at the level of individual production steps within a production process.

I also describe in the main text the algorithm that solves for the input quantities required by a production process given a set of input prices. Part of this process involves iterating over trial input values at each node. Given that reasonable input values for each node will be dependent on its specific parameters, I propose a process that generates the same number of trial values for every node but allows the range of those values to vary. In that proposed process, the extremities of the ranges are determined by the ratio of a node's inputs when making one input much smaller (or larger) than the other while holding output constant.

Take a node α with production function $f_\alpha(x_1, x_2)$. For the results in the chapter, I calculate the smallest trial value for node α to be the value of x_1 such that the ratio $\frac{x_2}{x_1}$ equals 20 when output is one unit. I take the largest trial value to be the value of x_1 such that the ratio $\frac{x_1}{x_2}$ equals 20 when output is one unit. I divide the range between the smallest trial value and the largest trial value into 2,000 equal-sized intervals, and the endpoints of these intervals (including the smallest and largest values) are the full set of trial values.

I choose a ratio of 20 to allow for a substantial difference in the quantities of the two inputs that each production node uses. I create 2,000 trial values for each node so that there is enough granularity to capture changes in the usage of input quantities in response to varying input prices.

Step 2: Final Demand

The initial final demand for each industry, d_i, is drawn from the uniform distribution over [0.08, 0.12] in increments of thousandths. I choose this range so that there is some, but not too much, variation in final demand for industries' outputs (a difference of 50 percent between the smallest and largest possible values). After the shocks are applied in each price round, final demand for each industry is adjusted based on the formulas provided in section 1.4.3 in the main text.

Step 3(a): Initial Price Vector

I determine the initial (pre-shock) price vector in each price round by using the steps outlined in Algorithm 2. The synthetic datasets that I produce for the chapter are based on the algorithm when using the parameter values shown in Table A.2.

Table A.2: Parameter Values for Creation of Initial Price Vector Algorithm.

Parameter	Value
p_{min}	0.95
p_{max}	1.05
max_rounds	15
$pass\_through\_rate$	0.3
$markup$	0.01
$change\_threshold$	0.02

Essentially, the algorithm first creates a random price vector, where each element is a random draw from the uniform distribution over [0.95, 1.05] (in increments of hundredths). Industries then increase their output prices (if needed) so that they have positive profits, including a pre-specified one percent markup. This increases input prices for the other industries, which (after adjusting their input usage), pass through 30 percent of the overall change in their total input cost. This process repeats until the round-to-round change in all prices is below 2 percent or until 15 pass-through rounds have been completed, whichever comes first. After verifying that all industry profits are positive, the resulting price vector is taken to be the baseline, pre-shock set of prices.

Step 3(c): Shock Propagation

The upstream shock propagates through the network based on the steps outlined in Algorithm 3. The synthetic datasets that I produce for the chapter are based on the algorithm when using the parameter values shown in Table A.3. The algorithm is quite similar to the algorithm that produces the initial price vector, with two differences: (1) instead of being created, the price vector is now given; and (2) I do not require that industries' profits remain positive.

Table A.3: Parameter Values for Shock Propagation Algorithm.

Parameter	Value
max_rounds	15
pass_through_rate	0.3
change_threshold	0.02

A.5 Additional Details about Indicator Construction

I describe here the details of the construction of the matrices P, R, and C, which are used to build the indicators generated by the computational variant of the model.

As in the analytical variant of the model, let A (equation 2) be the direct requirements table, where an element a_{ji} of A is the dollar amount of industry j's output required to produce one dollar's worth of industry i's output. Also as before, let T (equation 1) be the non-price-adjusted "quantity" requirements table, such that an element t_{ji} of T is the quantity of industry j's output required to produce one unit of industry i's output.

A note regarding the number of walks, edge weights, and the Leontief inverse: if we take A as the input-output matrix, the ji-th element of A^n measures the weighted sums of all walks of length n from industry j to industry i, where the weights are given by the values in the input-output matrix (which correspond to the edge weights in the network representation of the economy). As an example, the sum $A + A^2 + A^3$ measures the weighted sums of all walks of length one, two, and three from each industry to every other industry. The Leontief inverse $(I - A)^{-1} = I + A + A^2 + ...$ measures the total direct and indirect exposures of all lengths for each industry through the production network.

I construct the matrix P, which is based on the matrix A, in three steps. First, to retain only those inter-industry dependencies with more than a minimal amount of economic relevance, I take the matrix A and set any elements $a_{ji} < 0.05$ equal to zero. Call the resulting matrix $A_{\geq 0.05}$. Second, I form the matrix A_{bin} by converting all non-zero values in the matrix $A_{\geq 0.05}$ to ones (while all other values remain as zeros). Finally, I create the matrix P as:

$$P = A_{bin} + (A_{bin})^2 + (A_{bin})^3$$

Each element p_{ji} of P then represents the number of walks—of lengths one, two, and three—from supplying industry j to consuming industry i in the production network.

I construct a row-normalized matrix R, which is based on T, by calculating each element r_{ji} as follows:

$$r_{ji} = t_{ji} / \left(\sum_{k=1}^{n} t_{jk} \right) \cdot 0.95$$

where n is the number of industries. Each element r_{ji} represents the relative importance of industry j as one of industry i's customers (from i's perspective), based on the quantities of industry i's output purchased by all of i's customers (when producing one unit of their own output). I scale each element down by five percent to limit the magnitude of the indicators constructed from R.

Finally, I construct a column-normalized matrix C, which is based on A, by calculating each element c_{ji} as follows:

$$c_{ji} = a_{ji} / \left(\sum_{k=1}^{n} a_{ki} \right) \cdot 0.95$$

where n is again the number of industries. Each element c_{ji} represents the relative importance of industry j as one of industry i's suppliers (from i's perspective), based on i's expenditures on all of its suppliers' outputs. As with the matrix R, I scale each element down by five percent to limit the magnitude of the indicators constructed from C.

A.6 Alternative Specifications for Indicator Assessment

I present below the profit and revenue regression results with and without the controls introduced in section 1.5.3. (Tables A.4 and A.5). These results are based on the same synthetic dataset as used for the results in the main text.

The coefficients on the Leontief and walks indicators have slightly different magnitudes when including the controls, though the results are all qualitatively similar. One exception is the coefficient for supply_walks in the profit regressions, which becomes statistically insignificant once the controls have been added.

The controls have more of an effect on the evolutionary indicators. Specifically, in the profit regressions that use the walks indicators, controlling for a substantial direct dependence on the shocked industry reverses the sign of the coefficient on up_comp. The reason that the coefficient is initially positive (without the control) is because the average industry sees its profits increase as a result of price pass-through. Given this, and in light of the predictions of the analytical model, we would expect industries with a greater number/magnitude of upstream substitute paths to have even higher (positive) outcomes than those with a larger number/magnitude of upstream complement paths. This is what the results without the controls indicate: although the coefficients on up_subs and up_comp are both positive, the former is larger in magnitude than the latter.

The addition of the controls has much less of an effect on the up_comp coefficient in the profit regressions that include the Leontief indicators. The likely reason for this is, by its construction, the upstream Leontief indicator already includes the magnitude of each industry's direct dependence on the shocked industry, in the sense that this value is part of the input-output table from which the Leontief inverse is calculated. In turn, introducing the controls—and specifically, the proxy for direct dependence on the shocked industry—provides a limited amount of new information.

I note that introducing the controls has a much smaller effect on the coefficients for up_subs in the profit regressions, and the proxy for demand elasticity is not statistically significant in either of the profit regressions in which it appears.

In the revenue regressions, both of the controls are statistically significant in the regressions in which they appear. Analogous to above, introducing the controls does not substantially change the coefficients on down_subs, but their inclusion does reverse the sign of the coefficients on down_comp. Also similar to above, this is best explained as follows: without the controls, the average industry sees its revenues increase as a result of price pass-through. From this perspective, we would expect the coefficients on both down_subs and down_comp to be positive, though the former should (given the logic of the analytical model) be larger in magnitude than the latter. This is what the results without the controls suggest. By including the controls—and specifically, through the inclusion of the proxy for demand elasticity—the observations are split in such a way that the average industry is predicted to experience not just relatively better outcomes, but better outcomes in absolute terms (compared to a reference point of zero) when associated with a greater number/magnitude of downstream substitute paths (as compared to downstream complement paths).

Unlike the profit regressions, in the revenue regressions, the down_comp sign reversal occurs when using either the walks or Leontief indicators. The reason is that neither the Leontief indicators nor the walks indicators capture anything about the elasticity of demand for an industry's output, and in this way, the control that proxies for elasticity of demand is adding new information in both cases.

Lastly, in the revenue regression using the walks indicators without the controls, the coefficient on up_comp is positive. Its magnitude is quite similar as in the corresponding profit regression, and as above, its positive value is explained as the result of greater pass-through. It becomes statistically insignificant with the inclusion of the controls.

Table A.4: Profit regression results with and without controls.

	(1)	(2)	(3)	(4)
log(up_leontief_inv)	-0.519***	-0.484***		
	(0.0232)	(0.0225)		
log(down_leontief_inv)	-0.0314***	-0.0341***		
	(0.00972)	(0.00977)		
log(supply_walks)			-0.102***	0.0101
			(0.00885)	(0.0100)
log(demand_walks)			-0.0462***	-0.0535***
			(0.0133)	(0.0128)
log(up_subs)	0.0593***	0.0585***	0.111***	0.0848***
	(0.0203)	(0.0202)	(0.0233)	(0.0219)
log(up_comp)	-0.132***	-0.149***	0.0539***	-0.119***
	(0.0103)	(0.0120)	(0.0128)	(0.0132)
log(down_subs)	-0.00765	-0.00107	-0.00348	0.00135
	(0.00929)	(0.0102)	(0.00945)	(0.0100)
log(down_comp)	0.00318	0.0395	0.00360	0.0276
	(0.00948)	(0.0351)	(0.0103)	(0.0377)
direct > average		-0.117***		-0.751***
		(0.0288)		(0.0439)
zero down_comp		0.0869		0.0548
		(0.0829)		(0.0899)
log(up_subs) - log(up_comp)	0.191***	0.207***	0.0573***	0.204***
	(0.0182)	(0.0196)	(0.0174)	(0.0212)
log(down_subs) - log(down_comp)	-0.0108	-0.0406	-0.00708	-0.0263
	(0.0113)	(0.0314)	(0.0127)	(0.0340)
N	8,400	8,400	8,400	8,400
adj. R^2	0.197	0.199	0.023	0.101

Standard errors clustered by configuration round-price round.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.5: Revenue regression results with and without controls.

