DIGITAL ACCESS TO SCHOLARSHIP AT HARVARD
DASH.HARVARD.EDU

HARVARD LIBRARY
Office for Scholarly Communication

Essays on Innovation and Growth

Permanent link
http://nrs.harvard.edu/urn-3:HUL.InstRepos:40046455

Terms of Use
This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. Submit a story .

Accessibility

Essays on Innovation and Growth

A dissertation presented

by

Danial Lashkari

to

The Committee on Higher Degrees in Political Economy and Government

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Political Economy and Government

Harvard University

Cambridge, Massachusetts

April 2017

© 2017 Danial Lashkari
All rights reserved.

Dissertation Advisors:
Professor Marc Melitz
Professor Pol Antràs

Author:
Danial Lashkari

Essays on Innovation and Growth

Abstract

This dissertation is composed of three essays on economic growth.

The first chapter presents a new multi-sector growth model that accommodates long-run demand and supply drivers of structural change. The model generates nonhomothetic Engel curves at all levels of development and is consistent with the decline in agriculture, the hump-shaped evolution of manufacturing and the rise of services over time. The economy converges to a constant aggregate growth rate that depends on sectoral income elasticities, capital intensities and rates of technological progress. The chapter also presents estimates for the demand system derived from the model using historical data on sectoral employment shares from twenty-five countries and household survey data from the US. The estimated model parsimoniously accounts for the broad patterns of sectoral reallocation observed among rich, miracle and developing economies in the post-war period. We find that income effects play a major role in generating structural change.

The second chapter develops a multi-sector endogenous growth model in which the direction of innovation across sectors is endogenous. The model provides a theoretical general equilibrium framework for studying the classical demand-pull and technology-push drivers of innovation. A robust prediction is that the rate of growth innovation growth is asymptotically higher in more income-elastic sectors. The chapter also provides a test of this prediction using the universe of U.S. patents and firm R&D investments for the period 1976-2007. The analysis lends empirical support for the main predictions of the model.

Finally, the third chapter constructs a theory of industry growth to study the implications of knowledge diffusion and selection for innovation policy. Firms' ideas determine their

productivity and stochastically evolve over time. Firms innovate to improve their ideas and endogenously exit if unsuccessful. Entrants adopt the ideas of incumbents. In this model, creative destruction operates through selection: when better ideas are innovated or adopted, they selectively replace worse ideas. Innovation externalities vary based on firm productivity: ideas generated by more productive firms create 1) longer-lasting positive externalities due to knowledge diffusion and 2) stronger negative externalities due to dynamic competition effects. Therefore, the net external effect of innovation is heterogeneous across firms. Quantitatively, this heterogeneity is large when the model is calibrated to firm-level data from US manufacturing and retail trade, and implies first-order considerations for the design of innovation policy.

Contents

Abstract . . . iii
Acknowledgments . . . xii

1 Structural Change with Long-run Income and Price Effects **1**
1.1 Introduction . . . 1
1.2 Model . . . 7
1.2.1 Preferences and the Household Problem . . . 7
1.2.2 Production and Competitive Equilibrium . . . 12
1.2.3 Constant Growth Path . . . 14
1.3 Quantitative Exploration of a Cross-Country Panel . . . 16
1.3.1 Empirical Strategy . . . 17
1.3.2 Data Description . . . 19
1.3.3 Estimation Results . . . 20
1.3.4 Correlation between Real and Nominal Value Added . . . 26
1.3.5 Comparison with Stone-Geary Preferences . . . 27
1.3.6 Beyond Three Sectors . . . 29
1.4 Micro and Macro Estimates for the U.S. . . . 31
1.4.1 Micro Estimation: Consumer Expenditure Survey . . . 31
1.4.2 Macro Estimation: Value Added and Expenditure Measures . . . 35
1.5 Conclusion . . . 38

2 Demand-Pull, Technology-Push, and the Sectoral Direction of Innovation **40**
2.1 Introduction . . . 40
2.2 Motivating Evidence . . . 46
2.3 Model . . . 49
2.3.1 Demand and Income Elasticities . . . 50
2.3.2 Production . . . 54
2.3.3 Innovation and Technology-Push . . . 55
2.3.4 Market Equilibrium . . . 57
2.3.5 Equilibrium Dynamics . . . 61
2.3.6 Constant Growth Path . . . 65

2.4 Conclusion . . . 72

3 Innovation, Knowledge Diffusion, and Selection **73**
3.1 Introduction . . . 73
3.2 Model . . . 82
3.2.1 Households and Demand . . . 82
3.2.2 Firms and Innovation . . . 83
3.2.3 Entry . . . 87
3.2.4 Stationary Constant Growth Paths . . . 89
3.2.5 Incumbent Problem . . . 92
3.2.6 Industry Problem . . . 94
3.2.7 Equilibrium Industry Growth . . . 96
3.2.8 Solving for the Model . . . 100
3.3 Optimal Allocations . . . 103
3.3.1 Socially Optimal Allocations . . . 104
3.3.2 Unpacking Innovation Externalities . . . 108
3.3.3 Computing the Socially Optimal Allocation . . . 111
3.3.4 Tax/Subsidy Implementation of Optimal Policies . . . 111
3.4 Quantitative Results . . . 114
3.4.1 Data . . . 114
3.4.2 Model Specification . . . 115
3.4.3 Calibration Strategy . . . 116
3.4.4 Calibration Results . . . 118
3.4.5 Optimal Allocations . . . 124
3.4.6 Effect of Static Competition . . . 129
3.5 Discussion and Future Work . . . 131

References **133**

Appendix A Appendix to Chapter 1 **144**
A.1 Nonhomothetic CES Preferences . . . 144
A.1.1 General Nonhomothetic CES Preferences . . . 145
A.1.2 Isoelastic Nonhomothetic CES Preferences . . . 149
A.1.3 Comparison to Generalized Stone-Geary and PIGL Preferences . . . 151
A.2 Proofs of Propositions and Lemmas . . . 153
A.3 Additional Figures . . . 159

Appendix B Appendix to Chapter 2 **169**
B.1 Proofs 169
B.2 Generalized Spillover Function 176
B.3 Schumpeterian Model 183

Appendix C Appendix to Chapter 3 **185**
C.1 Facts 185
C.2 Derivations and Proofs 188
C.2.1 Section 3.2 188
C.2.2 Section 3.3 202
C.2.3 Section 3.4 209

List of Tables

1.1 Baseline Estimates for the Cross-Country Sample 21
1.2 Correlation of Nominal and Real Value Added 27
1.3 Sectoral R^2 measures for Stone-Geary and nonhomothetic CES 29
1.4 10-Sector Regression 30
1.5 Consumer Expenditure Survey, 1980-2006 33
1.6 Instrumental Variables Strategy Consumption Expenditure Survey, 2001-2002 . . 34
1.7 Estimation by Quartiles and Sub-periods Consumer Expenditure Survey, 1980-2006 35
1.8 Consumption Expenditure and Value Added for the U.S. 37

2.1 Regressions of annual patent growth on the income elasticity parameters from Aguiar and Bils (2015a). The three columns on the left include the total number of patents, while the three columns on the right weight patents by their number of citations. 48
2.2 Regressions of the rate of R&D expenditure growth on the income elasticity parameters from Aguiar and Bils (2015a). 49

3.1 Targeted moments and the calibrated parameters for retail trade and manufacturing. The first three targeted moments each directly pin down a corresponding parameter in the model: the rate of productivity growth pins down the relative cost of entry ψ, the rate of employment growth pins down the rate of growth in the number of firms g_M, and the cost elasticity pins down the curvature parameter of innovation costs κ. The 5 moments of the firm life-cycle dynamics together calibrate the five remaining parameters of the model, characterizing the volatility of firm outcomes σ, the costs of innovation χ, and the process of entry and adoption (μ, η, γ_e). 119
3.2 The impact of a drastic fall in the costs of innovation (parameter χ) to 75% of its initial value on industry outcomes. For each sector, the leftmost column, under the heading "Before," shows the value of the variable under the calibrated parameters, the middle column, under the heading "After," shows the value after the change. The third column then presents the difference between the two values. IDR stands for Inter-Decile Range. 121

3.3 The characteristics of the optimal allocations of innovation and production, as well as a subsidy scheme that implements them, in retail trade and manufacturing. See Section 3.4.5 for details. 128

A.1 Residual Relative Log-Expenditure Shares on Residual Aggregate Consumption 162
A.2 Contribution of Relative Prices and Consumption 162
A.3 Partial Correlations 163
A.4 Heterogeneous Price Elasticity of Substitution 163
A.5 Value Added Regressions 163
A.6 Stone-Geary Estimation Results 164
A.7 Contribution of Prices and Consumption for OECD Countries 165
A.8 Contribution of Prices and Consumption for Non-OECD Countries 166
A.9 Growth Rates of Relative Prices in the Country Panel 166

List of Figures

1.1 Partial Correlations of Sectoral Expenditure and Aggregate Consumption . . 3

1.2 Baseline fit with common preference parameters $\{\sigma, \epsilon_a - \epsilon_m, \epsilon_s - \epsilon_m\}$ for six countries 24

1.3 Stone-Geary fit with common preference parameters $\{\sigma, \bar{c}_a, \bar{c}_s\}$ for six countries 24

2.1 Determination of relative technologies in a static two-sector model of endogenous growth and nonhomothetic demand. The $\mathfrak{D}$ curve shows the relative sectoral consumer demand, and the $\mathfrak{T}^{-1}$ curve the relative supply of innovation. So long as the innovation supply function is not perfectly inelastic (the dashed line), when real income C_{tot} grows, it shifts up the relative demand for the income elastic sector s, and the equilibrium level of relative technology shifts to the right. 44

3.1 The solution to the Bellman equation in the two-dimensional space of $(u_1(z), u_2(z))$ for a given choice of model parameters (details in the online appendix). The relative sizes of arrows on the path of $u(z)$ indicate the relative values of the derivative of u with respect to z along the path. The equilibrium point u^*, which lies on the 45° line, is shown with a red dot. The stable arm around the saddle-path stable equilibrium point is also shown in light gray. For very large firms, as $z \to \infty$, the path of $u(z)$ asymptotically converges to this arm. 102

3.2 The solution to the differential equation characterizing the stationary distribution in the space of $(h(z), H(z))$ for a given choice of model parameters and an atomic distribution of adoption F_a (details in the online appendix). The two equilibrium points and the point of entry are shown in red. The stable arm around the saddle-path stable equilibrium point $(H, h) = (1, 0)$ is shown in light gray. For very large firms, as $z \to \infty$, the path of (H, h) asymptotically converges to the stable arm. 103

3.3 Comparison between the optimal and market equilibrium allocations of innovation in (Right) manufacturing and (Left) retail, as a function of the relative efficiency of firms. RET and MAN stand for retail and manufacturing, respectively. 124

3.4 Comparison between the optimal and market equilibrium long-run distributions of efficiency in (Right) manufacturing and (Left) retail. RET and MAN stand for retail and manufacturing, respectively. 125
3.5 The difference between the optimal and market equilibrium allocations in (Right) innovation investments and (Left) long-run probability distribution function, as a function of the relative efficiency of firms. RET and MAN stand for retail and manufacturing, respectively. 126
3.6 (Left) Comparison between the subsidy scheme that achieves the optimal innovation investments in manufacturing and retail, as a function of the relative efficiency of firms. (Right) The same comparison but zooming in on the right tail of the productivity distribution. Here, the x-axis shows the percentile ranking of the productivity of firms at the top of the distribution of productivities. RET and MAN stand for retail and manufacturing, respectively. 127
3.7 Comparison between the exact and the linearly approximated optimal subsidy for (Left) retail and (Right) manufacturing. The gray vertical line indicates the rank corresponding to the maximum of the subsidy rate. 127
3.8 Comparison of the difference between the optimal and market equilibrium allocations with two different degrees of substitutability in firm outputs in (Right) retail and (Left) manufacturing, as a function of the relative efficiency of firms. RET and MAN stand for retail and manufacturing, respectively. . . 130
3.9 Comparison between the subsidy scheme that achieves the optimal innovation investments with two different degrees of substitutability in firm outputs in (Left) retail and (Right) manufacturing, as a function of the relative efficiency of firms. RET and MAN stand for retail and manufacturing, respectively. . . 131

A.1 Regression Fit for Japan using common world parameters $\{\sigma, \epsilon_a - \epsilon_m, \epsilon_s - \epsilon_m\}$ 159
A.2 Baseline Country Fit 160
A.3 Baseline Country Fit (continued) 161
A.4 Predicted Evolution of Expenditure Shares for the U.S. 165
A.5 Country Fit Using Stone-Geary Preferences 167
A.6 Country Fit Using Stone-Geary Preferences 168

Acknowledgments

This is the second PhD dissertation that I have completed in my life. I made the decision six years ago to transition from my old life in computer science and neuroscience into social sciences and specifically economics. The decision may have come with understandable personal costs, but the benefits have far outweighed those. Ever since I entered my new field, I have felt blessed for having the opportunity to study the questions that have preoccupied me for so long, using the mathematical and quantitative tools that I cherish and enjoy so much. Needless to say, I owe my success in fulfilling this transition and completing the current dissertation to many individuals who have supported and guided me along the way.[1]

First, I feel indebted to Alberto Alesina, who was my main mentor and champion throughout the transition, for his support and trust in my potential as an economist. I am also immensely grateful for the support and advice of my dissertation committee members, Marc Melitz, Pol Antràs, Robert Barro, Elhanan Helpman, and Dani Rodrik. In addition, I have also benefitted from the advice and guidance that Philippe Aghion, Robert Z. Lawrence, William Kerr, and Jane Mansbridge have offered to me on different aspects of my transition and throughout my studies at Harvard. I also thank Nicole Tareosian, the director of the doctoral program, for being extremely helpful in navigating the program and fulfilling the requirements throughout these years.

Two out of the three essays included in this dissertation are the outcomes of joint research with Diego Comin and Martí Mestieri. Working with Diego and Martí played a key role in shaping my experience as an economics researcher, and I am thankful to them for all that I have learned through this collaboration and for their trust and unwavering support.

I would also like to acknowledge a number of colleagues at Harvard and MIT, notably Natalie Bau, David R. Baqaee, Kirill Borusyak, Andrew Garin, Tomoko Harigaya, Rohan

[1]Of course, many others have played this very role in earlier stages of my academic journey. For brevity, I only focus here on my time at Harvard and refer the interested readers to a detailed account of those earlier stages in Lashkari (2011).

Kekre, Pooya Molavi, Alex Peysakhovich, Mikkel Plagborg-Møller, Francisco Queiró, Martin Rotemberg, and Tom Zimmermann, who have been instrumental in helping me adjust to life as an economist and develop my research ideas. I also fondly acknowlegde other friends and allies who have been by my side and have made my time at Harvard richer and more enjoyable, especially Sanaz Bahargam, Christine Hwang, Taraneh Meshkani, Maria Metzler, and George Yin.

In addition, numerous friends and colleagues have offered feedback and inputs that have helped shape this dissertation.

For the first chapter, I particularly thank Robert Barro and Kiminori Matsuyama for comments on a previous draft, as well as Timo Boppart, Francisco Buera, Berthold Herrendorf, Bart Hobijn, Chad Jones, Joe Kaboski, Pete Klenow, Dirk Krueger, Robert Z. Lawrence, Serguei Maliar, Alex Monge-Naranjo, Martin Rotemberg, Orie Shelef, Tomasz Swiecki, Chris Tonetti, Kei-Mu Yi, and participants in presentations at Berkeley, Central Bank of Armenia, EFJK, Harvard, Michigan, Northwestern, U. Penn, Santa Clara, SED, St. Andrews, the Philadelphia Fed, the San Francisco Fed, Stanford, the Theory and Measurement Conference Honoring the Scholarship of Robert M. Townsend and UBC for useful comments and feedback.[2]

For the second chapter, I thank Philippe Aghion and seminar participants at Harvard University, Arizona State University, and the Society for Economic Dynamics for insightful comments. I am also grateful to Kun Li, Enrico Berkes, and Ben Pugsley for help with the data for this chapter.

Finally, for the third chapter, I thank Davin Chor, Xavier Gabaix, Rohan Kekre, Pooya Molavi, Mikkel Plagborg-Mller, Francisco Queiró, and Martin Rotemberg.

[2]In addition, on behalf of my coauthors, I would like to acknowledge the generous support of the National Science Foundation and the Institute for New Economic Thinking, for Diego Comin, and the generous support of the Agence Nationale de la Recherche (JJCC - GRATE program), for Martí Mestieri.

To two dear friends, Siavosh and Jessica.

Chapter 1

Structural Change with Long-run Income and Price Effects[1]

1.1 Introduction

Economies undergo large scale sectoral reallocations of employment and capital as they develop, in a process commonly known as structural change (Kuznets, 1973; Maddison, 1980; Herrendorf *et al.*, 2014; Vries *et al.*, 2014). These reallocations lead to a gradual fall in the relative size of the agricultural sector and a corresponding rise in manufacturing. As income continues to grow, services eventually emerge as the largest sector in the economy. Leading theories of structural change attempt to understand these sweeping transformations through mechanisms involving either supply or demand. Supply-side theories focus on differences across sectors in the rates of technological growth and capital intensities, which create trends in the composition of consumption through price (substitution) effects (Baumol, 1967a; Ngai and Pissarides, 2007a; Acemoglu and Guerrieri, 2008). Demand-side theories, in contrast, emphasize the role of heterogeneity in income elasticities of demand across sectors (nonhomotheticity in preferences) in driving the observed reallocations accompanying income growth (Kongsamut *et al.*, 2001).

[1]Co-authored with Diego Comin and Martí Mestieri.

The shapes of sectoral Engel curves play a crucial role in determining the contribution of supply and demand channels to structural change.[2] If the differences in the slopes of Engel curves are large and persistent, demand channels can readily explain reallocation of resources toward sectors with higher income elasticities. For instance, steep upward Engel curves for services, flat Engel curves for manufacturing, and steep downward Engel curves for agricultural products can give rise to sizable shifts of employment from agriculture toward services. However, demand-side theories have generally relied on specific classes of nonhomothetic preferences, e.g., generalized Stone-Geary preferences, that imply Engel curves that level off quickly as income grows. Because of this rapid flattening-out of the slopes of Engel curves across sectors, these specifications limit the explanatory power of the demand channel in the long-run.

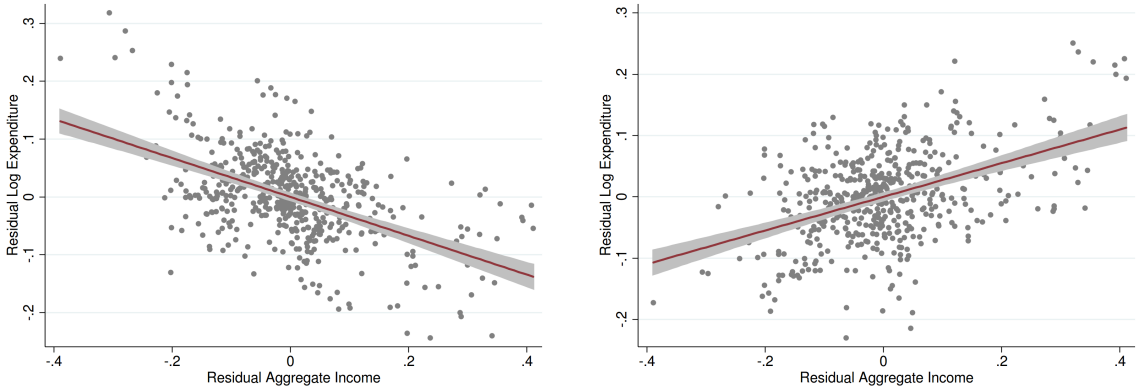
The empirical evidence suggests that Engel curves remain constant at different income levels and do not level off rapidly as income grows. At the micro level, Engel curves have been shown to be well approximated by log-linear functions with constant slopes.[3] We complement this evidence and show that income elasticities stay quite similar across different income brackets using the Consumption Expenditure survey (CEX). At the macro level, log-linear Engel curves also provide reasonable approximations to aggregate consumption variables. Figure (1.1) plots the relationship between the residual (log) expenditure share in agriculture (Figure 1.1a) and services (Figure 1.1b) relative to manufacturing on the y-axis and residual (log) income on the x-axis after controlling for relative prices. The depicted

[2]We define Engel curves as the relationship between sectoral consumption shares and aggregate real consumption holding prices constant.

[3]For instance, Aguiar and Bils (2015b) use the U.S. Consumer Expenditure Survey (CEX) to estimate Engel curves for 20 different consumption categories. Their estimates for the income elasticities are different from unity and vary significantly across consumption categories. Young (2012a) employs the Demographic and Health Survey (DHS) to infer the elasticity of real consumption of 26 goods and services with respect to income for 29 sub-Saharan and 27 other developing countries. He estimates the elasticity of consumption for the different categories with respect to the education of the household head and then uses the estimates of the return to education from Mincerian regressions to back out the income elasticity of consumption. Young also uses a log-linear Engel curve formulation and finds that the slopes of Engel curves greatly differ across consumption categories but appear stable over time. Olken (2010) discusses Young's exercise using Indonesia survey data and finds similar results for a small sample of three goods and services he studies. Young (2013a) also makes use of log-linear Engel curves to infer consumption inequality.

Figure 1.1: *Partial Correlations of Sectoral Expenditure and Aggregate Consumption*

(a) *Agriculture relative to Manufacturing* **(b)** *Services Relative to Manufacturing*



Notes: Data for OECD countries, 1970-2005. Each point corresponds to a country-year observation after partialling-out sectoral prices and country fixed effects. The red line depicts the OLS fit, the shaded regions, the 95% confidence interval.

log-linear fit shows that a constant slope captures a considerable part of the variation in the data.[4]

Motivated by this evidence, we develop a multi-sector model of structural transformation that accommodates for nonhomotheticity in the form of log-linear Engel curves, as well as trends in relative prices. The model builds on the standard framework used in recent empirical work on structural transformation (e.g., Herrendorf *et al.*, 2013). Our key departure from the standard framework is the introduction of a class of utility functions that generates heterogeneous, nonhomothetic sectoral demands for all levels of income, including when income grows toward infinity. These preferences, which we will refer to as nonhomothetic Constant Elasticity of Substitution (CES) preferences, have been studied by Gorman (1965), Hanoch (1975a), Sato (1975a), and Blackorby and Russell (1981) in the context of static, partial-equilibrium models. We show how to embed these preferences into a general equilibrium model of economic growth.

[4]The partial R^2 of the regressions shown in Figure 1.1 are 27% and 20%, respectively. In fact, if we split the sample into observations below and above the median income in the sample and estimate the relative Engel curves separately, we cannot reject the hypothesis of identical slopes of the Engel curves. See Table A.1 in the appendix. If we reported separately the Engel curves for agriculture, manufacturing and services, we would find a negative, zero and positive slope, respectively.

Nonhomothetic CES preferences present a number of advantatges. In addition to their tractability, they allow for income elasticity parameters to be independent of the elasticity of substitution between sectors, a feature that is unique to these preferences. Since our framework does not impose functional relationships between income and substitution elasticities, it lends itself to the task of decomposing the contributions of the demand and supply channels to structural change. In addition, our framework can accommodate an arbitrary number of sectors with heterogeneous and independent income elasticities. As a result, they generate Engel curves for different sectors that match the evidence discussed above: the logarithm of relative demand for the output of each sector has an approximately linear relationship with the logarithm of income. More specifically, this relationship is characterized by a sector-specific income elasticity parameter. We take advantage of these features to study a standard three-sector setting with agriculture, manufacturing and services. We then extend our analysis to a richer sectoral disaggregation (ten sectors) to explore reallocation patterns within manufacturing and services.

Our theory of structural change yields a number of theoretical and empirical results. First, the equilibrium in our model asymptotically converges to a path of constant real consumption growth. The asymptotic growth rate of real consumption depends on parameters characterizing both the supply and demand channels; it is a function of the sectoral income elasticities as well as sectoral growth rates of TFP and sectoral factor intensities. In this respect, our model differs from standard models using Stone-Geary preferences in which long-run growth is pinned down solely by the growth rate of TFP, and generalizes the findings of Ngai and Pissarides (2007a) and Acemoglu and Guerrieri (2008). Second, our theory can produce similar evolutions for nominal and real sectoral measures of economic activity, which is a robust feature of the data.[5] This is a consequence of the role of income elasticities in generating sectoral reallocation patterns. Third, our framework can generate hump-shaped patterns for the evolution of manufacturing consumption shares, which is a

[5] Herrendorf *et al.*, 2014 show that supply-side driven structural transformation cannot account for the similar evolution of nominal and real sectoral measures of activity.

well-documented feature in the data (Buera and Kaboski, 2012a).

To evaluate the model empirically, we use structural equations derived from our theory to estimate the elasticities that characterize our utility function. We use historical cross-country sectoral data and household expenditure data that vary in their geographies and periods covered, and in their measures of economic activity used to capture structural change. A major finding is that the estimates of the elasticity of substitution and the relative slopes of the Engel curves across sectors are robust to the sample of countries, time periods and economic measures of sectoral activity. This demonstrates that the patterns presented in Figure 1.1 not only characterize the Engel curves in the OECD but also apply more broadly to countries at other stages of development. We take this ability to parsimoniously account for structural transformation in a variety of contexts as evidence in favor of our model.[6]

Finally, we use our model to study the drivers of structural transformation. Both relative prices and income effects turn out to be significant contributors. However, in contrast to previous studies (e.g., Dennis and Iscan, 2009, Boppart, 2014a), we find that income effects are more important than sectoral substitution driven by relative price trends. A potential reason for this discrepancy is that in our framework income effects are not hard-wired to have only transitory effects on the structural transformation or to be correlated with price effects. Once we do not impose these constraints on income effects, our estimates are consistent with a predominant role of income effects in accounting for the structural transformation during the postwar period in a large sample of countries at different stages of development.

Our paper relates to a large literature that aims to quantify the role of nonhomotheticities on growth and development (see, among others, Matsuyama (1992), Echevarria, 1997, Gollin *et al.*, 2002, Duarte and Restuccia, 2010, Alvarez-Cuadrado and Poschke, 2011).[7] Buera

[6] A key parameter singled out in the literature is the price elasticity of substitution between consumption of different goods and services. Our baseline estimate of the elasticity of substitution is around 0.7. We find a very similar estimate using household level data from the Consumer Expenditure Survey (CEX), for which we can directly control for sectoral demand shocks and use an IV strategy. We also find that, compared to previous estimates based on Stone-Geary preferences (Herrendorf *et al.*, 2013), the estimate of the elasticity of substitution is more robust to using either value added or expenditure measures.

[7] An alternative formulation that can reconcile demand being asymptotically nonhomothetic with balanced

and Kaboski (2009) and Dennis and Iscan (2009) have noted the limits of the generalized Stone-Geary utility function to match long time series or cross-sections of countries with different income levels.

The paper that is the closest to ours is Boppart (2014a). Boppart studies the evolution of consumption of goods relative to services by introducing a sub-class of price-independent-generalized-linear (PIGL) preferences that also yield income effects in the long-run. There are several important differences between the PIGL preferences and nonhomothetic CES preferences. First, just like explicitly separable preferences such as Stone-Geary, PIGL preferences also presuppose specific parametric correlations for the evolution of income and price elasticities over time (Gorman, 1965). In contrast, nonhomothetic CES preferences do not build in any connection between price and income effects. Second, PIGL preferences can only accommodate two sectors with distinct income elasticities. In contrast, our framework allows for an arbitrary number of sectors.[8] These differences between the two models are further reflected in their empirical implications. Whereas we find a larger contribution for demand nonhomotheticity in accounting for structural change, Boppart concludes that supply and demand make roughly similar contributions.[9]

The rest of the paper is organized as follows. Section 1.2 presents the model. Section 1.3 contains the estimation and model evaluation for a panel of 25 countries for the period 1947-2005. Section 1.4 analyzes household expenditure data and aggregate macroeconomic time series for the United States. Section 1.5 concludes. Appendix A.1 presents some general properties of nonhomothetic CES. All proofs are in Appendix C.2.