	(1)	(2)	(3)	(4)
log(up_leontief_inv)	-0.160***	-0.133***		
	(0.0135)	(0.0168)		
log(down_leontief_inv)	-0.0613***	-0.0531***		
	(0.0159)	(0.0159)		
log(supply_walks)			-0.0686***	-0.0320**
			(0.0137)	(0.0147)
log(demand_walks)			-0.0944***	-0.0932***
			(0.0128)	(0.0128)
log(up_subs)	-0.0000362	0.000434	0.0145	0.00708
	(0.0124)	(0.0123)	(0.0128)	(0.0125)
log(up_comp)	0.0115	-0.000712	0.0599***	0.00524
	(0.0122)	(0.0133)	(0.0118)	(0.0134)
log(down_subs)	0.331***	0.294***	0.323***	0.287***
	(0.0161)	(0.0174)	(0.0156)	(0.0167)
log(down_comp)	0.102***	-0.136***	0.0936***	-0.150***
	(0.0128)	(0.0352)	(0.0132)	(0.0352)
direct > average		-0.0866**		-0.239***
		(0.0345)		(0.0296)
zero down_comp		-0.578***		-0.598***
		(0.0820)		(0.0819)
log(up_subs) -	-0.0115	0.00115	-0.0453***	0.00185
log(up_comp)	(0.0141)	(0.0150)	(0.0138)	(0.0151)
log(down_subs) -	0.229***	0.430***	0.229***	0.437***
log(down_comp)	(0.0178)	(0.0312)	(0.0177)	(0.0312)
N	8,400	8,400	8,400	8,400
adj. R^2	0.109	0.114	0.100	0.112

Standard errors clustered by configuration round-price round.

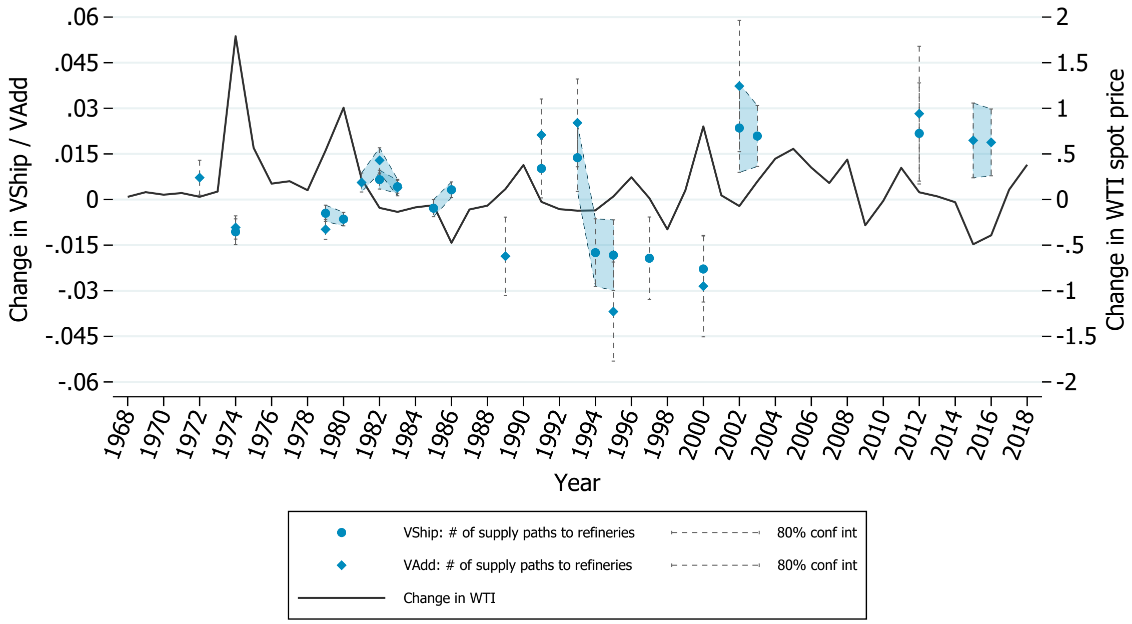
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

B Appendix for Chapter 2

B.1 Oil Episodes Context

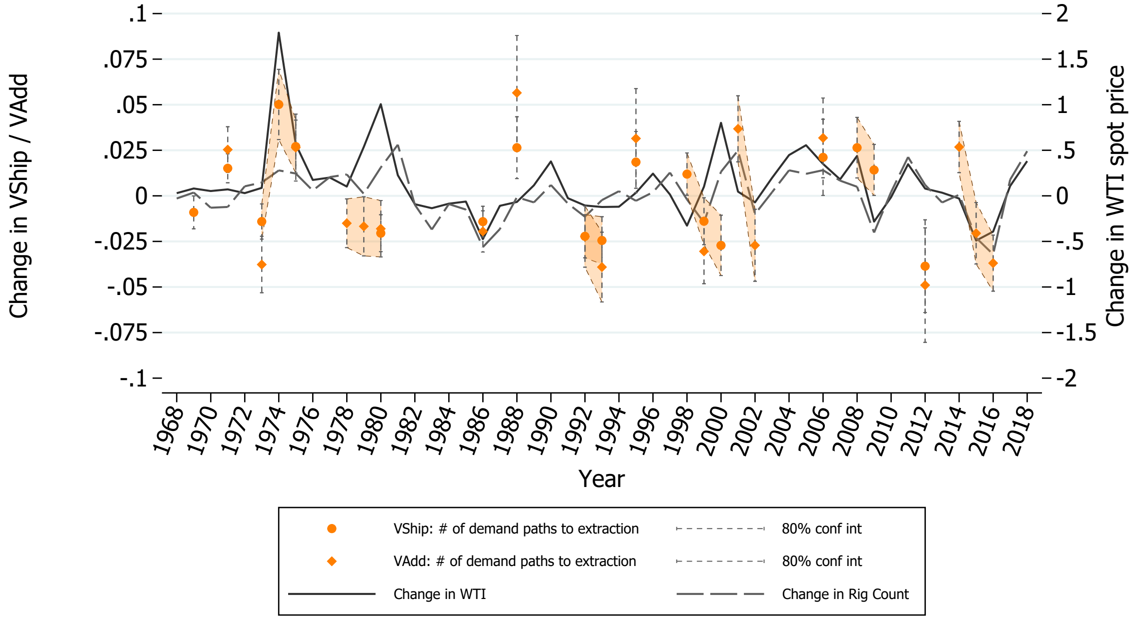
In the results in the main text, I only show points that are significant at the 95 percent level. For comparison purposes, here I plot points that are significant at the 80 percent level (Figures B.1-B.7).

Figure B.1: Indicator Results: Supply Paths to Petroleum Refineries.



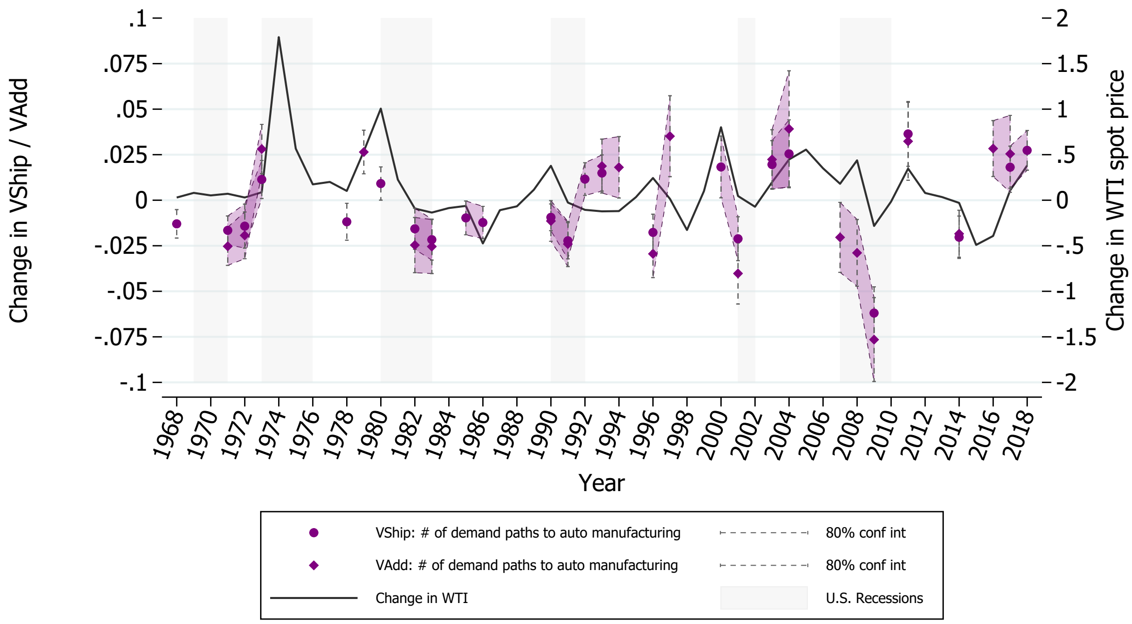
Note: the figure shows the change in value of shipments / value added based on linear combinations of $\beta_{refining} + \beta_{refining,t}$ for $t \in [1968, 2018]$ moving from the 25th percentile to the 75th percentile. Only estimates significant at the 80 percent level are pictured. Successive points are connected by confidence areas.

Figure B.2: Indicator Results: Demand Paths to Oil and Gas Extraction.



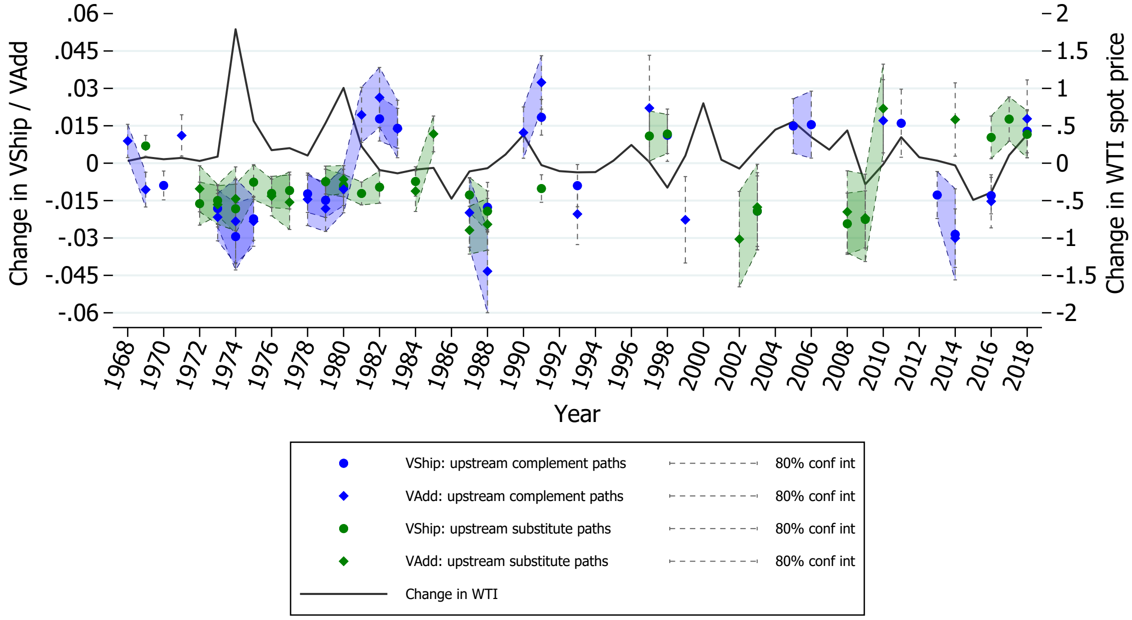
Note: the figure shows the change in value of shipments / value added based on linear combinations of $\beta_{extract} + \beta_{extract,t}$ for $t \in [1968, 2018]$ moving from the smallest value greater than zero to the 90th percentile. Only estimates significant at the 80 percent level are pictured. Successive points are connected by confidence areas.

Figure B.3: Indicator Results: Demand Paths to Auto Manufacturing.



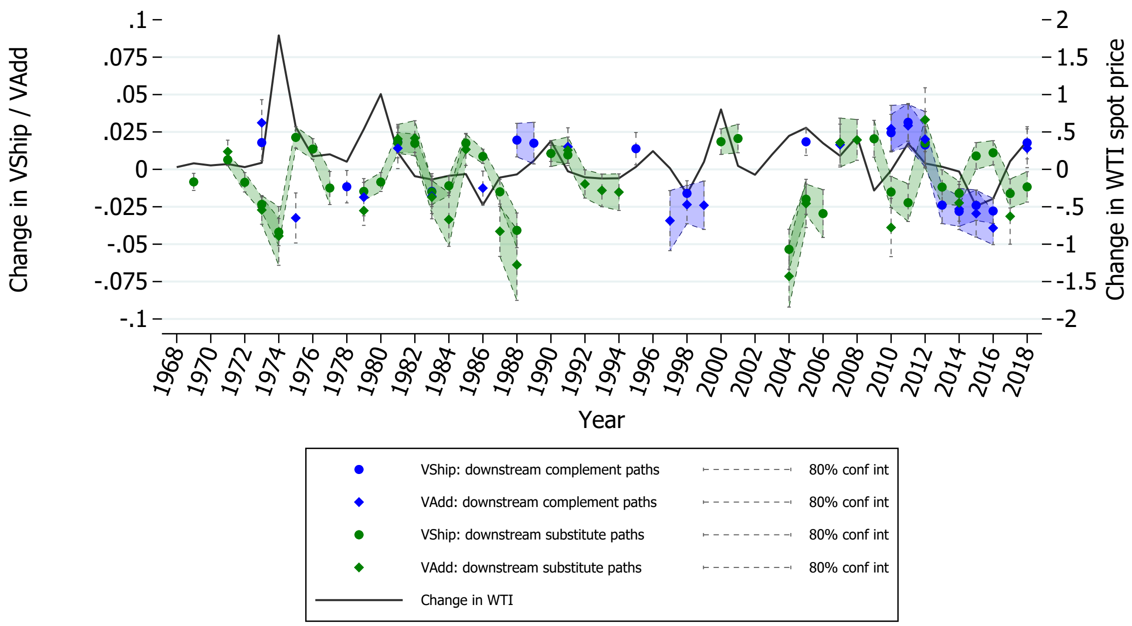
Note: the figure shows the change in value of shipments / value added based on linear combinations of $\beta_{auto} + \beta_{auto,t}$ for $t \in [1968, 2018]$ moving from the smallest value greater than zero to the 75th percentile. Only estimates significant at the 80 percent level are pictured. Successive points are connected by confidence areas.

Figure B.4: Indicator Results: Upstream Chains of Substitutability and Complementarity.



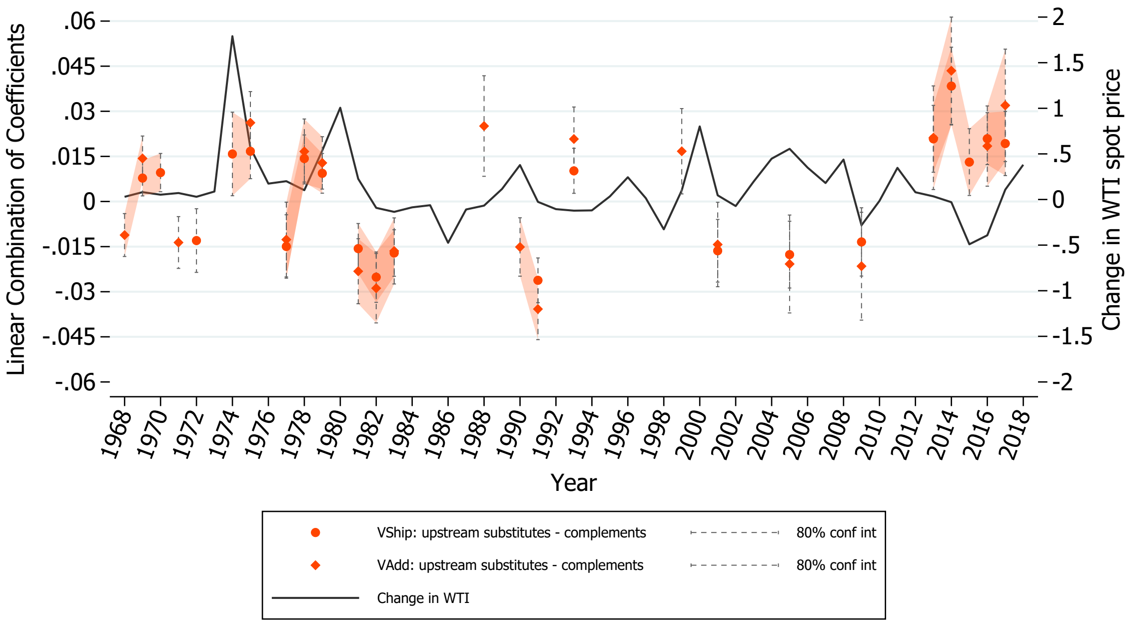
Note: the figure shows the change in value of shipments / value added based on linear combinations of $\beta_{ucomp} + \beta_{ucomp,t}$ and $\beta_{usub} + \beta_{usub,t}$ for $t \in [1968, 2018]$ moving from the 25th percentile to the 75th percentile. Only estimates significant at the 80 percent level are pictured. Successive points are connected by confidence areas.

Figure B.5: Indicator Results: Downstream Chains of Substitutability and Complementarity.



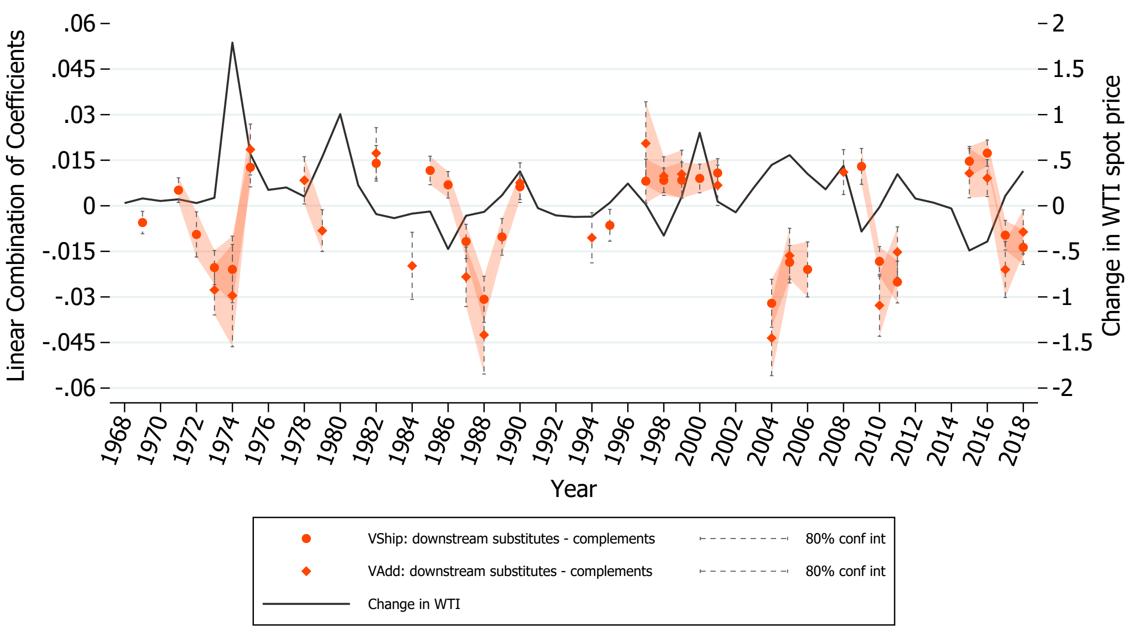
Note: the figure shows the change in value of shipments / value added based on linear combinations of $\beta_{dcomp} + \beta_{dcomp,t}$ and $\beta_{dsub} + \beta_{dsub,t}$ for $t \in [1968, 2018]$ moving from the 25th percentile to the 75th percentile. Only estimates significant at the 80 percent level are pictured. Successive points are connected by confidence areas.