---

growth path is given by hierarchical preferences (e.g., Foellmi and Zweimüller, 2006, 2008 and Foellmi *et al.*, 2014). In recent work, Święcki (2014) estimates a demand system that features non-vanishing income effects in combination with subsistence levels à la Stone-Geary. However, his demand system also imposes a parametric relation between income and price effects.

[8]One can extend PIGL preferences to more than two sectors by nesting other functions, e.g., CES aggregators, as composites within the two-good utility function (Boppart, 2014a). However, the resulting utility function does not allow for heterogeneity in income elasticity among the goods within each nested composite.

[9]In terms of the scope of the empirical exercise, while Boppart (2014a) estimates his model with U.S. data and considers two sectors, the empirical evaluation of our model includes, in addition to the U.S., a wide range of other rich and developing countries and more than two sectors.

1.2 Model

In this section, we develop the model that guides our empirical investigation of structural transformation in Sections 1.3 and 1.4, and characterize its asymptotic properties. The model closely follows workhorse models of structural transformation (e.g., Buera and Kaboski, 2009; Herrendorf *et al.*, 2013, 2014). We replace the standard aggregators of sectoral consumption goods with a nonhomothetic CES aggregator. This single departure from the standard workhorse model delivers the main theoretical results of the paper and the demand system later used in the estimation. On the production side, the model combines two distinct potential drivers of sectoral reallocation previously highlighted in the literature: heterogeneous rates of technological growth (Ngai and Pissarides, 2007a) and heterogeneous capital-intensity across sectors (Acemoglu and Guerrieri, 2008). We show that our empirical framework can account for both of these supply-side channels through the price effect.

1.2.1 Preferences and the Household Problem

A representative household has the following intertemporal preferences over goods and services produced in I different sectors

$$\sum_{t=0}^{\infty} \beta^t \left(\frac{C_t^{1-\theta} - 1}{1-\theta} \right), \tag{1.1}$$

where $\beta \in (0,1)$ is the discount factor, and θ is the reciprocal of the elasticity of intertemporal substitution. Aggregate consumption, C_t, combines sectoral goods, $\{C_{it}\}_{i=1}^{I}$, according to the implicitly defined function

$$\sum_{i=1}^{I} \Omega_i^{\frac{1}{\sigma}} C_t^{\frac{\epsilon_i - \sigma}{\sigma}} C_{it}^{\frac{\sigma-1}{\sigma}} = 1, \tag{1.2}$$

where $\sigma \in (0,1)$ is the elasticity of substitution, and Ω_i's are constant weights for all $i \in \mathcal{I} \equiv \{1, \ldots, I\}$.[10] Each sectoral good i is identified with a parameter $\epsilon_i \geq 1$, which

[10] We focus in the empirically relevant case $\sigma \in (0,1)$ (gross complements). However, the preferences are also well-defined when $\sigma > 1$ (gross substitutes). In general, if $\sigma > 0$ and $\Omega_i > 0$ for all $i \in \mathcal{I}$ and if $\epsilon_i > \sigma$ when $0 < \sigma < 1$, or $\epsilon_i < \sigma$ when $\sigma > 1$, then the aggregator C_t introduced in equation (1.2) is globally monotonically increasing and quasi-concave, yielding a well-defined utility function over the bundle of goods $(C_{1t}, \cdots, C_{It})$,

is a measure of the income elasticity of demand for that good. Equation (1.2) introduces a nonhomothetic generalization of the standard Constant Elasticity of Substitution (CES) aggregator, which corresponds to the special case where $\epsilon_i = 1$ for all sectors. Intuitively, as aggregate consumption C_t increases, the weight given to the consumption of good i varies at a rate controlled by parameter ϵ_i. As a result, the household's demand for sectoral good i features a constant elasticity in terms of the aggregate consumption C_t, which is in turn determined by household income.

A number of unique features of the nonhomothetic CES aggregator makes it a natural choice for our model. In particular, consider the static expenditure minimization problem with sectoral prices $\{p_i\}_{i=1}^{I}$ and aggregate consumption C_t defined as in equation (1.2). The resulting demand function has the following properties.

1. The elasticity of the relative demand for two different goods with respect to aggregate consumption is constant, i.e.,

$$\frac{\partial \log\left(C_i/C_j\right)}{\partial \log C} = \epsilon_i - \epsilon_j. \tag{1.3}$$

2. The elasticity of substitution between goods of different sectors is uniquely defined and constant[11]

$$\frac{\partial \log\left(C_i/C_j\right)}{\partial \log\left(P_j/P_i\right)} = \sigma. \tag{1.4}$$

The first property ensures that the nonhomothetic features of these preferences do not systematically vary as income grows. As discussed in the introduction and in Section 1.4, available data on sectoral consumption, both at the macro and micro levels, suggest stable

see Hanoch (1975a). The additional restriction $\epsilon_i \geq 1$ ensures strict concavity, which simplifies the analysis of the dynamics below.

[11]Note that for preferences defined over I goods when $I > 2$, alternative definitions for elasticity of substitution do not necessarily coincide. In particular, equation (1.4) defines the so-called Morishima elasticity of substitution, which in general is not symmetric . This definition may be contrasted from the Allen (or Allen-Uzawa) elasticity of substitution defined as $\frac{E \cdot \partial C_i/\partial P_j}{C_i C_j}$, where E is the corresponding value of expenditure. Blackorby and Russell (1981) prove that the only preferences for which the Morishima elasticities of substitution between any two goods are symmetric, constant, and identical to Allen-Uzawa elasticities have the form of equation (1.2), albeit with a more general dependence of weights on C.

and heterogeneous income elasticities across sectors. Therefore, we find it reasonable to specify preferences that do not result in systematically vanishing patterns of nonhomotheticity, as, for instance, would be implied by the choice of Stone-Geary preferences. Similarly, the second property ensures that the patterns of inter-sectoral substitution have a constant price elasticity and, thus, do not systematically vary as income grows. This property is unique to this class of nonhomothetic CES preferences.[12] The combination of these two properties uniquely defines the aggregator in equation (1.2).

To complete the characterization of the household behavior, we assume that the representative household inelastically supplies one unit of perfectly divisible labor, and starts at period 0 with an initial endowment $\mathcal{A}_0$ of assets. The household takes the sequence of wages, real interest rates, and prices of goods and services $\left\{w_t, r_t, \{p_{it}\}_{i=1}^{I}\right\}_{t=0}^{\infty}$ as given, and chooses a sequence of assets stocks $\{\mathcal{A}_t\}_{t=1}^{\infty}$ and aggregate consumption $\{C_t\}_{t=0}^{\infty}$ to maximize its utility defined in Equation (1.1), subject to the per-period budget constraint

$$\mathcal{A}_{t+1} + \sum_{i=1}^{I} p_{it} C_{it} \leq w_t + (1 + r_t)\, \mathcal{A}_t, \tag{1.5}$$

where we have normalized the price of assets to 1. The next lemma characterizes the solution to the household problem.

Lemma 1. *(Household Behavior) Consider a household with preferences and budget constraint as described by equations (1.1), (1.2), (1.5), and the No-Ponzi condition* $\lim_{t\to\infty} \mathcal{A}_t \left(\prod_{t'=1}^{t-1} \frac{1}{1+r_{t'}}\right) = 0$. *Given a sequence of prices* $\left\{w_t, r_t, \{p_{it}\}_{i=1}^{I}\right\}_{t=0}^{\infty}$ *and an initial stock of assets* $\mathcal{A}_0$, *the problem has a unique solution, fully characterized by the following conditions.*

1. *The intratemporal allocation of consumption goods satisfies*

$$C_{it} = \Omega_i \left(\frac{p_{it}}{P_t}\right)^{-\sigma} C_t^{\epsilon_i}, \tag{1.6}$$

[12] Nonhomothetic CES preferences inherit this property because they belong to the class of implicitly additively separable preferences (Hanoch, 1975a). In contrast, any preferences that are explicitly additively separable in sectoral goods imply parametric links between income and substitution elasticities (this result is known as Pigou's Law, see Snow and Warren (2015) and the references therein). Appendix A.1 illustrates how such links appear in specific case of Stone-Geary and price-independent generalized linear (PIGL) preferences, two types of specifications recently used in studies of structural change.

where P_t is the aggregate price index

$$P_t \equiv \frac{E_t}{C_t} = \frac{1}{C_t}\left[\sum_{i=1}^{I} \Omega_i C_t^{\epsilon_i - \sigma} p_{it}^{1-\sigma}\right]^{\frac{1}{1-\sigma}}, \tag{1.7}$$

and $E_t \equiv \sum_{i=1}^{I} p_{it} C_{it}$ denotes consumption expenditure at time t.

2. *The intertemporal allocation of real aggregate consumption satisfies the Euler equation*

$$\left(\frac{C_{t+1}}{C_t}\right)^{-\theta} = \frac{1}{\beta\left(1+r_t\right)} \frac{P_{t+1}}{P_t} \frac{\bar{\epsilon}_{t+1} - \sigma}{\bar{\epsilon}_t - \sigma}, \tag{1.8}$$

and the transversality condition

$$\lim_{t\to\infty} \beta^t \left(1+r_t\right) \frac{\mathcal{A}_t}{E_t} C_t^{1-\theta} \frac{1-\sigma}{\bar{\epsilon}_t - \sigma} = 0, \tag{1.9}$$

where we have defined $\bar{\epsilon}_t \equiv \sum_{i=1}^{I} \omega_{it} \epsilon_i$ with ω_{it} denoting the expenditure share in sector i, $p_{it} C_{it} / E_t$.

The key insight from Lemma 1 is that the household problem can be decomposed into two sub-problems: one involving the allocation of consumption and savings over time, and one involving the allocation of consumption across sectors. First, consider the *intertemporal* consumption-savings problem. The household solves for the sequence of $\{\mathcal{A}_{t+1}, C_t\}_{t=0}^{\infty}$ that maximizes utility (1.1) subject to the constraint

$$\mathcal{A}_{t+1} + E\left(C_t; \{p_{it}\}_{i=1}^{I}\right) \leq w_t + \mathcal{A}_t\left(1+r_t\right), \tag{1.10}$$

where $E\left(C_t; \{p_{it}\}_{i=1}^{I}\right)$ is the total expenditure function for the nonhomothetic CES preferences, defined in equation (1.7). Because of nonhomotheticity, consumption expenditure is a nonlinear function of real aggregate consumption, and the price index reflects changes in the sectoral composition of consumption as income grows.[13] The household incorporates this relationship in its Euler equation (1.8), where we see a wedge between the marginal

[13] An equivalent definition for the price index in terms of total and sectoral expenditure is

$$P_t^{1-\sigma} = \sum_i \left(\Omega_i E_t^{\epsilon_i - 1} p_{it}^{1-\sigma}\right)^{\frac{1-\sigma}{\epsilon_i - \sigma}} \omega_{it}^{\frac{\epsilon_i - 1}{\epsilon_i - \sigma}}. \tag{1.11}$$

cost of real consumption and the aggregate price index. The size of this wedge, given by $(\bar{\epsilon}_t - \sigma) \,/\, (1-\sigma)$, depends on the average income elasticities across sectors, $\bar{\epsilon}_t = \sum_{i=1}^{I} \omega_{it}\epsilon_i$, and varies over time. In the case of homothetic CES where $\epsilon_i \equiv 1$, this wedge disappears.

The second part of the household problem involves the *intratemporal* problem of allocating consumption across different goods. Equation (1.6) corresponds to the sectoral demand implied by the nonhomothetic CES aggregator. Lemma 1 establishes that given aggregate consumption, C_t, allocated to period t, sectoral demand simply follows the solution to the static allocation problem.

Note that Equation (1.6) restates the two main features of the nonhomothetic CES aggregator expressed in equations (1.3) and (1.4): constant and independent elasticities of income and substitution for different goods. We can rewrite this relation in terms of the logarithm of relative real consumption and consumption expenditure shares between sectors i and j as a function of the logarithm of relative prices and aggregate consumption,

$$\log\left(\frac{C_{it}}{C_{jt}}\right) = -\sigma \log\left(\frac{p_{it}}{p_{jt}}\right) + (\epsilon_i - \epsilon_j)\log C_t + \log\left(\frac{\Omega_i}{\Omega_j}\right), \tag{1.12}$$

$$\log\left(\frac{\omega_{it}}{\omega_{ji}}\right) = (1-\sigma)\log\left(\frac{p_{it}}{p_{jt}}\right) + (\epsilon_i - \epsilon_j)\log C_t + \log\left(\frac{\Omega_i}{\Omega_j}\right), \tag{1.13}$$

where the second equation simply states the relationship in terms of expenditure shares, which plays an important role in our theory, as we will see later. Equation (1.12) once again highlights the key features of the demand system implied by this nonhomothetic CES preferences. Interpreting C_{it} as the Hicksian demand for good i with aggregate consumption C_t under prices p_{it}'s, we find a perfect separation of the price and the income effects. The first term on the right hand side shows the price effects characterized by constant elasticity of substitution σ. More interestingly, the second term on the right hand side shows the change in relative sectoral demand as consumers move across indifference curves. This income effect is characterized by constant sectoral income elasticity parameters ϵ_i's. If

$\epsilon_i > \epsilon_j$, demand for good i rises relative to good j as consumers become wealthier.[14]

Equations (1.12) and (1.13) also show how our model can generate a positive correlation between relative sectoral consumption in real and expenditure terms, as it is observed in the data. As in the case with homothetic CES aggregators, the combination of the price effect and gross complementarity ($\sigma < 1$) imply that relative real sectoral consumption should negatively correlate with relative sectoral prices. To see why, note that relative real consumption is decreasing in relative prices with an elasticity of $-\sigma$, while relative expenditure is increasing with an elasticity of $1 - \sigma$. However, our demand system has an additional force, income effects, which makes both time series co-move in aggregate consumption. Thus, if income effects are sufficiently strong, both time series can be positively correlated. In Section 1.3.4 we show that this is the case when we estimate our demand system.

1.2.2 Production and Competitive Equilibrium

The supply side of the economy allows for two distinct sources of heterogeneity in sectoral production. Our model combines the heterogeneous sectoral productivity growth framework of Ngai and Pissarides (2007a) with the heterogeneous sectoral factor intensity model of Acemoglu and Guerrieri (2008).

A representative firm in each consumption sector produces sectoral output under perfect competition. In addition, a representative firm in a perfectly competitive investment sector produces investment good, Y_{0t}, that is used in the process of capital accumulation. We

[14]The expenditure elasticity of demand for sectoral good i is given by

$$\eta_{it} \equiv \frac{\partial \log C_{it}}{\partial \log E_t} = 1 + \frac{1-\sigma}{\bar{\epsilon}_t - \sigma}\left(\epsilon_i - \bar{\epsilon}_t\right), \tag{1.14}$$

which, as Engel aggregation requires, averages to 1 when sectoral weights are given by expenditure shares. If sector i has an income elasticity parameter, ϵ_i, that exceeds the economy-wide average elasticity parameter, $\bar{\epsilon}_t$, at time t, then sector i has an expenditure elasticity greater than 1 at that point in time. The expenditure elasticity of relative demand is

$$\frac{\partial \log\left(C_{it}/C_{jt}\right)}{\partial E_t} = \frac{\epsilon_i - \epsilon_j}{\bar{\epsilon}_t - \sigma}, \tag{1.15}$$

which parallels equation (1.3) now expressed in terms of expenditure, rather than real aggregate consumption.

assume Cobb-Douglas production functions with time-varying Hicks-neutral sector-specific productivities,

$$Y_{it} = A_{it} K_{it}^{\alpha_i} L_{it}^{1-\alpha_i}, \qquad i \in \{0\} \cup \mathcal{I},$$

where K_{it} and L_{it} are capital and labor used in the production of output Y_{it} in sector i at time t (we have identified the sector producing investment good as $i = 0$) and $\alpha_i \in (0, 1)$ denotes sector-specific capital intensity. The aggregate capital stock of the economy, K_t, accumulates using investment goods and depreciates at rate δ, $Y_{0t} = K_{t+1} - (1 - \delta) K_t$.

We focus on the features of the competitive equilibrium of this economy that motivate our empirical specifications.[15] Firm profit maximization and equalization of the prices of labor and capital across sectors pin down prices of sectoral consumption goods,

$$p_{it} = \frac{p_{it}}{p_{0t}} = \frac{\alpha_0^{\alpha_0} (1 - \alpha_0)^{1-\alpha_0}}{\alpha_i^{\alpha_i} (1 - \alpha_i)^{1-\alpha_i}} \left(\frac{w_t}{R_t} \right)^{\alpha_0 - \alpha_i} \frac{A_{0t}}{A_{it}}, \tag{1.16}$$

where, since the units of investment good and capital are the same, we normalize the price of investment good, $p_{0t} \equiv 1$. Equation (1.16) shows that price effects capture both supply-side drivers of sectoral reallocation: heterogeneity in productivity growth rates and heterogeneity in capital intensities.

Goods market clearing ensures that household sectoral consumption expenditure equals the value of sectoral production output, $P_{it} C_{it} = P_{it} Y_{it}$.[16] Competitive goods markets and profit maximization together imply that a constant share of sectoral output is spent on the wage bill,

$$L_{it} = (1 - \alpha_i) \frac{P_t C_t}{w_t} \omega_{it}, \tag{1.17}$$

where ω_{it} is the share of sector i in household consumption expenditure.

---

[15] Given initial stock of capital K_0 and a sequence of sectoral productivities $\left\{ \{A_{it}\}_{i=1}^{I} \right\}_{t \geq 0}$, a competitive equilibrium is defined as a sequence of allocations $\left\{ C_t, K_{t+1}, Y_{0t}, L_{0t}, K_{0t}, \{Y_{it}, C_{it}, K_{it}, L_{it}\}_{i=1}^{I} \right\}_{t \geq 0}$ and a sequence of prices $\left\{ w_t, R_t, \{p_{it}\}_{i=1}^{I} \right\}_{t \geq 0}$ such that *(i)* agents maximize the present discounted value of their utility given their budget constraint, *(ii)* firms maximize profits and *(iii)* markets clear.

[16] In our empirical applications, we account for sectoral trade flows.

Equation (1.17), together with equations (1.13) and (1.16) summarize the main insights from the theory that we employ in our empirical strategy. First, equation (1.17) implies

$$\frac{L_{it}}{L_{jt}} = \frac{1-\alpha_i}{1-\alpha_j}\frac{\omega_{it}}{\omega_{jt}}, \qquad i,j \in \mathcal{I}. \tag{1.18}$$

Equation (1.18) shows that the paths of relative sectoral employment shares follow those of relative consumption expenditure shares. Second, equation (1.13) characterizes the paths of relative consumption expenditure shares as a function of relative prices and aggregate real consumption. Thus, equations (1.13) and (1.18) together predict the evolution of relative employment shares across sectors. We will use these two equations extensively in our empirical exercise in the next two sections.

1.2.3 Constant Growth Path

Before moving into the empirical sections, we characterize the asymptotic dynamics of the economy when sectoral total factor productivities grow at heterogeneous but constant rates. In particular, let us assume that sectoral productivity growth is given by

$$\frac{A_{it+1}}{A_{it}} = 1+\gamma_i, \qquad i \in \{0\} \cup \mathcal{I}. \tag{1.19}$$

Under this assumption, the competitive equilibrium of the economy converges to a path of constant per-capita consumption growth. Along this path, nominal consumption, investment, and the stoack of capital all grow at a rate disctated by the rate of growth of the investment sector γ_0. Denoting the rate of growth of real consumption by γ^*, it is straightforward to show that along a constant growth path the share of each sector i in consumption expenditure also exhibits constant growth, characterized by constants

$$1+\xi_i \equiv \lim_{t\to\infty}\frac{\omega_{it+1}}{\omega_{it}} = \frac{(1+\gamma^*)^{\epsilon_i-\sigma}}{\left[(1+\gamma_0)^{\frac{\alpha_i}{1-\alpha_0}}(1+\gamma_i)\right]^{1-\sigma}}. \tag{1.20}$$

Given the fact that expenditures shares have to be positive and sum to 1, Equation (1.20) allows us to find the rate of growth of real consumption as a function of sectoral income

elasticity, factor intensity, and the rates of technical growth. The next proposition presents these results that characterize the asymptotic dynamics of the competitive equilibrium.

Proposition 1. *Let γ^* be defined as*

$$\gamma^* = \min_{i \in \mathcal{I}} \left[(1+\gamma_0)^{\frac{\alpha_i}{1-\alpha_0}} (1+\gamma_i) \right]^{\frac{1-\sigma}{\epsilon_i - \sigma}} - 1. \tag{1.21}$$

Assume that γ^* satisfies the following condition

$$(1+\gamma_0)^{-\frac{\alpha_0}{1-\alpha_0}} < \beta (1+\gamma^*)^{1-\theta} < \min \left\{ \frac{(1+\gamma_0)^{-\frac{\alpha_0}{1-\alpha_0}}}{\alpha_0 + (1-\alpha_0)(1+\gamma_0)^{-\frac{1}{1-\alpha_0}}(1-\delta)}, 1 \right\}. \tag{1.22}$$

Then, for any initial level of capital stock, K_0, there exists a unique competitive equilibrium along which consumption asymptotically grows at rate γ^*,[17]

$$\lim_{t \to \infty} \frac{C_{t+1}}{C_t} = 1 + \gamma^*. \tag{1.23}$$

Along the this constant growth path, (i) the real interest rate is constant, $r^* \equiv (1+\gamma_0)^{1/(1-\alpha_0)}/\beta(1+\gamma^*)^{1-\theta} - 1$, (ii) nominal expenditure, total nominal output, and the stock of capital grow at rate $(1+\gamma_0)^{\frac{1}{1-\alpha_0}}$, (iii) only the subset of sectors $\mathcal{I}^*$ that achieve the minimum in equation (1.21) employ a non-negligible fraction of workers.

Equation (1.21) shows how the long-run growth rate of consumption is affected by income elasticities, ϵ_i, rates of technological progress, γ_i, and sectoral capital intensities, α_i. To build intuition, consider the case in which all sectors have the same capital intensity, and preferences are homothetic. Then, since $\sigma \in (0,1)$, equation (1.21) implies that the long-run growth rate of real consumption is pinned down by the sectors with the lowest technological progress, as in Ngai and Pissarides (2007a). Consider now the case in which there is also heterogeneity in income elasticities. In this case, sectors with higher income elasticity and higher technological progress can co-exist in the long-run with sectors with low technological progress and low income elasticity. The intuition is that, as agents become richer, they want to consume more goods that have higher technological progress, as they are

[17] Here we follow the terminology of Acemoglu and Guerrieri (2008) in referring to our equilibrium path as a constant growth path. Kongsamut *et al.* (2001) refer to this concept as generalized balanced growth path.

more income-elastic. Finally, the role of heterogeneity in capital shares also affects the long-run equilibrium through its effect on relative prices in an analogous way to technological progress.

Which sectors survive in the long run? At all points in time, all sectors produce a positive amount of goods, and its production grows over time. In relative terms, however, only the subset of sectors $\mathcal{I}^*$ satisfying equation (1.21) will comprise a non-negligible share of total consumption expenditure in the long-run.

We study the transitional dynamics of the economy in the special case where all sectors have a common capital intensity $\alpha \equiv \alpha_i$ for all i. Let us normalize each of the aggregate variables by their respective rates of growth, introducing normalized consumption expenditure $\tilde{E}_t \equiv (1+\gamma_0)^{-\frac{t}{1-\alpha}} E_t$, per-capita stock of capital $\tilde{k}_t \equiv (1+\gamma_0)^{-\frac{t}{1-\alpha}} K_t/L_t$, and real per-capita consumption $\tilde{C}_t \equiv (1+\gamma^*)^{-t} C_t$. Using the assets market clearing condition, we can translate Equations (1.8) and (1.10) into equations that characterize the evolution of the normalized aggregate variables

$$\tilde{k}_{t+1} = (1+\gamma_0)^{-\frac{1}{1-\alpha}} \left[\tilde{k}_t^{\alpha} - \tilde{E}_t + \tilde{k}_t (1-\delta)\right], \tag{1.24}$$

$$\left(\frac{\tilde{C}_{t+1}}{\tilde{C}_t}\right)^{1-\theta} = \frac{1+r^*}{1+\alpha \tilde{k}_t^{\alpha-1} - \delta} \frac{\tilde{E}_{t+1}}{\tilde{E}_t} \frac{\overline{\epsilon}_{t+1} - \sigma}{\overline{\epsilon}_t - \sigma}, \tag{1.25}$$

where the normalized consumption expenditure $\tilde{E}_t$ is a function of $\tilde{C}_t$ and the two functions

$$\left(\frac{\tilde{E}_{t+1}}{\tilde{E}_t}\right)^{1-\sigma} = \sum_{i=1}^{I} \omega_{it} \left(\frac{\tilde{C}_{t+1}}{\tilde{C}_t}\right)^{\epsilon_i - \sigma} (1+\xi_i), \tag{1.26}$$

$$\frac{\overline{\epsilon}_{t+1} - \sigma}{\overline{\epsilon}_t - \sigma} = \left(\frac{\tilde{E}_t}{\tilde{E}_{t+1}}\right)^{1-\sigma} \sum_{i=1}^{I} \omega_{it} \left(\frac{\tilde{C}_{t+1}}{\tilde{C}_t}\right)^{\epsilon_i - \sigma} (1+\xi_i) \left(\frac{\epsilon_i - \sigma}{\overline{\epsilon}_t - \sigma}\right), \tag{1.27}$$

depend both on $\tilde{C}_t$ and $\tilde{C}_{t+1}$. The proposition establishes that the equilibrium path exists, is unique, and is therefore fully characterized by the dynamic equations above.

1.3 Quantitative Exploration of a Cross-Country Panel

In this section we explore the ability of nonhomothetic CES preferences to account for the broad patterns of structural transformation observed across countries in agriculture,

manufacturing and services during the postwar period. We discipline our model by using the fact that the same parameters of the utility function $\{\sigma, \epsilon_i\}_{i\in\mathcal{I}}$ for all countries. After estimating these parameters, we gauge the ability of our model to account for the very different experiences of advanced, miracle and developing economies. We conclude the section by conducting a battery of exercises that revisit critical findings in the structural change literature, and extending our analysis to more disaggregated sectoral data.

1.3.1 Empirical Strategy

Our empirical strategy uses the solution of the intratemporal problem and the production decisions of firms to estimate the preference parameters of the nonhomothetic CES aggregator (1.2). Taking the logarithms of the sectoral demands (1.13) and using that the ratio of sectoral expenditures is proportional in equilibrium to the ratio of sectoral labor allocations, (1.18), we obtain our estimating equations

$$\log\left(\frac{L^c_{a,t}}{L^c_{m,t}}\right) = \zeta^c_{am} + (1-\sigma)\log\left(\frac{p^c_{a,t}}{p^c_{m,t}}\right) + (\epsilon_a - \epsilon_m)\log C^c_t + \nu^c_{am,t}, \tag{1.28}$$

$$\log\left(\frac{L^c_{s,t}}{L^c_{m,t}}\right) = \zeta^c_{sm} + (1-\sigma)\log\left(\frac{p^c_{s,t}}{p^c_{m,t}}\right) + (\epsilon_s - \epsilon_m)\log C^c_t + \nu^c_{sm,t}, \tag{1.29}$$

where a, m and s denote agriculture, manufacturing and services, respectively, and t, time. The superscript c denotes a country, and $\nu^c_{am,t}$ and $\nu^c_{sm,t}$ are the error terms. We allow for country-sector dyad fixed effects, ζ^c_{am} and ζ^c_{sm}, as there may be systematic differences in measurement across countries. These country-sector dyad fixed effects also absorb potential cross-country differences in sectoral taste parameters, Ω^c_i, and differences in factor shares in the production function, α_i. Note that there are two cross-equation restrictions. The price elasticity σ is restricted to be the same across sectors and countries. Income elasticities, ϵ_i, are also restricted to be the same across countries for a given sector s.