Figure B.6: Difference between Upstream Substitutability and Complementarity Indicators.



Note: the figure shows linear combinations of $\beta_{usub} + \beta_{usub,t} - \beta_{ucomp} - \beta_{ucomp,t}$ for $t \in [1968, 2018]$. Only estimates significant at the 80 percent level are pictured. Successive points are connected by confidence areas.

Figure B.7: Difference between Downstream Substitutability and Complementarity Indicators.

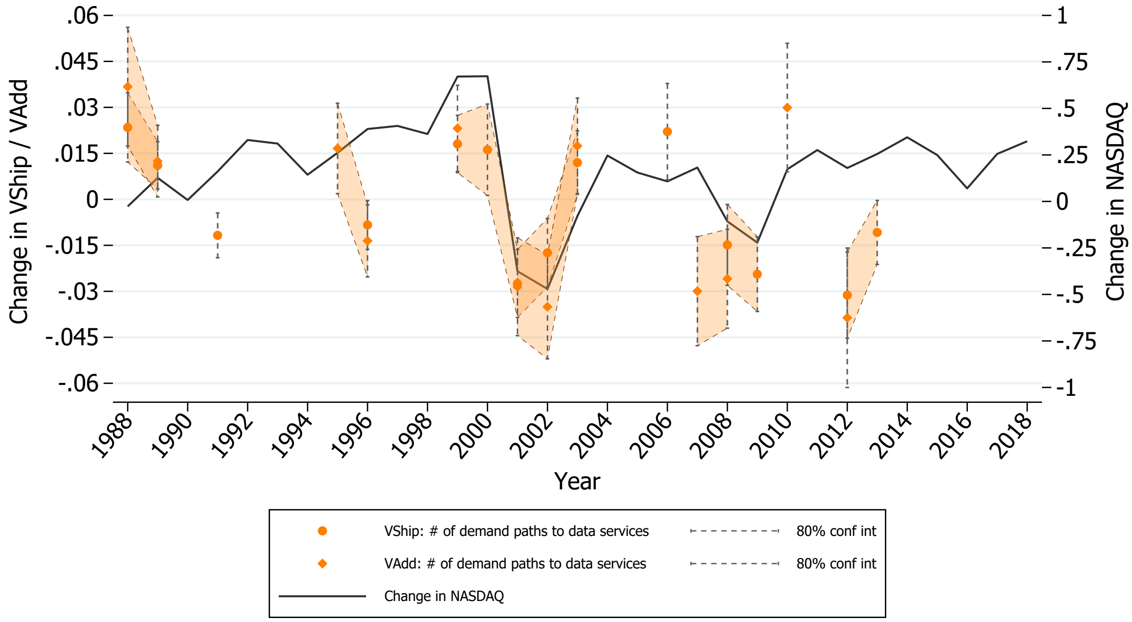


Note: the figure shows linear combinations of $\beta_{dsub} + \beta_{dsub,t} - \beta_{dcomp} - \beta_{dcomp,t}$ for $t \in [1968, 2018]$. Only estimates significant at the 80 percent level are pictured. Successive points are connected by confidence areas.

B.2 Information Technology Context

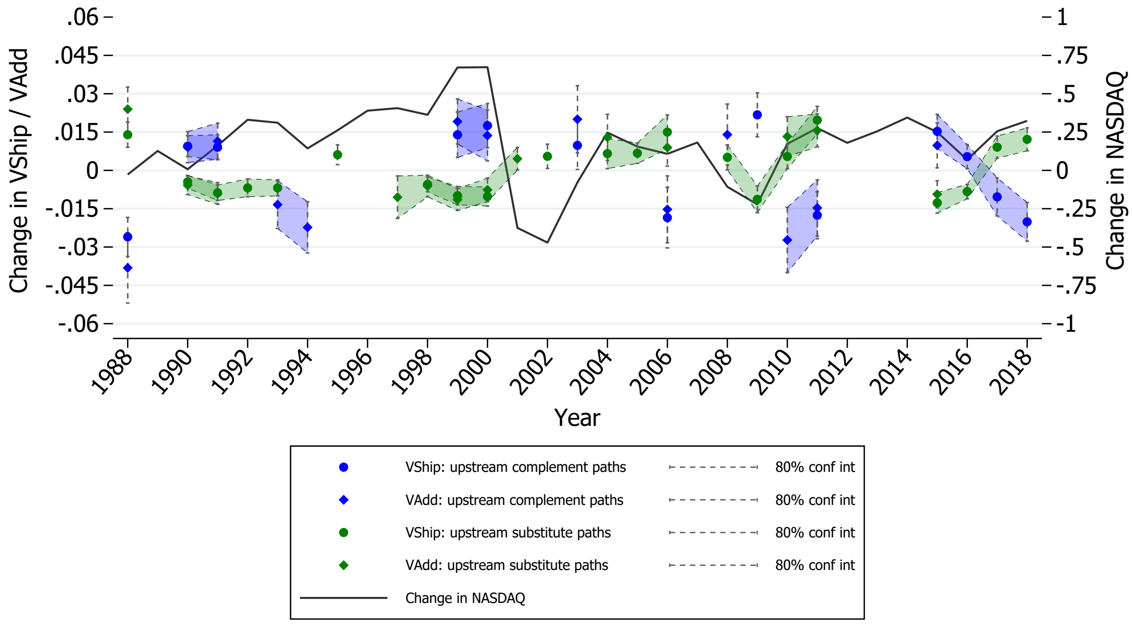
In the results in the main text, I only show points that are significant at the 95 percent level. For comparison purposes, here I plot points that are significant at the 80 percent level (Figures B.8-B.12).

Figure B.8: Indicator Results: Demand Paths to Data Processing Services.



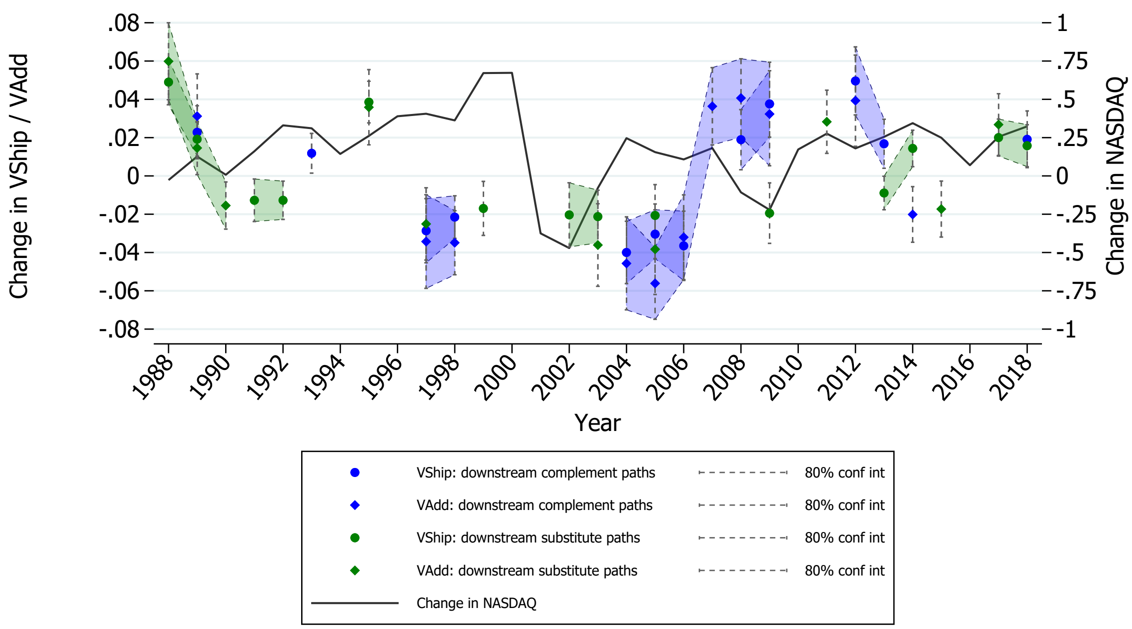
Note: the figure shows the change in value of shipments / value added based on linear combinations of $\beta_{data} + \beta_{data,t}$ for $t \in [1988, 2018]$ moving from the smallest value greater than zero to the 90th percentile. Only estimates significant at the 80 percent level are pictured. Successive points are connected by confidence areas.

Figure B.9: Indicator Results: Upstream Chains of Substitutability and Complementarity.



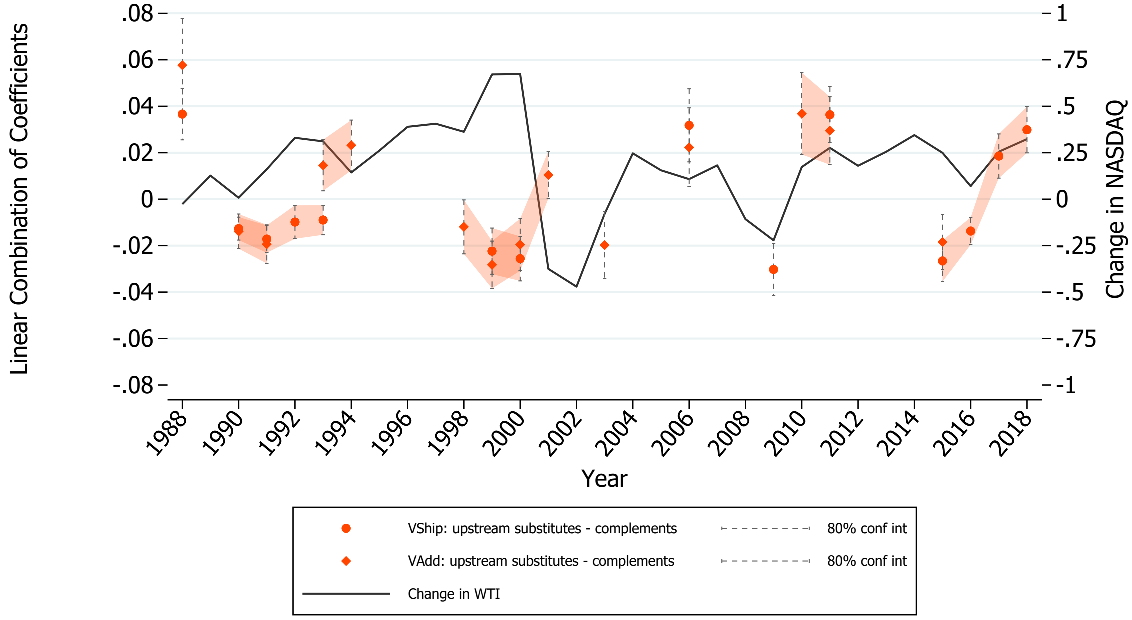
Note: the figure shows the change in value of shipments / value added based on linear combinations of $\beta_{ucomp} + \beta_{ucomp,t}$ and $\beta_{usub} + \beta_{usub,t}$ for $t \in [1988, 2018]$ moving from the 25th percentile to the 75th percentile. Only estimates significant at the 80 percent level are pictured. Successive points are connected by confidence areas.