To construct our dependent variable, we could either use expenditure shares or employment shares. In the baseline specification, we use employment shares. This allows us to circumvent the problem that the prices that we use as regressors also enter expenditure

shares, which could introduce bias.[18] To account for the fact that some goods can be imported and exported, thus affecting the sectoral composition of employment, we control for the share of net sectoral exports over total production in sector i, time t and country c.[19]

Identification The identification strategy relies on the intra-period allocation of consumption that follows from the solution of the intratemporal allocation problem (1.13). That is, conditional on the observed levels of aggregate consumption C_t^c and sectoral prices p_{it}^c, we use our demand system to estimate relative consumption across sectors. Given the presence of country-sector dyad fixed effects, ζ_{am}^c, ζ_{sm}^c, the relevant variation used to identify the elasticities is the within country-sector time variation. To the extent that we have a long time series for C_t^c and p_{it}^c, we have a super-consistent estimator of the elasticities (Hamilton, 1994).

Changes over time in aggregate and sectoral productivity contribute to the identification of the price and income elasticities. For example, sectoral productivity growth differences affect relative prices and introduce variation in the estimating equations, (1.28) - (1.29).[20] Sectoral and aggregate productivity shocks can also affect the level of total consumption, thereby introducing additional variation in the estimation through C_t^c.

Aggregate demand shocks, such as a rise in the propensity to spend, are captured through the aggregate consumption term $\log C_t^c$ in (1.28) - (1.29) and also contribute to the identification of our demand system.[21] However, sectoral taste shocks that induce consumers

[18] In practice, as we discuss below, we obtain similar estimates with either dependent variable.

[19] This particular specification of the trade controls follows from our theoretical model. To derive this result, note that $p_{it}^c C_{it}^c = p_{it}^c Y_{it}^c - NX_{it}^c$, where NX_{it}^c denotes the nominal value of net exports in sector i, time t and country c. Using equation (1.17), the amount of labor needed to produce the amount consumed in sector s needs to be adjusted by $NX_{it}^c / p_{it}^c Y_{it}^c$. We have also used more reduced-form controls, such as controlling directly for net exports or exports and imports separately, obtaining similar results.

[20] Using the log of the ratio of sectoral prices in our estimation has the advantage that we can directly use nominal prices and any cross-country systematic difference in the measurement of prices is going to be captured in the fixed effect.

[21] In particular, suppose that the preference term for sector i, Ω_i, in (1.2) becomes stochastic rather than constant, $\tilde{\Omega}_{it} = \Omega_i + \beth_t$ with $\beth_t$ being white noise. Our identification strategy remains valid in this more general case, as it relies on the within period problem, which remains unaltered.

to spend more in one sector for a given level of aggregate consumption and sectoral prices are not well-captured in this specification. Given that we have already country-sector dyad fixed effects, we cannot add an additional time fixed effect that would control for preference shocks. To the extent that these shocks are uncorrelated with other type of shocks, they enter as classical measurement error in the estimating equations (1.28) − (1.29), and our estimates are still consistent. If, on the other hand, sectoral preference shocks are correlated with other shocks (e.g., aggregate demand or productivity shocks) our estimation is going to produce biased estimates.

We deal with potential biases coming from sectoral taste shocks in two different ways. First, we estimate the elasticities separately for OECD and non-OECD countries and show that we cannot reject the null that they are statistically the same. While the estimates could in principle be biased in both cases, this would require sectoral taste shocks to be correlated with aggregate demand or productivity shocks in the same way for these two groups of countries, which we deem unlikely. Second, in Section 1.4.1 we use household-level data, which allows us to include sector-year fixed effects that absorb sectoral demand shocks and to use an IV strategy. Reassuringly, we find that the estimates for the elasticity of substitution and of the income elasticities are similar to our estimates from aggregate data.

1.3.2 Data Description

We use the GGDC 10-Sector Database for sectoral value added data (Vries *et al.*, 2014). It provides a long-run internationally comparable dataset on sectoral measures for 10 countries in Asia, 9 in Europe, 9 in Latin America, 10 in Africa and the United States. The variables covered in the data set are annual series of production value added (nominal and real) and employment for 10 broad sectors starting in 1947. In our baseline exercise we aggregate the ten sectors into agriculture, manufacturing and services.[22] In Section 1.3.6, we estimate our

[22]The ten sectors are agriculture, mining, manufacturing, construction, public utilities, retail and wholesale trade, transport and communication, finance and business services, other market services and government services. We classify as manufacturing: mining, manufacturing and construction, while the rest are classified as services (except agriculture). Figures A.2 and A.3 depict the time series for each country.

model for 10 sectors.

For real consumption, we use the time series on consumption per capita from the Barro-Ursua Macroeconomic Data.[23] Their data has the advantage of using the Fisher chained price index, which allows us to have a meaningful measure of real consumption.[24] These data do not include government services into consumption, which we exclude from our estimation. Also, the only African country covered is South Africa. Our final sample consists of 25 countries that span very different growth trajectories during the postwar period.[25]

1.3.3 Estimation Results

We estimate jointly (1.28) and (1.29) imposing the cross-equation restrictions that price and income elasticities have to be the same across countries. Table 1.1 reports the results of estimating the system of equations for the whole sample of 25 countries and for OECD and non-OECD countries separately.

Columns (1) to (3) in Table 1.1 report our estimates for the entire sample of countries. Column (1) reports the estimates without using country-sector fixed effects (and thus using cross-country variation in levels to identify the parameters). Column (2) reports our estimates using country-sector fixed effects (whereby using only within country variation to identify the elasticities) and column (3) includes trade controls. Our estimates of the price elasticity of substitution across sectors range from 0.66 to 0.75 and are precisely estimated in all three specifications (standard errors are clustered at the country level). In fact, we cannot

[23] It can be obtained at http://scholar.harvard.edu/barro/publications/barro-ursua-macroeconomic-data

[24] Fisher price indices approximate to second order the ideal price index of any continuously differentiable utility, thus approximating ours. We have also run Monte-Carlo simulations to asses the extent of the bias induced by using this approximation around the estimated parameters. We find that the error is less than 1% of the estimated parameters. Deaton and Muellbauer (1980) find the same result for the AIDS demand system: using the exact form of the price index or an approximation by a superlative price index makes little difference for the estimation results. In light of the small bias introduced, we prefer using the linear estimation model to a non-linear estimation of the price index, for which we would have an incidental parameters problem (see Wooldridge, 2001).

[25] These are Denmark, France, Italy, the Netherlands, Spain, Sweden, UK, USA, West Germany, India, Indonesia, Japan, South Korea, Malaysia, Philippines, Singapore, Taiwan, Argentina, Brazil, Chile, Colombia, Mexico, Peru, Venezuela and South Africa.

Table 1.1: *Baseline Estimates for the Cross-Country Sample*

Dep. Var.:	World			OECD		Non-OECD	
Rel. Emp.	(1)	(2)	(3)	(4)	(5)	(6)	(7)
σ	0.66	0.75	0.75	0.69	0.71	0.69	0.71
	(0.19)	(0.11)	(0.10)	(0.17)	(0.16)	(0.08)	(0.07)
$\varepsilon_a - \varepsilon_m$	-0.81	-1.09	-1.04	-0.99	-0.89	-1.16	-1.10
	(0.24)	(0.10)	(0.10)	(0.19)	(0.21)	(0.14)	(0.11)
$\varepsilon_s - \varepsilon_m$	0.32	0.32	0.32	0.40	0.51	0.23	0.23
	(0.08)	(0.10)	(0.10)	(0.19)	(0.17)	(0.09)	(0.09)
Obs.	1006	1006	1005	436	436	570	569
R^2	0.14	0.83	0.83	0.77	0.80	0.84	0.85
$c \cdot sm$ FE	N	Y	Y	Y	Y	Y	Y
Trade Controls	N	N	Y	N	Y	N	Y

Note: Standard errors clustered by country.

reject the null that these three estimates are statistically identical. However, we can reject the null that they are equal to one at 5%, implying that these three sectors are complements. The estimates for the difference of income elasticities yield sensible results that are very stable across the three specifications and significant at conventional levels. We find that the difference in income elasticities between agriculture and manufacturing, $\epsilon_a - \epsilon_m$, is negative and ranges from -1.09 to -0.81, while the difference between services and manufacturing, $\epsilon_s - \epsilon_m$, is positive around 0.32 in all three specifications.[26]

As we have discussed, sectoral taste shocks that are systematically correlated with other shocks can be a threat to the consistency of our estimates of elasticity parameters. To mitigate this concern, we report the estimated elasticities for OECD countries in columns (5) and (6) and non-OECD countries in columns (7) and (8) with and without trade controls, respectively. We find that the estimates of all elasticities are similar for the two sub-samples. In fact, we cannot reject the null that the estimates for the elasticities are the same for both sub-samples at conventional levels. For example, the p-value of testing the joint hypothesis of price and income elasticities being different for OECD and non-OECD countries is 0.45.

[26]Note that the scale of the difference of elasticities does not matter for the real allocation of resources. The reason is that there is one degree of freedom in the definition of real income elasticities.

We take this as reassuring evidence. In the next section, we also estimate our demand system using household data, which allows us to have more demanding specifications, finding similar results.

Table 1.1 also reports the R^2 of these regressions. We find that prices and aggregate consumption, along with the country-sector fixed effects, account for 83% of the variation in our panel. The fit is equally good if we split the sample between OECD and non-OECD countries (77% and 84%, respectively). The inclusion of explicit sectoral net exports controls improves the fit to 80% for OECD countries and 85% for non-OECD countries. This modest improvement suggests that the change over time of domestic demand of domestic goods is the main driver of our results. Note, however, that this does not necessarily mean that trade plays a minor role. The reason is that that our sectoral price measures are constructed from measures which include imported inputs. Thus, part of the productivity enhancing effects of trade are reflected in our price measures.

Case Studies Figure 1.2 plots the actual and the predicted employment shares implied by our estimates from column (3) of Table 1.1 for six countries, Mexico, Colombia, Japan, Taiwan, the U.S. and Spain.[27] This figure confirms the good fit of the model despite the parsimony of using the same price and income elasticities for all countries. In particular, the model captures the evolution of employment shares in all sectors for countries at very different stages of development.

For Japan and Taiwan, we see that our fitted model generates a hump-shape for employment in manufacturing that tracks the patterns observed in the data. For Taiwan, the predicted initial level of the employment share in manufacturing is 21%, it goes up to 39% and back to 35% at the end of the period. The observed levels are 20%, 43% and 37%. For Japan, the employment share in manufacturing is 26% in the initial period, it goes up to 38% in the mid 1970s and it is 30% by the end of our sample. The fitted time series starts at

[27] Figures A.2 and A.3 show the predicted series of the employment shares and the actual time series for all countries in our sample.

26%, goes up to 35% and declines to 33% by the end of the period.[28]

To gain better intuition on the role played by relative prices, consumption and net exports, we illustrate ithe case of Japan in more detail in Figure A.1 in Appendix A.3. Panel (a) shows the overall fit of the data using the estimated parameters from the entire sample (column 3). Panels (b), (c) and (d) show the time series of relative prices, consumption and sectoral net exports. We further report the partial fit generated by each time series (and the country-sector fixed effects) at the estimated parameter values. Panel (e) shows the partial fit generated by the relative price time series. We see that, at the estimated parameter values, relative prices account for relatively little of the variation. In contrast, the evolution of aggregate consumption accounts for much of the structural transformation (see panel (f)). In particular, income effects drive the observed hump-shape in manufacturing. Intuitively, as the Japanese economy became richer in the 1950s, it reallocated labor away from agriculture to both manufacturing and services. Subsequent income growth led to the expansion of services which absorbed employment from manufacturing. Finally, panel (g) shows that changes in sectoral net exports did not play a significant role in accounting for the structural transformation in Japan.

Contribution of Relative Price and Income Effects After discussing the case of Japan, we analyze the drivers of structural transformation for our entire sample. We compute the share of the variation in employment shares that is generated by price and income effects in our fitted model.[29] We find that the variation generated by prices in the median year accounts for 14% of the employment growth in agriculture, 43% in manufacturing and 17% in services. The remaining variation is accounted for by income effects. If we

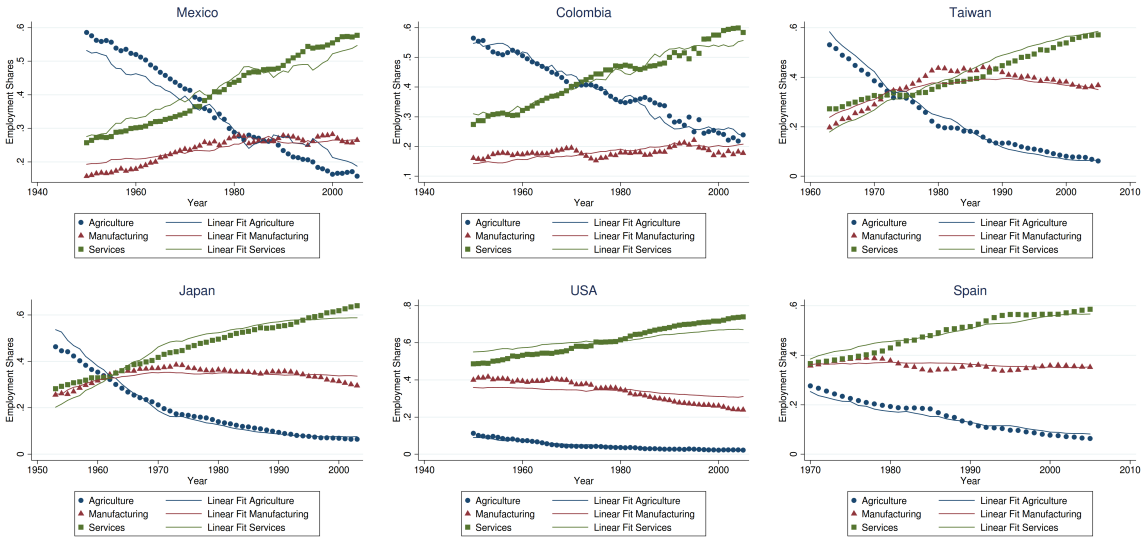
[28]For the U.S., we see that the evolution of the employment shares in services and manufacturing are steeper than predicted by our model. This is the case also for other OECD countries. This reflects the fact that the income elasticity of services is greater for these set of countries, as column (5) in Table 1.1 shows.