Figure B.10: Indicator Results: Downstream Chains of Substitutability and Complementarity.



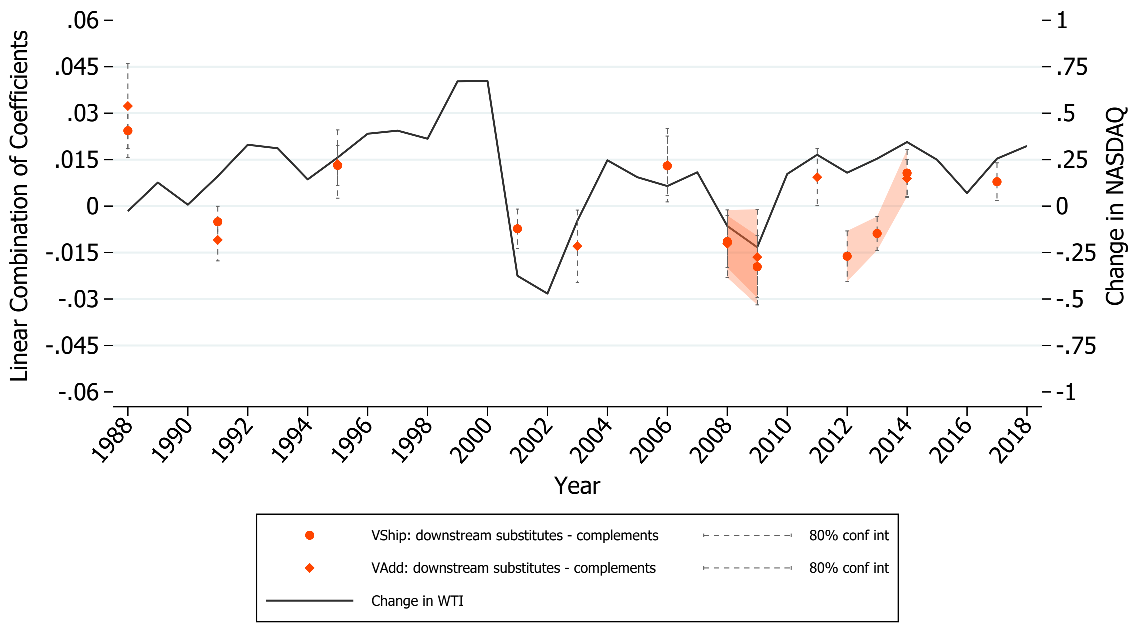
Note: the figure shows the change in value of shipments / value added based on linear combinations of $\beta_{dcomp} + \beta_{dcomp,t}$ and $\beta_{dsub} + \beta_{dsub,t}$ for $t \in [1988, 2018]$) moving from the 25th percentile to the 75th percentile. Only estimates significant at the 80 percent level are pictured. Successive points are connected by confidence areas.

Figure B.11: Difference between Upstream Substitutability and Complementarity Indicators.



Note: the figure shows linear combinations of $\beta_{usub} + \beta_{usub,t} - \beta_{ucomp} - \beta_{ucomp,t}$ for $t \in [1988, 2018]$. Only estimates significant at the 80 percent level are pictured. Successive points are connected by confidence areas.

Figure B.12: Difference between Downstream Substitutability and Complementarity Indicators.



Note: the figure shows linear combinations of $\beta_{dsub} + \beta_{dsub,t} - \beta_{dcomp} - \beta_{dcomp,t}$ for $t \in [1988, 2018]$. Only estimates significant at the 80 percent level are pictured. Successive points are connected by confidence areas.

C Appendix for Chapter 3

C.1 Creation of Oil Regression Scenarios

As described in the main text, I modify the oil episodes regression specification in Chapter 2 by interacting two binary variables, U and D, with the three upstream indicators and with two of the downstream indicators, respectively. This updated specification is:

$$
\begin{aligned}
y_{it} = \alpha &+ Y_t + 2\text{digit}_i \cdot Y_t + \beta_{emp} \cdot \text{emp}_{it} + \beta_{energy} \cdot \text{energy}_{it} + \\
&\beta_{direct} \cdot \log(\text{direct}_{h(i),t-2}) + \beta_{direct,t} \cdot \log(\text{direct}_{h(i),t-2}) \cdot Y_t + \\
&\beta_{refining} \cdot \log(\text{supply\_paths\_refining}_{i,g(t)}) \cdot U + \beta_{refining,t} \cdot \log(\text{supply\_paths\_refining}_{i,g(t)}) \cdot Y_t \cdot U + \\
&\beta_{extract} \cdot \log(\text{demand\_paths\_extract}_{i,g(t)}) + \beta_{extract,t} \cdot \log(\text{demand\_paths\_extract}_{i,g(t)}) \cdot Y_t + \\
&\beta_{auto} \cdot \log(\text{demand\_paths\_auto}_{i,g(t)}) + \beta_{auto,t} \cdot \log(\text{demand\_paths\_auto}_{i,g(t)}) \cdot Y_t + \\
&\beta_{dcomp} \cdot \log(\text{down\_comp\_paths}_{h(i),t-2}) \cdot D + \beta_{dcomp,t} \cdot \log(\text{down\_comp\_paths}_{h(i),t-2}) \cdot Y_t \cdot D + \\
&\beta_{dsub} \cdot \log(\text{down\_sub\_paths}_{h(i),t-2}) \cdot D + \beta_{dsub,t} \cdot \log(\text{down\_sub\_paths}_{h(i),t-2}) \cdot Y_t \cdot D + \\
&\beta_{ucomp} \cdot \log(\text{up\_comp\_paths}_{h(i),t-2}) \cdot U + \beta_{ucomp,t} \cdot \log(\text{up\_comp\_paths}_{h(i),t-2}) \cdot Y_t \cdot U + \\
&\beta_{usub} \cdot \log(\text{up\_sub\_paths}_{h(i),t-2}) \cdot U + \beta_{usub,t} \cdot \log(\text{up\_sub\_paths}_{h(i),t-2}) \cdot Y_t \cdot U + \epsilon_{it}
\end{aligned}
$$

where y_{it} is the percentage change in the value of shipments for 6-digit-NAICS industry i in year t, Y_t is a set of year fixed-effects, 2digit_i is a set of 2-digit-NAICS fixed effects, emp_{it} is the industry-level change in employment, energy_{it} is the industry-level change in energy usage, direct_{it} is the fraction of industry i's inputs purchased from petroleum products in year t, $g(t)$ maps year t to one of {1967, 1987, 1997} (whichever most closely precedes t), $h(i)$ maps each 6-digit-NAICS industry i to its 3-digit-NAICS counterpart, and the indicators are calculated as in section 2.4.1 of Chapter 2.

I initially set the values of U and D to be one, such that running the regression yields the same coefficients as presented in the results in Chapter 2. I use this regression to predict the change in value of shipments for each 6-digit-NAICS industry i and year t; call this value $\hat{y}_{it,11}$.

I then set U to zero and re-predict the values across the industries and years; call these values $\hat{y}_{it,01}$. Similarly, I set D to zero (leaving U equal to one) to generate the predictions $\hat{y}_{it,10}$. Finally, I set both U and D to zero and generate the predictions $\hat{y}_{it,00}$.

I am then able to calculate three differences, each of which represents the marginal effect of, respectively, the upstream network dynamics, the downstream network dynamics, or both:

$$\hat{y}_{it,up} = \hat{y}_{it,11} - \hat{y}_{it,01}$$

$$\hat{y}_{it,down} = \hat{y}_{it,11} - \hat{y}_{it,10}$$

$$\hat{y}_{it,both} = \hat{y}_{it,11} - \hat{y}_{it,00}$$

Given that the machine learning models use GDP predictors at the 3-digit-NAICS level, I take weighted averages of these 6-digit-NAICS values (with weights given by value of shipments) to create 3-digit-NAICS versions that can be used to modify the predictor gdp_change.

The modification sets for each scenario are created by using these aggregated values for selected years from across the Chapter 2 study period. Specifically, the upstream-only modification sets are constructed using the 3-digit-NAICS aggregates of $\hat{y}_{it,up}$ for $t \in \{1974, 1975, 1979, 1988\}$. The downstream-only modification sets are constructed using the 3-digit-NAICS aggregates of $\hat{y}_{it,down}$ for $t \in \{1975, 1985, 1990, 1997\}$. The combination upstream/downstream modification set is constructed using the 3-digit-NAICS aggregates of $\hat{y}_{it,both}$ for $t = 1975$.

C.2 Percentage Changes for Downstream-Only/Combination Scenarios

I present in the main text (Table 3.7) the percentage changes applied to the predictor gdp_change to construct the modification sets for the upstream-only oil regression scenario. I present below an analogous table that shows the values for the downstream-only and combination upstream/downstream oil scenarios (Table C.1).

C.3 Additional Results

I include below Tables C.2-C.6 and Figure C.1, which present (1) the overall occupational rankings (and groupings), (2) the exchange occupations with the highest contributions (positive, negative, and in absolute value), (3) results of the regression of occupational rankings against the contribution measures of individual exchange occupations, and (4) the relationship between the occupational rankings and the contribution measures for green/brown exchange occupations.

Table C.1: Percent Changes Based on Downstream-Only and Upstream/Downstream Indicators in Oil Regression Results.