[29]To perform the decomposition, from the demand system (1.13), we have that the growth rate of employment in sector i relative to j is

$$\gamma_{L_i} - \gamma_{L_j} = (1-\sigma)(\gamma_{p_i} - \gamma_{p_j}) + (\epsilon_i - \epsilon_j)\gamma_C, \tag{1.30}$$

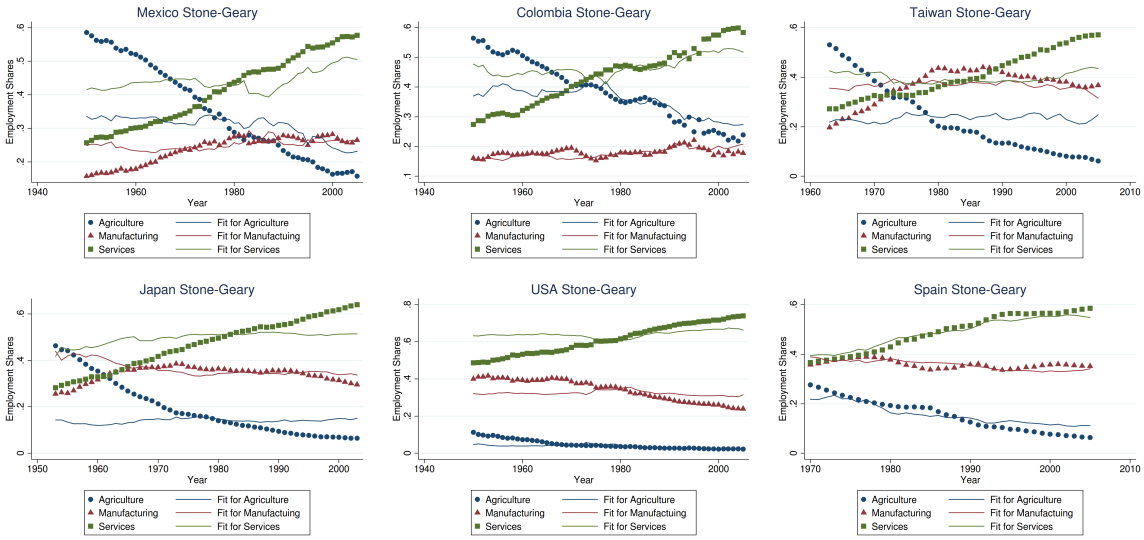
where γ_{L_i}, γ_{p_i}, γ_C denote the growth rate of the employment share in sector i, the growth rate in the price of sector i and the growth rate of aggregate consumption. We compare the relative contribution of these two terms.

Figure 1.2: *Baseline fit with common preference parameters* $\{\sigma, \epsilon_a - \epsilon_m, \epsilon_s - \epsilon_m\}$ *for six countries*



Note: Predicted fit (shown in solid lines) and actual data for six countries of the sample. Fit corresponding to the estimates for the entire sample reported in column (3) of Table 1.1.

Figure 1.3: *Stone-Geary fit with common preference parameters* $\{\sigma, \bar{c}_a, \bar{c}_s\}$ *for six countries*



Note: Predicted fit of Stone-Geary preferences (shown in solid lines) and actual data for six countries of the sample. The fit corresponds to the estimates for the entire sample. Relative to the fit of nonhomothetic CES shown in Figure 1.2, we see that the fit is overall worse, except perhaps for Spain for which it is comparable. For example, the Stone-Geary estimation fails to capture the hump-shaped evolution of manufacturing in Taiwan and Japan.

restrict attention to the OECD, we find a similar pattern with prices accounting for 19%, 39% and 20% of employment share growth in agriculture, manufacturing and services, respectively. Hence, we conclude that the bulk of the variation in sectoral employment growth is generated by income effects.

In the Appendix, we study the robustness of this result to using other approaches to assess the relative contribution of the drivers of structural transformation. Table A.2 in the appendix shows that the likelihood-ratio tests of including price and consumption data to a model with only country-sector fixed effects are significant. However, the increase in the likelihood is much higher when we add consumption data than when we add relative price data.[30]

Heterogeneous Price Elasticity of Substitution Across Sectors So far, we have assumed that the price elasticities across sectors, σ, are identical. We can test whether this identifying restriction is a good approximation of our data or whether allowing for differential price elasticities across sectors improves the fit significantly. We run the baseline regressions (1.28) and (1.29) without imposing that the coefficient on relative prices is the same across regressions,

$$\log\left(\frac{L^c_{a,t}}{L^c_{m,t}}\right) = \zeta^c_{am} + (1-\sigma_{am})\log\left(\frac{p^c_{a,t}}{p^c_{m,t}}\right) + (\epsilon_a - \epsilon_m)\log C^c_t + \nu^c_{am,t}, \tag{1.31}$$

$$\log\left(\frac{L^c_{s,t}}{L^c_{m,t}}\right) = \zeta^c_{sm} + (1-\sigma_{sm})\log\left(\frac{p^c_{s,t}}{p^c_{m,t}}\right) + (\epsilon_s - \epsilon_m)\log C^c_t + \nu^c_{sm,t}, \tag{1.32}$$

where σ_{sm} is not restricted to be equal to σ_{am}. Table A.4 in the appendix reports our estimates. We find that the price elasticities in the two regressions are very similar. For services relative

[30]This conclusion differs from Boppart (2014a) who studies the evolution of services relative to the rest of the economy in the U.S. during the postwar period and finds that the contribution of price and income effects are roughly of equal sizes. If we confine our analysis to the U.S. and lump together agriculture and manufacturing into one sector, we still find that price effects generate less than a third of the variation. The key difference between our analyses is in the demand systems. In our specification, the demand price elasticity is constant. In contrast, Boppart's demand system implies that the price elasticity of services relative to the rest of consumption is declining as the economy grows. As noted by Buera and Kaboski (2009), as relative services expenditure and value added grows at a faster rate than services' relative price, a declining price elasticity automatically increases the explanatory power of relative prices.

to manufacturing, σ_{sm}, we find a point estimate of 0.78 with standard error of 0.18 (clustered at the country level). For agriculture relative to manufacturing, σ_{am}, we find a point estimate of 0.67 with standard error of 0.12. Thus, we cannot reject the null that the coefficients on prices are statistically different from each other. Moreover, the income elasticity estimates remain unchanged. This suggests that the CES is a good approximation for analyzing these three sectors.

Estimation with Value Added Shares Some statistical agencies impute all investment employment to manufacturing, while its service component has been increasing over time (Herrendorf *et al.*, 2013). By measuring sectoral activity with employment shares, we are implicitly adopting this assumption. Following Herrendorf et al., we study the robustness of our findings to estimating the baseline regressions (1.28)- (1.29) using value added shares as dependent variables. Table A.5 in the appendix reports the estimation results. The main observation is that the estimates are robust to using value added shares as dependent variable. In particular, the estimate of the price elasticity declines insignificantly to .51 (from .75 with employment shares), the income elasticity of agriculture (relative to manufacturing) is -1.17 (vs. -1.04 with employment shares) and the income elasticity of services (relative to manufacturing) is 0.1 (vs. 0.32 with employment shares).

1.3.4 Correlation between Real and Nominal Value Added

A salient feature of structural transformation in the data is that the sectoral time-series patterns are similar regardless of whether we document them in nominal or real terms (Herrendorf *et al.*, 2014). To investigate our model's ability to account for this fact, we use our estimated preference parameters $\{\sigma, \epsilon_a - \epsilon_m, \epsilon_s - \epsilon_m\}$ from column (3) in Table 1.1 and nominal and real sectoral demands, equations (1.12) and (1.13), to generate the predicted evolution of nominal and real sectoral demands.

Table 1.2 reports the correlation between nominal and real shares both in our estimated model and in the data. We find that the model is able to generate correlations similar to

Table 1.2: *Correlation of Nominal and Real Value Added*

	Correlation	
	Data	Model
Agriculture/Manufacturing	0.95	0.93
Services/Manufacturing	0.80	0.71

Note: the correlation between real and nominal value added is generated using the estimated income and price elasticities for the entire sample reported in column (3) of Table 1.1.

the data. In particular, the correlation between the nominal and real relative demand of agricultural goods to manufactures is 0.93 in our model, while in the data it is .95. For services, the model generates a correlation of .71 while in the data it is .80.

The success in matching the correlation between nominal and real measures of activity is important. First, note that it is an out-of-sample check on our model, since our analysis has not targeted the evolution of real sectoral shares (recall that the left-hand-side of our estimating equations (1.28)-(1.29) were employment shares). More significantly, simultaneously accounting for the evolution of real and nominal sectoral shares highlights the critical importance of using a nonhomothetic CES framework. If we had used an homothetic framework, the correlation generated by the model would have been negative because the price elasticity of substitution is smaller than one.[31] In our framework, the estimated income effects are sufficiently strong to overcome the relative price effect.

1.3.5 Comparison with Stone-Geary Preferences

Given the prevalence of Stone-Geary-like preferences in quantitative models of structural transformation, we compare the fit of our model with one that replaces our nonhomothetic CES preferences with Stone-Geary preferences. To this end, we consider the aggregator

$$C_t^c = \left[\Omega_a^c \left(C_{a,t}^c + \bar{c}_a\right)^{\frac{\sigma-1}{\sigma}} + \Omega_m^c \left(C_{m,t}^c\right)^{\frac{\sigma-1}{\sigma}} + \Omega_c^c \left(C_{s,t}^c + \bar{c}_s\right)^{\frac{\sigma-1}{\sigma}}\right]^{\frac{\sigma}{\sigma-1}}, \tag{1.33}$$

[31] To see that, note that the relative trend in nominal values $\frac{\omega_{it}}{\omega_{jt}}$ would be proportional to $\left(\frac{p_{it}}{p_{jt}}\right)^{1-\sigma}$. For real values, $\frac{C_{it}}{C_{jt}}$, would be proportional to $\left(\frac{p_{it}}{p_{jt}}\right)^{-\sigma}$. As $0 < \sigma < 1$, both trends would move in opposite directions.

where C_t^c denotes aggregate consumption of country c at time t, $C_{i,t}^c$ denotes its consumption of sector i, $\bar{c}_a$ and $\bar{c}_s$ are constants that govern the nonhomotheticity of these preferences, $\sigma > 0$ and $\Omega_i^c > 0$ are preference parameters that are country specific.[32] We follow the estimation procedure described in Herrendorf *et al.* (2013).As with nonhomothetic CES preferences, we estimate three parameters common across countries $\{\sigma, \bar{c}_a, \bar{c}_s\}$ that govern the price and income elasticities, and $\{\Omega_i^c\}_{i\in\mathcal{I},c\in C}$ which are country specific parameters. As expected, we find that $\bar{c}_a < 0$ and $\bar{c}_s > 0$.

Figure 1.3 shows the fit of the Stone-Geary model for the same countries as in Figure 1.2 using the common parameters $\{\sigma, \bar{c}_a, \bar{c}_s\}$ and the country-specific preference shifters $\{\Omega_i^c\}_{i\in\mathcal{I},c\in C}$.[33] We see that the overall fit is better with nonhomothetic CES preferences. For example, the fitted model is not able to reproduce the hump-shaped pattern for manufacturing of Japan and Taiwan. This is confirmed for the full sample. In Table 1.3, we compare the sectoral R^2 measures generated with Stone-Geary preferences and nonhomothetic CES.[34] We find that in all sectors the fit improves with nonhomothetic CES. The difference in the $R^{2\prime}s$ ranges from 12 to 16 percentage points.

The intuition for the improvement of the fit with nonhomothetic CES is that, with Stone-Geary preferences, income effects are very low for rich countries. For high levels of income, the subsistence levels responsible for introducing the nonhomotheticity $\{\sigma, \bar{c}_a, \bar{c}_s\}$ are negligible. For example, the value of $\frac{p_{it}\bar{c}_i}{P_t C_t}$ for the U.S. at our estimated parameters is less than 0.05% at any point in time. Thus, the only remaining sources of variation left to explain the variation in employment shares are prices and trade shares.[35]

[32]Since these preferences are not implicitly additive, the price and income elasticities are not independent. In Appendix A.1.3 we show that the elasticity of substitution between i and j is $\sigma_{ij} = \sigma\eta_i\eta_j$, where η's denote income elasticities.

[33]Figures **??** and A.6 in the appendix show the fit for all countries in our sample.

[34]Formally, we compute $R_i^2 = 1 - \sum_{j=1}^N (y_{ij} - \hat{y}_{ij})^2 / \sum_{j=1}^N (y_{ij} - \bar{y}_i)^2$ where N denotes the total number of observations, $\bar{y}_i$ denotes the sample average of y_i and $i \in \{a, m, s\}$.

[35]An equivalent intuition provided by Dennis and Iscan (2009) is that the subsistence constants $\{\bar{c}_a, \bar{c}_s\}$ should not be stable over time to have income effects play a greater role and improve the model fit.

Table 1.3: *Sectoral R^2 measures for Stone-Geary and nonhomothetic CES*

	R^2	
	Stone-Geary	Nonhomothetic CES
Agriculture	0.84	0.98
Manufacturing	0.75	0.87
Services	0.74	0.90

Note: R^2 is computed for each sector as $R_i^2 = 1 - \sum_{j=1}^{N}(y_{ij} - \hat{y}_{ij})^2 / \sum_{j=1}^{N}(y_{ij} - \bar{y}_i)^2$ where N denotes the total number of observations, $\bar{y}_i$ denotes the sample average of y_i and $i \in \{a, m, s\}$.

1.3.6 Beyond Three Sectors

Jorgenson and Timmer (2011) have pointed out that in order to understand how structural transformation progresses in rich countries, it is important to zoom in the service sector, as it represents the majority of rich economies' consumption shares (see also Buera and Kaboski, 2012b). Our framework lends itself to this purpose, as it can accommodate an arbitrary number of sectors. In this section, we use the richness of the GDDC database to extend our estimation to 10 sectors: (1) agriculture, forestry and fishing, (2) mining and quarrying, (3) manufacturing, (4) public utilities, (5) construction, (6) wholesale and retail trade, hotels and restaurants, (7) transport, storage and communication, (8) finance, insurance, real state, (9) community, social and personal services, (10) government services.[36]

We estimate a demand system analogous to the one used in our baseline estimation, where we use manufacturing as a reference sector[37]

$$\log\left(\frac{L_{i,t}^c}{L_{m,t}^c}\right) = \zeta_{im}^c + (1-\sigma)\log\left(\frac{p_{i,t}^c}{p_{m,t}^c}\right) + (\epsilon_i - \epsilon_m)\log C_t^c + \nu_{i,t}^c, \tag{1.34}$$

with i denoting any of our sectors and c, a country index. Our panel estimates are reported

[36]The data set also contains information on dwellings that are not constructed within the period, but this information is very sparse and we abstract from them. Recall also that our aggregate consumption data does not contain government services. We include them in our regressions because the important items of health and education are included in this category (along with public administration, defense and social work) and we believe it is informative to have a sense of the magnitude of the income elasticity. Excluding government services from our regressions does not change the other estimates significantly.