U.S. BEA 3-digit-NAICS Industry		% Change (Downstream-Only)				% Change (Upstream/Downstream)
Name	Code	1975	1985	1990	1997	1975
Food and beverage and tobacco products manufacturing	311-312	-4.31%	-4.61%	-4.84%	-5.34%	-4.50%
Textile mills and textile product mills	313-314	-3.64%	-4.62%	-3.95%	-7.55%	-6.65%
Apparel, leather, and allied product manufacturing	315-316	-2.51%	-2.97%	-2.33%	-3.57%	-6.19%
Wood product manufacturing	321	-4.70%	-5.59%	-5.32%	-9.32%	-5.21%
Paper manufacturing	322	-1.94%	-2.94%	-2.09%	-6.28%	-3.71%
Printing and related support activities	323	-3.36%	-3.99%	-2.91%	-5.35%	-3.75%
Petroleum and coal products manufacturing	324	-10.00%	-10.00%	-10.00%	-10.00%	-10.00%
Chemical manufacturing	325	-3.51%	-4.40%	-3.90%	-7.67%	-4.74%
Plastics and rubber products manufacturing	326	-2.35%	-3.61%	-3.09%	-6.20%	-6.23%
Nonmetallic mineral product manufacturing	327	-1.71%	-2.49%	-2.18%	-5.10%	-2.67%
Primary metal manufacturing	331	-10.00%	-10.00%	-10.00%	-10.00%	-10.00%
Fabricated metal product manufacturing	332	-0.94%	-2.29%	-1.65%	-5.48%	-1.91%
Machinery manufacturing	333	-1.94%	-2.51%	-2.05%	-5.35%	-1.73%
Computer and electronic product manufacturing	334	-1.79%	-2.98%	-3.10%	-4.20%	0.00%
Electrical equipment, appliance, and component manufacturing	335	-4.05%	-4.21%	-4.23%	-4.31%	-6.33%
Furniture and related product manufacturing	337	0.00%	0.00%	0.00%	0.00%	-1.86%
Miscellaneous manufacturing	339	-3.77%	-4.88%	-5.20%	-6.76%	-6.31%

Note: this table shows the percentage changes for the modification sets corresponding to the downstream-only oil scenario and the combination upstream/downstream oil scenario. The values have been shifted and scaled within each modification set so that they fall between -10 percent and zero.

Table C.2: Occupational Rankings by Scenario and Clustering Algorithm Group Assignments.

	Scenario Rankings					
Occupation Title	1	2	3	4	5	Group
Wholesale and Retail Buyers, Except Farm Products	1	1	2	4	2	Similar
Industrial and Refractory Machinery Mechanics	2	3	4	24	5	Downstream/Combination
Other Installation, Maintenance, and Repair Workers	3	5	57	80	82	(Own Group)
First-Line Supervisors of Production and Operating Workers	4	4	18	6	12	Upstream/Combination
Bus and Truck Mechanics and Diesel Engine Specialists	5	7	3	2	3	Similar
Roofers	6	2	1	1	1	Similar
Woodworkers including Model Makers and Patternmakers, Other	7	8	8	3	4	Similar
Cabinetmakers and Bench Carpenters	8	6	21	15	20	Upstream/Combination
Railroad Conductors and Yardmasters	9	10	6	7	16	Upstream/Combination
Financial Analysts	10	9	15	16	19	Similar
Law Enforcement Workers, Other	11	13	20	51	28	Downstream/Combination
Computer Control Programmers and Operators	12	11	9	5	6	Similar
Molders and Molding Machine Setters, Operators, and Tenders, Metal and Plastic	13	12	16	10	13	Similar
Construction Workers, Other	14	15	19	17	18	Similar
Packaging and Filling Machine Operators and Tenders	15	18	52	18	29	(Own Group)
Cement Masons, Concrete Finishers, and Terrazzo Workers	16	20	10	22	7	Upstream/Combination
Structural Metal Fabricators and Fitters	17	17	14	11	8	Similar
Cutting, Punching, and Press Machine Setters, Operators, and Tenders, Metal and Plastic	18	14	13	21	10	Similar
Chemical Processing Machine Setters, Operators, and Tenders	19	29	43	46	42.5	Differences (smaller)
Helpers--Installation, Maintenance, and Repair Workers	20	21	24	14	17	Similar
Grinding, Lapping, Polishing, and Buffing Machine Tool Setters, Operators, and Tenders, Metal and Plastic	21	24	46	47	47	Differences (smaller)
Stationary Engineers and Boiler Operators	22	16	11	12	11	Similar
Construction Laborers	23	25	29	26	30	Similar
Aerospace Engineers	24	39	23	43	35	Differences (smallest)
Textile, Apparel, and Furnishings Workers, Other	25	23	26	20	23	Similar
Industrial Engineers, including Health and Safety	26	26	37	31	38	Upstream/Combination
Chemists and Materials Scientists	27	27	40	48	50	Differences (smaller)
Construction and Building Inspectors	28	31	38	35	42.5	Upstream/Combination
Mechanical Engineers	29	28	33	25	25	Similar
Electricians	30	22	65	60	69	Differences (larger)
Sheet Metal Workers, Metal-Working	31	32	34	44	39	Similar
Electrical and Electronics Engineers	32	35	31	30	33	Similar
Architectural and Engineering Managers	33	30	39	37	37	Similar
Pumping Station Operators	34	36	50	53	54	Differences (smaller)
Sawing Machine Setters, Operators, and Tenders, Wood	35	42	47	36	42.5	Upstream/Combination

Table C.2: Occupational Rankings by Scenario and Clustering Algorithm Group Assignments. (Continued)

	Scenario Rankings					
Occupation Title	1	2	3	4	5	Group
Insulation Workers	36	33	30	27	27	Similar
Extruding, Forming, Pressing, and Compacting Machine Setters, Operators, and Tenders	37	19	12	32	32	Differences (smallest)
General and Operations Managers	38	37	35	33	34	Similar
Millwrights	39	39	42	38.5	42.5	Similar
Crushing, Grinding, Polishing, Mixing, and Blending Workers	40	44	44	40	45	Similar
Customer Service Representatives	41	41	36	42	40	Similar
Machine Feeders and Offbearers	42	46	48	57	55	Similar
Personal Financial Advisors	43	45	45	41	46	Similar
Shipping, Receiving, and Traffic Clerks	44	50	58	63	61	Differences (smaller)
Environmental Engineers	45	48	59	58	60	Differences (smaller)
Plant and System Operators, Other	46	39	28	28	22	Differences (smaller)
Food Batchmakers	47	43	41	52	48	Similar
Maintenance Workers, Machinery	48	47	53	55	52	Similar
Constructions Managers	49	52	27	29	26	Differences (smaller)
Cutting Workers	50	61	61	59	66	Upstream/Combination
Electrical Power-Line Installers and Repairers	51	68	62	65	58	Similar
Architects, Except Naval	52	59	55	49	51	Similar
Industrial Truck and Tractor Operators	53	56	79	78	74	Differences (smaller)
Furnace, Kiln, Oven, Drier, and Kettle Operators and Tenders	54	51	63	64	65	Differences (smaller)
Painting Workers and Dyers	55	34	7	8	15	Differences (larger)
Dredge, Excavating, and Loading Machine Operators	56	53	32	34	36	Differences (smaller)
Derrick, Rotary Drill, and Service Unit Operators, and Roustabouts, Oil, Gas, and Mining	57	49	25	19	21	Differences (larger)
Laborers and Freight, Stock, and Material Movers, Hand	58	62	54	50	57	Similar
Earth Drillers, Except Oil and Gas	59	63	56	67	63	Similar
First-Line Supervisors of Mechanics, Installers, and Repairers	60	70	75	77	76	Differences (smaller)
Industrial Production Managers	61	57	51	70	77	Differences (smallest)
Forest and Conservation Workers	62	72	17	9	9	Differences (largest)
Chemical Technicians	63	55	73	75	75	Differences (smaller)
Upholsterers	64	64	70	69	71	Similar
Machinists	65	74	76	72	62	Upstream/Combination
Refuse and Recyclable Material Collectors	66	60	60	38.5	53	Downstream/Combination
Material Moving Workers, Other	67	67	22	23	24	Differences (larger)
Public Relations Specialists	68	77	77	71	78	Similar
Chemical Engineers	69	58	5	13	14	Differences (largest)
Environmental Scientists and Geoscientists	70	54	74	56	56	Differences (smallest)

Table C.2: Occupational Rankings by Scenario and Clustering Algorithm Group Assignments. (Continued)

	Scenario Rankings					
Occupation Title	1	2	3	4	5	Group
Locomotive Engineers and Operators	71	79	80	82	81	Similar
Production, Planning, and Expediting Clerks	72	65	66	66	64	Similar
Woodworking Machine Setters, Operators, and Tenders, Except Sawing	73	69	49	45	49	Differences (smaller)
Designers	74	66	64	68	70	Similar
Healthcare Practitioners and Technical Occupations, Other	75	75	67	54	59	Differences (smaller)
Extraction Workers, Other	76	73	68	76	73	Similar
Paper Goods Machine Setters, Operators, and Tenders	77	71	69	73	72	Similar
Sales Representatives, Wholesale and Manufacturing	78	76	83	81	83	Similar
Assemblers and Fabricators, Other	79	78	78	79	79	Similar
Structural Iron and Steel Workers	80	81	72	74	67	Upstream/Combination
Maintenance and Repair Workers, General	81	80	71	61	68	Differences (smaller)
Inspectors, Testers, Sorters, Samplers, and Weighers	82	82	84	84	84	Similar
Civil Engineers	83	84	81	62	31	(Own Group)
Mining Machine Operators	84	83	82	83	80	Similar

Table C.3: Mean Value of Scaled Contribution Measure for Selected Exchange Occupations: Top Positive Contributions.