[37]Note that in this case, the manufacturing sector is more narrowly defined than in the baseline estimation as it excludes mining and construction.

Table 1.4: *10-Sector Regression*

	World	OECD	Non-OECD
Price Elasticity σ	0.82	0.84	0.82
	(0.04)	(0.04)	(0.06)
Income Elasticity (relative to Manufacturing)			
Mining	-1.43	-1.14	-1.78
	(0.17)	(0.19)	(0.28)
Agriculture	-0.96	-0.82	-1.15
	(0.10)	(0.17)	(0.15)
Public Utilities	-0.02	0.12	-0.20
	(0.08)	(0.07)	(0.14)
Transp., Storage, Communications	0.10	0.32	-0.17
	(0.10)	(0.11)	(0.15)
Government Services	0.11	0.50	-0.36
	(0.21)	(0.20)	(0.26)
Construction	0.18	0.26	0.12
	(0.09)	(0.08)	(0.13)
Wholesale and Retail	0.37	0.56	0.12
	(0.13)	(0.17)	(0.20)
Community, Social and Personal Serv.	0.44	0.83	-0.03
	(0.14)	(0.26)	(0.07)
Finance, Insurance, Real State	0.94	1.19	0.64
	(0.15)	(0.25)	(0.17)
R^2	0.95	0.91	0.93

Note: All sectoral elasticities computed relative to Manufacturing. Standard errors clustered at the country level. All regressions include a sector-country fixed effect. Source: GDDC 10-Sector database (Vries *et al.*, 2014).

in Table 1.4. The overall fit is good, with an R^2 above 0.9 in all regressions (this includes the country-sector fixed effect). Column (1) shows that we find an elasticity of substitution of 0.82 which is reasonably close to the 0.75 we found in our baseline, three-sector, estimation. We find that the smallest income elasticities correspond to mining and agriculture, while the highest correspond to service sectors, such as finance, insurance, real state and government services. Columns (2) and (3) show that the ranking of sectors in terms of their income elasticity is very similar when we estimate OECD and non-OECD countries separately.

1.4 Micro and Macro Estimates for the U.S.

This section analyzes the U.S. case in more depth, for which we have more detailed data. We perform two exercises. First, we estimate our demand system using household data from the Consumption Expenditure Survey. This allows us to use an instrumental variable approach and control for sectoral preference shocks using time-sector fixed effects. Second, we estimate the parameters of the utility function using data on aggregate time series for the United States. Building on the work of Herrendorf *et al.* (2013), we specify the utility function both over final expenditure and value added and analyze the robustness of the estimates to these alternative definitions of the utility function.

1.4.1 Micro Estimation: Consumer Expenditure Survey

In this section we use household expenditure data to estimate our demand system. We use U.S. household quarterly consumption data for the period 1980-2006 from Consumption Expenditure Survey (CEX) data as constructed in Heathcote *et al.* (2010). We follow Heathcote et al. and focus on a sample of households with a present household head aged between 25 and 60. We also use the same consumption categories, except that we separate food from the rest of non-durables consumption.[38] We estimate the demand system using expenditure shares for each household on the left hand side, equation (1.13). To control for household fixed characteristics, we estimate the demand system using household fixed effects,

$$\log\left(\frac{\omega_{i,t}^h}{\omega_{nd,t}^h}\right) = (1-\sigma)\log\left(\frac{p_{i,r,t}}{p_{nd,r,t}}\right) + (\epsilon_i - \epsilon_{nd})\log C_t^h + \zeta^h + \zeta_{i,t} + \nu_{i,t}^h. \tag{1.35}$$

[38]Consumption measures are divided by the number of adult equivalents in the household. We use the categorization of Heathcote *et al.* (2010) for expenditures. The consumption categories in non-durables are: alcoholic beverages, tobacco, personal care, fuels, utilities and public services, public transportation, gasoline and motor oil, apparel, education, reading, health services and miscellaneous. Our data for services comes from entertainment expenditures. These includes among others fees for recreational lessons, TV and music related expenditures, pet-related expenditures, toys, games. Durables comprises vehicles (purchases/services derived from it and car maintenance and repair) and household equipment. Housing comprises the rents or imputed rents (if the dwelling is owned) as well as from "other dwellings" (primarily vacation homes). For each household we have a maximum of 4 observations (one per semester). The consumption data comes from the Family Characteristics and Income files except for years 1982 and 1983 for which the Detailed Expenditures files were used. See Heathcote *et al.* (2010) and Krueger and Perri (2006) for further discussion on the construction of the data set and its characteristics.

The superscript h denotes a household, and nd denotes non-durables –which we use as reference in our regressions. Prices $p_{i,r,t}$ come from the corresponding sectoral CPI-Us of the BLS. They vary across regions r for each expenditure category i and time t. Aggregate consumption expenditure C_t^h is deflated using a household specific CPI, as suggested by our theory.[39] Household fixed effects are denoted by ζ^h, while time-sector fixed-effects are denoted by $\zeta_{i,t}$. Note that this time-sector fixed effect absorbs sectoral preference shocks. The error term is $\nu_{i,t}^h$.

In an analogous manner to the cross-country panel, the identification comes from within household variation in total consumption expenditure over time. Note that since the BLS CPI-Us vary regionally, we can identify the price elasticity of consumption shares even after including time-sector fixed effects. Variation in the prices is arguably exogenous to households.[40]

Baseline Results Table 1.5 reports the results of estimating (1.35) with and without including consumption durables, and with and without sector-year fixed effects. The estimates are very similar across all four specifications, suggesting that sector-specific demand shocks are not a significant source of bias in our demand system. The point estimate of the price elasticity is around 0.64, which is within the range of estimates we obtained in the cross-country panel (Table 1.1). With respect to the income elasticities, food has the lowest elasticity, followed by housing, services and we find the highest income elasticity for durables. As the purchase of durables is lumpy and most of our consumption data for durables imputes the service flow, this latter elasticity should be interpreted with caution.[41]

[39] We use the Törnqvsit price index constructed from household expenditure shares, which the superlative price index for nonhomothetic preferences (Diewert, 1976).

[40] One additional possible concern is the fact that even after controlling for household fixed characteristics there is an unobserved and persistent shock driving both aggregate consumption expenditure and some particular consumption category, most likely durable goods and housing. Following Heathcote *et al.* (2010), we impute the flow services obtained from housing and vehicles, which should attenuate these concerns. We also report our estimates excluding durables and show that the estimated elasticities are very similar.

[41] The service flow is measured using market rent values when available or the potential rent commanded in the market. This latter object is imputed from a regression analysis that presents a number of important challenges, such as property values being reported by only a subset of households and having missing

Table 1.5: *Consumer Expenditure Survey, 1980-2006*

	(1)	(2)	(3)	(4)
σ	0.64	0.64	0.63	0.63
	(0.01)	(0.01)	(0.02)	(0.2)
Food	-0.46	-0.44	-0.49	-0.48
	(0.02)	(0.02)	(0.02)	(0.02)
Housing	-0.31	-0.31	-0.27	-0.26
	(0.02)	(0.02)	(0.02)	(0.02)
Services	0.57	0.52	0.62	0.57
	(0.03)	(0.03)	(0.03)	(0.03)
Durables			0.94	0.93
			(0.06)	(0.06)
Time FE	N	Y	N	Y
Observations	346631	346631	241470	241470

Note: Std. Errors clustered at the household level. Elasticity estimates are relative to non-durables consumption. Data from Heathcote *et al.* (2010).

IV Strategy One possible concern with the previous regression is that total consumption expenditure is an endogenous choice of the household that may be correlated with some omitted variable. To address this concern, we use the instrumental variable approach developed by Johnson *et al.* (2006). The instrument is based on the fact that the timing of the 2001 Federal income tax rebates for each household was a function of the last digit of the recipient's social security number. Thus, it was effectively random.[42] Column (1) in Table 1.6 reports the OLS estimated as in our baseline estimation for the period of interest (2001-2002). Albeit some point estimates of the income elasticities differ from the baseline model, the relative ranking of the income elasticities remains the same. We also report the OLS for the households with information on tax rebates in column (2) because the households for which we can merge the tax rebate sample may not be random. We find similar estimates. Finally, column (3) reports the IV estimates. We instrument consumption expenditure with

observations for certain years (see the appendix in Heathcote *et al.*, 2010).

[42]Because the data requirements to construct household expenditures in Heathcote *et al.* (2010) are different than in Johnson *et al.* (2006), we can merge 60% of the tax rebate data with our baseline data. This represents around 20% of our data data for years 2001-2002. We thank Bart Hobijn for suggesting this instrument to us.

a dummy for whether or not a household received the rebate at a given point in time. The estimated price elasticity remains unaltered at .64 and is precisely estimated. The income elasticities are not precisely estimated, but remain similar to the other specifications.[43]

Table 1.6: *Instrumental Variables Strategy*
Consumption Expenditure Survey, 2001-2002

	(1)	(2)	(3)
σ	0.62	0.64	0.64
	(0.03)	(0.06)	(0.06)
Food	-0.24	-0.17	-0.15
	(0.03)	(0.06)	(0.65)
Housing	-0.00	-0.09	-0.36
	(0.03)	(0.05)	(0.53)
Services	0.36	0.42	0.64
	(0.04)	(0.11)	(1.04)
			1st Stage
Tax Rebate Indicator			0.02
			(0.006)
Time FE	Y	Y	Y
IV Sample Only	N	Y	Y
IV	N	N	Y
Observations	20921	4779	4779

Note: Std. Errors clustered at the household level. Elasticity estimates are relative to non-durables consumption. All estimates contain household and time-sector fixed effects. Data from Heathcote *et al.* (2010) and Johnson *et al.* (2006).

Quartiles and Time Split Estimation A key property of the nonhomothetic CES preferences is that the income elasticity parameter ε_i is constant at different income levels. As argued in the introduction, this property is supported by evidence from prior empirical work based on mirco data (e.g., see Aguiar and Bils, 2015b). Next, we complement this evidence using CEX data. We estimate the baseline regression (1.35) by income quartiles for the period 1980-2006. Columns (1) to (4) in Table 1.7 report the elasticity estimates for the

[43]If we use the value of the rebate rather than an indicator as an instrument, we obtain similar results. However, as the magnitude of the rebate was not random, we prefer to use only the indicator.

Table 1.7: *Estimation by Quartiles and Sub-periods*
Consumer Expenditure Survey, 1980-2006

	Quartiles				Pre 93	Post 93
	(1)	(2)	(3)	(4)	(5)	(6)
σ	0.63	0.76	0.75	0.67	0.65	0.70
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Food	-0.44	-0.31	-0.42	-0.48	-0.38	-0.44
	(0.02)	(0.07)	(0.09)	(0.02)	(0.01)	(0.01)
Housing	-0.20	-0.44	-0.38	-0.23	-0.24	-0.24
	(0.02)	(0.06)	(0.07)	(0.02)	(0.01)	(0.01)
Services	0.47	0.74	0.78	0.68	0.72	0.51
	(0.03)	(0.15)	(0.18)	(0.04)	(0.02)	(0.02)

Note: Std. Errors clustered at the household level. Elasticity estimates are relative to non-durables consumption. (1) Corresponds to the first quartile, (2), to the second, etc. All estimates contain household and time-sector fixed effects. Data from Heathcote *et al.* (2010).

first to the fourth quartiles of the CEX. We find that the value of all elasticities is quite stable at different income levels. Likewise, splitting the sample between the pre-1993 and post-1993 periods and estimating the elasticities separately also yields very similar estimates.

1.4.2 Macro Estimation: Value Added and Expenditure Measures

We conclude the empirical exploration by estimating our demand system with aggregate U.S. consumption time series. We use the data from the Bureau of Economic Analysis as constructed by Herrendorf *et al.* (2013) for agriculture, manufacturing and services. Aggregate consumption data are decomposed in two different ways: final expenditure and value added. Consumption expenditure classifies sectors according to final good expenditure. Value added decomposes each dollar of final expenditure into the share of value added attributable to agriculture, manufacturing and services using U.S. input-output tables. For example, purchases of food from supermarkets is included in agriculture in the final expenditure computation, while it is broken down into food (agriculture), food processing (manufacturing) and distribution (services) when using the value added formulation. Thus, final expenditure and value added are two alternative classifications of

the same underlying data.

Herrendorf *et al.* (2013) estimate a Stone-Geary utility demand system specified over expenditure and value added time series. They find that this distinction yields quantitatively very different results. Using value added measures, the elasticity of substitution across sectors is not statistically different from 0. Buera and Kaboski (2009) report a similar finding for the period 1870-2000. When estimating the model with final expenditure they find that the elasticity of substitution is around 0.85. Herrendorf *et al.* (2013) convincingly argue that the elasticity of substitution should be greater when using expenditure measures because they embed goods from the three sectors. However, they do not provide any justification for aggregate consumption being a Leontief aggregator of sectoral outputs. In fact, Buera and Kaboski (2009) consider the Leontief estimate an "implausibly low elasticity of substitution."

We re-do the exercise of Herrendorf *et al.* (2013) using the nonhomothetic CES demand system rather than Stone-Geary, estimating

$$\log\left(\frac{\omega_{at}}{\omega_{mt}}\right) = \zeta_{am} + (1-\sigma)\log\left(\frac{p_{at}}{p_{mt}}\right) + (\epsilon_a - \epsilon_m)\log C_t + \nu_{amt}, \tag{1.36}$$

$$\log\left(\frac{\omega_{st}}{\omega_{mt}}\right) = \zeta_{sm} + (1-\sigma)\log\left(\frac{p_{st}}{p_{mt}}\right) + (\epsilon_s - \epsilon_m)\log C_t + \nu_{smt}, \tag{1.37}$$

where ω_{it} denotes consumption expenditure or value added in sector i at time t.[44] Our estimates are reported in Table 1.8. Our point estimate of the elasticity of substitution for the expenditure data is 0.88 which is very close to estimate of 0.85 reported by Herrendorf *et al.* (2013). As Herrendorf et al., we find that the elasticity of substitution is larger for expenditure data than value added. However, our estimated elasticity for value added measures is 0.57 with a standard deviation of 0.1. Thus, the preferences implied by our estimate differ significantly from the Leontief specification found using the Stone-Geary setting.