		Mean Value of Contribution					
Exchange Occupation Title	# of Focus Occ.	1	2	3	4	5	Scenarios Mean
Engineers, Other	9	0.018	0.069	0.293	0.192	0.308	0.176
Machinists	11	0.008	0.140	0.138	0.166	0.137	0.118
Automotive Service Technicians and Mechanics	15	0.020	0.129	0.127	0.114	0.122	0.102
Financial Managers	5	0.030	0.109	0.100	0.110	0.102	0.090
Maintenance and Repair Workers, General	14	0.015	0.087	0.120	0.078	0.118	0.084
Production, Planning, and Expediting Clerks	6	0.015	0.095	0.077	0.103	0.074	0.073
Heating, Air Conditioning, and Refrigeration Mechanics and Installers	10	-0.001	0.068	0.130	0.046	0.036	0.056
Construction Managers	11	-0.001	-0.022	0.102	0.100	0.079	0.051
First-Line Supervisors of Office and Administrative Support Workers	13	0.010	0.023	0.057	0.093	0.070	0.050
Construction Laborers	26	0.010	0.075	0.055	0.033	0.072	0.049
Carpenters	23	-0.001	0.061	0.059	0.036	0.050	0.041
Designers	12	0.007	0.064	0.005	0.045	0.041	0.033
Heavy Vehicle and Mobile Equipment Service Technicians and Mechanics	7	0.004	0.022	0.014	0.087	0.029	0.031
Software Developers, Applications and Systems Software	11	0.005	0.055	0.022	0.046	0.027	0.031
Managers in Marketing, Advertising, and Public Relations	9	0.016	-0.015	0.028	0.077	0.045	0.030
First-Line Supervisors of Mechanics, Installers, and Repairers	7	0.004	0.023	0.026	0.048	0.046	0.029
Office Clerks, General	11	0.007	0.064	0.015	0.027	0.021	0.027
Customer Service Representatives	14	0.006	0.052	0.017	0.028	0.032	0.027
Construction Equipment Operators except Paving, Surfacing, and Tamping	7	-0.013	-0.121	0.102	0.076	0.089	0.027
Other Teachers and Instructors	6	0.007	0.020	0.041	0.028	0.025	0.024
Human Resources, Training, and Labor Relations Specialists	7	0.004	0.021	0.016	0.033	0.030	0.021
Management Analysts	8	-0.002	-0.020	0.038	0.048	0.039	0.021
Shipping, Receiving, and Traffic Clerks	10	0.001	-0.028	0.057	0.020	0.048	0.020
Data Entry Keyers	5	0.005	0.011	0.028	0.023	0.027	0.019
Retail Salespersons	20	0.004	0.054	-0.001	0.013	0.023	0.019
Accountants and Auditors	9	0.001	0.064	0.011	0.010	-0.007	0.016
Food Service and Lodging Managers	7	-0.004	0.018	0.017	0.009	0.029	0.014
Mechanical Engineers	9	-0.001	-0.025	0.079	0.005	0.006	0.013
Agricultural Workers, Other	6	-0.003	0.020	0.020	0.012	0.007	0.011
Welding, Soldering, and Brazing Workers	14	0.005	0.059	-0.034	0.009	0.013	0.010

Table C.4: Mean Value of Scaled Contribution Measure for Selected Exchange Occupations: Top Negative Contributions.

		Mean Value of Contribution					
Exchange Occupation Title	# of Focus Occ.	1	2	3	4	5	Scenarios Mean
Life, Physical, and Social Science Technicians, Other	5	-0.011	-0.195	-0.148	-0.144	-0.140	-0.127
Engineering Technicians, Except Drafters	14	-0.014	-0.070	-0.180	-0.123	-0.145	-0.106
Packaging and Filling Machine Operators and Tenders	8	-0.009	-0.080	-0.065	-0.111	-0.090	-0.071
Waiters and Waitresses	8	-0.011	-0.027	-0.070	-0.119	-0.091	-0.064
Electrical and Electronics Engineers	8	0.000	0.009	-0.109	-0.081	-0.139	-0.064
Laborers and Freight, Stock, and Material Movers, Hand	31	-0.008	-0.064	-0.109	-0.060	-0.075	-0.063
Industrial and Refractory Machinery Mechanics	14	-0.005	-0.108	-0.052	-0.054	-0.072	-0.058
First-Line Supervisors of Sales Workers	28	0.000	-0.071	-0.071	-0.068	-0.062	-0.054
Electricians	11	-0.003	-0.049	-0.036	-0.073	-0.066	-0.045
Bookkeeping, Accounting, and Auditing Clerks	10	-0.008	-0.021	-0.039	-0.074	-0.072	-0.043
Correspondent Clerks and Order Clerks	5	-0.020	-0.005	-0.061	-0.056	-0.032	-0.035
Packers and Packagers, Hand	9	-0.009	-0.040	-0.063	-0.007	-0.050	-0.034
Computer Scientists and Systems Analysts/Network systems Analysts/Web Developers	15	0.008	-0.014	-0.062	-0.040	-0.052	-0.032
Sales and Related Workers, All Other	5	-0.014	-0.094	-0.039	-0.006	-0.008	-0.032
Driver/Sales Workers and Truck Drivers	31	-0.010	-0.077	-0.020	-0.029	-0.021	-0.031
General and Operations Managers	9	-0.001	-0.032	-0.014	-0.039	-0.064	-0.030
Stock Clerks and Order Fillers	15	-0.008	-0.049	-0.036	-0.028	-0.022	-0.029
First-Line Supervisors of Construction Trades and Extraction Workers	18	-0.006	-0.060	-0.002	-0.016	-0.037	-0.024
Cashiers	16	-0.004	-0.031	-0.010	-0.044	-0.028	-0.023
Sales Representatives, Services, All Other	7	-0.005	-0.018	-0.039	-0.020	-0.027	-0.022
Unemployed	79	-0.001	-0.008	-0.032	-0.019	-0.051	-0.022
Painters, Construction and Maintenance	8	0.002	0.001	-0.028	-0.04	-0.028	-0.019
Sales Representatives, Wholesale and Manufacturing	14	-0.005	-0.022	-0.020	-0.026	-0.016	-0.018
Security Guards and Gaming Surveillance Officers	7	-0.005	-0.006	-0.024	-0.028	-0.028	-0.018
Bus and Truck Mechanics and Diesel Engine Specialists	6	0.004	-0.009	0.002	-0.050	-0.023	-0.015
Metal Workers and Plastic Workers, Other	13	0.006	0.031	-0.026	-0.045	-0.042	-0.015
Chief Executives and Legislators/Public Administration	13	-0.005	-0.009	-0.028	-0.014	-0.012	-0.014
Grounds Maintenance Workers	15	0.000	-0.008	-0.018	-0.007	-0.017	-0.010
Secretaries and Administrative Assistants	15	0.008	0.041	-0.042	-0.035	-0.024	-0.010

Table C.5: Mean Absolute Value of Scaled Contribution Measure for Selected Exchange Occupations.

		Mean Absolute Value of Contribution					
Exchange Occupation Title	# of Focus Occ.	1	2	3	4	5	Scenarios Mean
Engineers, Other	9	0.036	0.453	0.457	0.374	0.434	0.351
Unemployed	79	0.038	0.412	0.364	0.302	0.316	0.286
Construction Equipment Operators except Paving, Surfacing, and Tamping	7	0.028	0.251	0.360	0.254	0.263	0.231
Managers, Other (including Postmasters)	43	0.035	0.343	0.246	0.161	0.181	0.193
Other Production Workers	34	0.020	0.253	0.234	0.172	0.187	0.173
Industrial Engineers, including Health and Safety	10	0.046	0.283	0.234	0.147	0.148	0.171
Electrical and Electronics Engineers	8	0.026	0.270	0.243	0.124	0.177	0.168
Retail Salespersons	20	0.038	0.155	0.218	0.206	0.218	0.167
Heating, Air Conditioning, and Refrigeration Mechanics and Installers	10	0.034	0.302	0.200	0.156	0.143	0.167
Civil Engineers	9	0.021	0.226	0.237	0.148	0.185	0.163
First-Line Supervisors of Sales Workers	28	0.035	0.214	0.211	0.178	0.176	0.163
Laborers and Freight, Stock, and Material Movers, Hand	31	0.027	0.215	0.197	0.148	0.167	0.151
Construction Laborers	26	0.033	0.193	0.173	0.152	0.147	0.140
Life, Physical, and Social Science Technicians, Other	5	0.027	0.217	0.159	0.146	0.140	0.138
Managers in Marketing, Advertising, and Public Relations	9	0.028	0.187	0.140	0.158	0.147	0.132
Machinists	11	0.026	0.167	0.142	0.171	0.139	0.129
Mechanical Engineers	9	0.018	0.217	0.198	0.091	0.117	0.128
Automotive Service Technicians and Mechanics	15	0.024	0.146	0.161	0.156	0.149	0.127
First-Line Supervisors of Construction Trades and Extraction Workers	18	0.030	0.222	0.153	0.106	0.123	0.127
Assemblers and Fabricators, Other	24	0.029	0.185	0.145	0.131	0.128	0.124
Chief Executives and Legislators/Public Administration	13	0.017	0.155	0.180	0.128	0.124	0.121
Engineering Technicians, Except Drafters	14	0.022	0.121	0.183	0.123	0.145	0.119
General and Operations Managers	9	0.021	0.137	0.130	0.109	0.151	0.110
Maintenance and Repair Workers, General	14	0.020	0.148	0.138	0.095	0.130	0.106
Carpenters	23	0.019	0.117	0.139	0.119	0.134	0.106
Financial Managers	5	0.035	0.130	0.120	0.118	0.119	0.104
Packaging and Filling Machine Operators and Tenders	8	0.015	0.103	0.121	0.124	0.123	0.097
Driver/Sales Workers and Truck Drivers	31	0.022	0.163	0.112	0.094	0.094	0.097
Packers and Packagers, Hand	9	0.023	0.134	0.136	0.064	0.126	0.097

Table C.5: Mean Absolute Value of Scaled Contribution Measure for Selected Exchange Occupations. (Continued)

		Mean Absolute Value of Contribution					
Exchange Occupation Title	# of Focus Occ.	1	2	3	4	5	Scenarios Mean
Grounds Maintenance Workers	15	0.015	0.131	0.118	0.088	0.118	0.094
Accountants and Auditors	9	0.013	0.125	0.111	0.097	0.118	0.093
Metal Workers and Plastic Workers, Other	13	0.023	0.092	0.112	0.112	0.122	0.092
Industrial and Refractory Machinery Mechanics	14	0.021	0.197	0.104	0.063	0.075	0.092
Production, Planning, and Expediting Clerks	6	0.028	0.120	0.097	0.110	0.087	0.088
Constructions Managers	11	0.011	0.120	0.111	0.102	0.083	0.085
Computer Scientists and Systems Analysts/Network systems Analysts/Web Developers	15	0.013	0.111	0.113	0.078	0.098	0.082
Electricians	11	0.018	0.124	0.090	0.093	0.085	0.082
Shipping, Receiving, and Traffic Clerks	10	0.021	0.142	0.100	0.060	0.087	0.082
Pipelayers, Plumbers, Pipefitters, and Steamfitters	10	0.031	0.124	0.090	0.088	0.074	0.081
Cashiers	16	0.023	0.130	0.064	0.100	0.084	0.080
Waiters and Waitresses	8	0.028	0.064	0.077	0.128	0.100	0.079
Software Developers, Applications and Systems Software	11	0.014	0.094	0.093	0.090	0.101	0.078
Welding, Soldering, and Brazing Workers	14	0.013	0.126	0.097	0.085	0.060	0.076
Architectural and Engineering Managers	5	0.020	0.128	0.123	0.029	0.060	0.072
Sales and Related Workers, All Other	5	0.020	0.141	0.064	0.058	0.069	0.071
First-Line Supervisors of Office and Administrative Support Workers	13	0.018	0.061	0.089	0.097	0.080	0.069
Secretaries and Administrative Assistants	15	0.016	0.070	0.090	0.082	0.088	0.069
First-Line Supervisors of Production and Operating Workers	23	0.020	0.137	0.060	0.069	0.055	0.068
Customer Service Representatives	14	0.022	0.087	0.074	0.069	0.081	0.067
Industrial Truck and Tractor Operators	8	0.023	0.099	0.092	0.038	0.070	0.064
Janitors and Building Cleaners	23	0.015	0.071	0.094	0.067	0.069	0.063
Inspectors, Testers, Sorters, Samplers, and Weighers	17	0.025	0.115	0.092	0.040	0.040	0.062
Heavy Vehicle and Mobile Equipment Service Technicians and Mechanics	7	0.016	0.058	0.060	0.098	0.065	0.059
Bookkeeping, Accounting, and Auditing Clerks	10	0.012	0.038	0.072	0.089	0.081	0.058
Management Analysts	8	0.014	0.083	0.074	0.052	0.065	0.058
Correspondent clerks and order clerks	5	0.025	0.101	0.065	0.058	0.038	0.057
Sales Representatives, Services, All Other	7	0.020	0.062	0.064	0.063	0.065	0.055

Table C.6: Coefficients from Regression of Scenario Rankings on Exchange Occupation Contribution Measures.

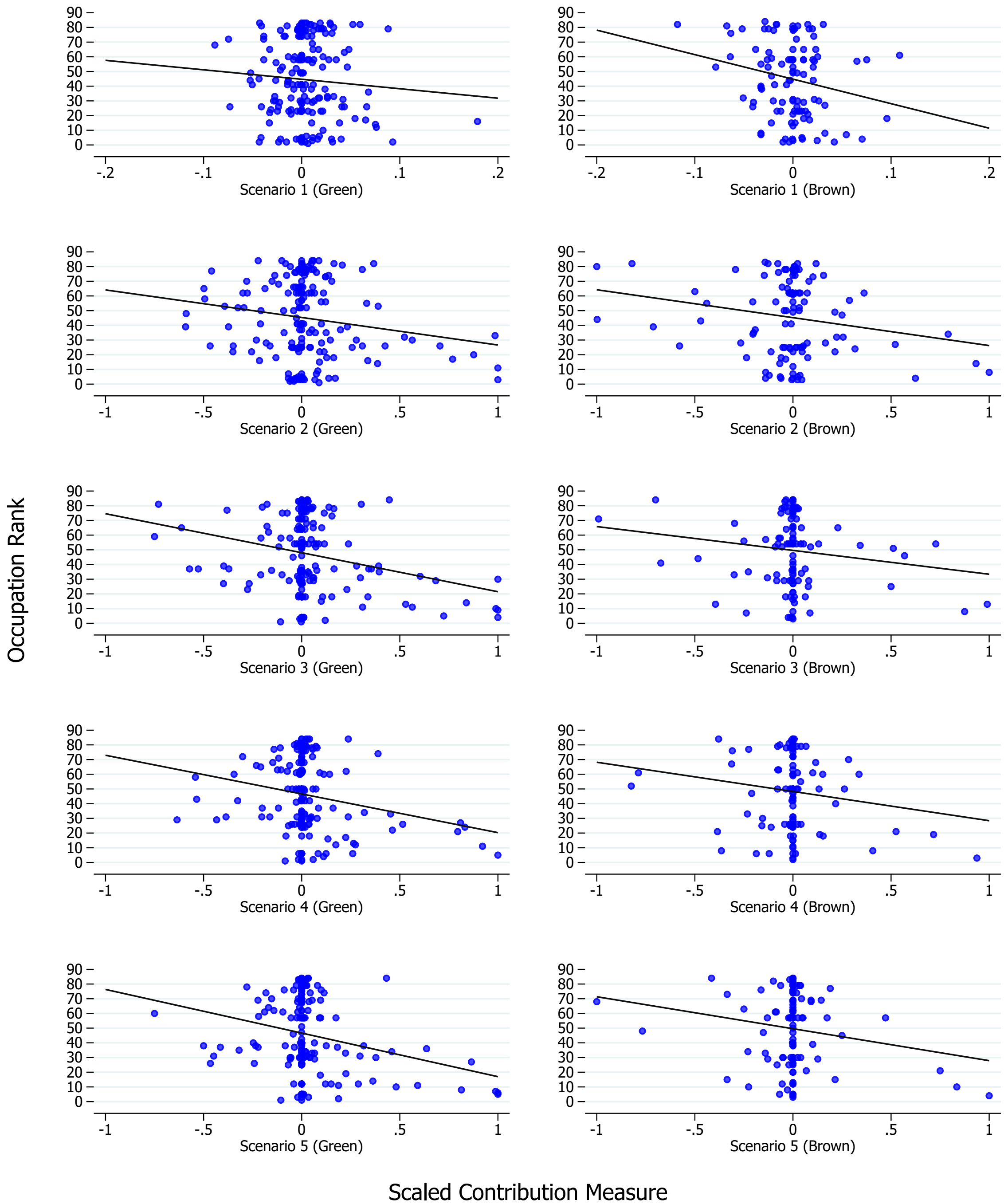
		Scenario Coefficient				
Exchange Occupation Title	# of Focus Occ.	1	2	3	4	5
Unemployed	79	-150***	-21***	-25***	-25***	-26***
Managers, Other (including Postmasters)	43	-152*	-14*	-11	-13	-23*
Other Production Workers	34	-181	-17*	-11	-15	-14
Laborers and Freight, Stock, and Material Movers, Hand	31	-265**	-19	-35**	-20	-29
First-Line Supervisors of Sales Workers	28	55	13	-15	-30*	-29
Assemblers and Fabricators, Other	24	-248*	-33*	-17	-11	-16
Carpenters	23	-313	-20	-49*	-61**	-48*
Retail Salespersons	20	-320***	-44**	-36***	-37**	-35***
Automotive Service Technicians and Mechanics	15	-403*	-58*	-56**	-57***	-58**
Maintenance and Repair Workers, General	14	-111	-47	-57**	-42	-57**
Engineering Technicians, Except Drafters	14	-383	-53	-50**	-46	-44*
Chief executives and legislators/public administration	13	28	-23	-29	-38	-67*
Designers	12	394	19	530**	34	-17
Heating, Air Conditioning, and Refrigeration Mechanics and Installers	10	-338**	-32*	-30	-49*	-58*
Accountants and Auditors	9	-832***	-65**	-77**	-84**	-69**
Engineers, Other	9	-142	-22**	-23	-30**	-20*
Civil Engineers	9	-152	-13	-26	-54**	-36*
Construction Equipment Operators except Paving, Surfacing, and Tamping	7	-426***	-30	-35**	-47**	-44**
Security Guards and Gaming Surveillance Officers	7	-3165*	-404	-183	7	157
Heavy Vehicle and Mobile Equipment Service Technicians and Mechanics	7	-1014*	-188	68	32	128
Farmers, Ranchers, and Other Agricultural Managers	6	1143	353	884*	-21	264
Agricultural Workers, Other	6	-985	-399**	-278	-1024	-566
Architectural and Engineering Managers	5	-179	-49	-62	67	63**
Computer, Automated Teller, and Office Machine Repairers	5	-2298	-2706*	-396	-545	-359
Data Entry Keyers	5	-480*	-211***	-365***	-346	-425*

Table C.6: Coefficients from Regression of Scenario Rankings on Exchange Occupation Contribution Measures. (Continued)

		Scenario Coefficient				
Exchange Occupation Title	# of Focus Occ.	1	2	3	4	5
Sales and Related Workers, All Other	5	-498	-35	-192*	-126	-98
Sewing Machine Operators	4	9175	315**	303	1342	0
Grinding, Lapping, Polishing, and Buffing Machine Tool Setters, Operators, and Tenders, Metal and Plastic	4	707	28*	50	-136	-91
Childcare Workers	4	-86	454	551	1671	947*
Telecommunications Line Installers and Repairers	4	-408	-56	-82	-112	-258**
Door-to-Door Sales Workers, News and Street Vendors, and Related Workers	3	-3341*	499	18	-51	38
Advertising Sales Agents	3	8059	3748	7205*	31819	7185
Aircraft Mechanics and Service Technicians	3	-265	54	-130*	-235	-119
Graders and Sorters, Agricultural Products	3	5384**	535**	327	2850	596
Carpet, Floor, and Tile Installers and Finishers	3	161	-69	136	136	142***
Electric Motor, Power Tool, and Related Repairers	3	2330**	1498	-23	103	-34*
First-Line Supervisors of Food Preparation and Serving Workers	3	-776	-205	-445**	-556	-1103
Painting Workers and Dyers	3	8654	482***	1024	440*	739
	Count	**15**	**16**	**16**	**11**	**19**

Note: this table displays coefficients from a regression of the rankings in each scenario against the contribution measures of individual exchange occupations (for those exchange occupations that have at least one statistically significant relationship across the scenarios). Coefficients marked with *, **, and *** are statistically significant at the ten, five, and one percent levels, respectively.

Figure C.1: Relationship of Green/Brown Scaled Contribution Measures to Overall Rankings.



Note: this figure illustrates the relationships between the scaled contribution measures for green and brown exchange occupations and the overall occupational rankings. The lines in each plot are fitted through the points in the associated scenario. All of the relationships are significant at the ten percent level, with the exception of the relationship for green exchange occupations in scenario one.