The intuition for this difference in the estimate of the price elasticity is as follows. Expenditure shares in services measured in value-added terms raise at a faster rate than the

[44] As we are using relative consumption shares rather than employment, there is no need to control for international trade in this regression because it is subsumed in consumption expenditure.

Table 1.8: *Consumption Expenditure and Value Added for the U.S.*

	σ	$\varepsilon_a - \varepsilon_m$	$\varepsilon_s - \varepsilon_m$	Obs.	R^2_{am}	R^2_{sm}
Value Added	0.57	-0.63	0.62	63	0.77	0.93
	(0.10)	(0.12)	(0.06)			
Expenditure	0.88	-0.63	0.55	63	0.98	0.91
	(0.03)	(0.03)	(0.03)			

relative price of services. The Stone-Geary demand system imposes that the income effects become less important as aggregate consumption grows. This implies that the estimation has to load the late increase in service expenditure to increases in the relative price of services. Thus, as the relative prices of services grows at a slower rate than value added, the estimation selects the minimal price elasticity to maximize the explanatory power of relative prices. In contrast, the nonhomothetic CES does not impose declining income effects. As a result, both income and price effects help account for the secular increase in expenditure shares in services and the estimation does not need to select Leontief demand system.

Our estimates of the income elasticities are very similar using measures of sectoral activity. In fact, the point estimates are identical for $\epsilon_a - \epsilon_m$, with an elasticity of -0.63. The estimates for $\epsilon_s - \epsilon_m$ are 0.62 for value added data and 0.55 for expenditure data and we cannot reject the null that they are statistically identical. Thus, our estimates imply that the role for nonhomotheticities is very similar regardless of whether utility is specified in terms of value added or expenditure.

Forecasting U.S. Expenditure Shares What do the estimated price and income elasticities imply for the evolution of the sectoral composition of the U.S. economy?[45] Assuming that relative prices and aggregate consumption grow at the average rate of the postwar period, we forecast expenditure shares in 2025 and 2050 using the estimated price and income elasticities. The projected evolution of expenditure shares is depicted in figure A.4 in the appendix. In the last year of our sample (2010) expenditure shares were 6% in agriculture, 20.5% in manufacturing and 74.2% in services. The projected shares in 2025 are 3.8% in

[45]We thank Paco Buera and Alex Monge-Naranjo for suggesting this exercise.

agriculture, 18.1% manufacturing and 78.1% in services. In 2050, our projected shares are 2.1%, 14.5% and 83.4% in agriculture, manufacturing and services respectively. This exercise suggests that the process of structural transformation in the U.S. may continue in the next decades, with manufacturing and agriculture still accounting for a non-negligible part of the economy.

1.5 Conclusion

This paper presents a tractable model of structural transformation that accommodates both long-run demand and supply drivers of structural change. Our main contribution is to introduce the nonhomothetic CES utility function to growth theory and to show their empirical relevance. These preferences generate nonhomothetic Engel curves at any level of development, which are in line with the evidence that we have from both rich and developing countries. Moreover, for this class of preferences, price and income elasticities are independent and they can be used for an arbitrary number of sectors. We argue that these are desirable theoretical and empirical properties.

In contrast to generalized Stone-Geary utility functions, nonhomothetic CES utility functions do not assume that preferences become asymptotically homothetic as income grows. In contrast to PIGL preferences, recently employed in a growth model by Boppart (2014a), nonhomothetic CES utility functions accommodate an arbitrary number of sectors with heterogeneous income elasticities. In contrast to both of these models, our demand system does not assume specific functional interrelation between price and income elasticity. This property makes our specification particularly suitable for the exercise of separating the contributions of income and prices to changes in relative sectoral demand. Relative to models with differential trends in relative prices and homothetic constant elasticity of substitution preferences, our model has the advantage that can accommodate trends in both real and nominal measures.

From a quantitative stand-point, we show that our model captures well the broad sectoral reallocation patterns of structural change, in spite of the parsimony of our approach (we

only make use of three elasticity parameters). We estimate our model applied to three sectors (agriculture, manufacturing and services) using a panel of 25 countries for the postwar period. The sample covers countries with very different levels and trajectories of development. The model fit captures the evolution of these three sectors of the economy. In particular, it generates a hump-shaped evolution for the manufacturing sector in all cases where this pattern appears in the data.

To conclude, we believe that the proposed preferences provide a tractable departure from homothetic preferences. We think that they can be used in many applied general equilibrium settings that currently use homothetic CES and monopolistic competition as their workhorse model, such as international trade (see Matsuyama, 2015, for a subsequent application of these preferences with monopolistic competition in international trade). These preferences can be also combined with production functions without constant shares to study skilled-biased technological change.

Chapter 2

Demand-Pull, Technology-Push, and the Sectoral Direction of Innovation[1]

2.1 Introduction

Modern economies go through immense structural transformations as their incomes grow: the shares of manufacturing industries in employment and output dramatically decline, while they continuously rise for services. What are the implications of these transformations for understanding the future of growth? Historically, manufacturing industries have shown the fastest rates of technological growth, have hired the most number of R&D workers, and have created the highest volumes of patents, our best available indicator of innovation. In contrast, service industries are often associated with slow productivity growth and low intensity of innovation activity. These observations have caused concerns about the sustainability of innovation and productivity growth among academics and policy makers.[2] These

[1] Co-authored with Diego Comin and Martí Mestieri.

[2] Baumol (1967b) famously coined the term "cost disease" to describe a chronic problem of the personal services industry that limits the potential for productivity growth in this sector. More recently, Gordon (2016) has forcefully argued that the main innovations driving the productivity growth of the manufacturing sector in the 20th century were unique to this era and will not be matched by the more recent innovations in the information and telecommunication industries (see also The Economist's special issue Eco, 2013). Clark (2016) discusses the close connection between the phenomenon of structural transformation and Gordon's projections about the future of growth.

concerns stem from a view of growth that emphasizes the role of technological possibilities as the key determinants of the direction of innovation. If technological possibilities determine which industries are innovation-intensive and which are not, growth will inevitably slow down when the share of economic activity in the former group falls.

As early as Schmookler (1966), however, economists have noted that, due to the non-rivalry of ideas, there are strong incentives for innovation efforts to shift to sectors with larger market demand. Accordingly, to the extent that consumer preferences act as drivers of sectoral market size, they may also act as determinants of the sectoral intensity of innovation. In particular, changes in consumer income can result in shifts in patterns of consumption across sectors. Indeed, recent work has shown that the income elasticities of goods and services produced by different industries show robust and persistent differences (Aguiar and Bils, 2015a), and that these differences explain a major part of the long-run reallocations of output and employment across sectors (Boppart, 2014b; Comin *et al.*, 2015). These observations provide us with an alternative view for the future of growth: one in which innovation will follow the patterns of structural transformation, and the pull force of demand will raise the rates of innovation and productivity growth in highly income elastic sectors such as services.

In this paper, we study how demand nonhomotheticity and technological possibilieis together shape, first, cross-sectoral variations in market size and, second, the sectoral direction of innovation. We first empirically document that the income elasticities of the outputs of US industries significantly correlate with the rates of growth of patenting and R&D expenditure, as proxies for the output and inputs of innovation. For this exercise, we use income elasticity estimates that Aguiar and Bils (2015a) have obtained using household-level Consumption Expenditure Survey (CEX) data in the US. We further develop a multi-sector model of endogenous growth with nonhomothetic demand and persistent variations in relative income elasticities of sectoral outputs. We find that for, a general class of innovation possibilities frontiers, the income elasticities of sectoral outputs ultimately determine technological growth across industries. The asymptotic equilibrium paths of our

economy give rise to the types of correlations that we empirically document in the paper.

To see the core mechanism of our theory and to better appreciate potential interactions between the pull force of demand and push force of technology, consider a static closed economy with two sectors: services (s) and manufacturing (m). Relative output of the two sectors Y_s/Y_m equals the relative demand, which in general depends both on relative prices (P_s/P_m) and income (real consumption C_{tot}). Equalization of the marginal product of productive factors imply that relative sectoral prices (negatively) depend on the relative states of technologies (N_s/N_m) in the two sector. We can show these relationships as

$$\frac{Y_s}{Y_m} = \mathfrak{D}\left(\frac{P_s}{P_m}; C_{tot}\right) \qquad \text{and} \qquad \frac{P_s}{P_m} = \mathfrak{P}\left(\frac{N_s}{N_m}\right).$$

A body of evidence shows that broad categories of goods are gross complements; therefore, the relative expenditure of households $\mathfrak{D}$ is increasing in relative prices P_s/P_m. Together, these demand-side relations suggest relative output is negatively related to relative sectoral technologies. In addition, we can capture the forces of technological possibilities through an innovation supply function

$$\frac{N_s}{N_m} = \mathfrak{T}\left(\frac{Y_s}{Y_m}\right),$$

that shows how relative technology may (positively) respond to the relative size of output and demand. Substituting for the relative prices and intersecting supply and demand, the model endogenizes equilibrium technology (N_s^*/N_m^*) as a function of income C_{tot}. Figure **??** shows how a rise in income C_{tot} affects the equilibrium bias of technology when sector s is more income elastic relative to m. So long as the market size elasticity of the supply of innovation exceeds zero, as income rises the relative state of technology improves in sector s.

Our theory shows how the core mechanism of this static model generalizes to a dynamic setting, and endogenously determines both the *rates of growth* and the levels of technology in an infinite-horizon multisector growth model.[3] We use nonhomothetic CES preferences

[3]In this sense, our model can be distinguished from theories of factor-biased technical change (Acemoglu, 2002, 2007), in which sectors reach endogenously different levels of technologies, but still grow at the same

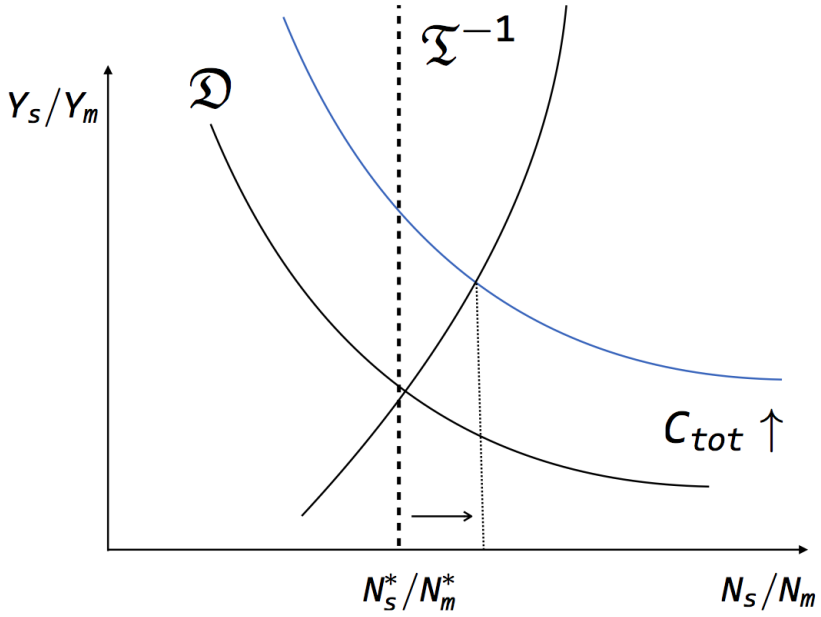
(Hanoch, 1975b; Comin *et al.*, 2015) to account for variations in the income elasticity of sectoral outputs and formulate the pull forces in demand $\mathfrak{D}$. We assume that entrepreneurs innovate to create novel sector-specific varieties, and that the masses of these varieties determine the states of sectoral technologies in a Romer-style setting. We account for the push forces in our formulation of $\mathfrak{T}$ through heterogeneous costs of innovation for different sectors and allow for intersectoral knowledge spillovers.

Along the equilibrium path of our model, two conditions fully characterize the allocation of sectoral innovation, offering simple and intuitive expressions for their dynamics. First, the entrepreneurial arbitrage condition, or simply the free entry condition, equates the value of entry across sectors. This implies that the productivity of R&D (in creating new varieties) and the value of each new variety are inversely related across sectors. Second, the R&D investment arbitrage equation equates the return to investing in the value of monopoly assets across different sectors. This implies that the sum of growth in value of monopolies and their dividend to value ratios are equalized across sectors. Together, these two conditions imply that the allocation of R&D at any point in time depends on two distinct distributions across sectors: the distribution of (production) market size, and the distribution of (knowledge) asset values. The distribution of market size captures the role of demand-side pull forces, and the distribution of asset values captures that of technology-side push forces. The allocation of R&D workers across sectors is given by a linear combination of these two distributions at any point in time.

We further characterize the asymptotic rates of technological growth for different sectors, along equilibrium paths that converge to a constant rate of growth in real income. We find that relative income elasticities asymptotically pin down both the relative shares of R&D investments and the relative rates of technological growth across sectors. Sectors that produce more income elastic goods asymptotically grow fasters in terms of innovation and productivity. We show that these asymptotic results explain the empirical regularities that we document using firm patenting and R&D expenditure, as proxies for the growth of

rates. See below for further discussions of the distinctions between the two models.

Figure 2.1: *Determination of relative technologies in a static two-sector model of endogenous growth and nonhomothetic demand. The $\mathfrak{D}$ curve shows the relative sectoral consumer demand, and the $\mathfrak{T}^{-1}$ curve the relative supply of innovation. So long as the innovation supply function is not perfectly inelastic (the dashed line), when real income C_{tot} grows, it shifts up the relative demand for the income elastic sector s, and the equilibrium level of relative technology shifts to the right.*



innovation across sectors. In Appendix B.2, we show that our asymptotic results generalize to a broad class of potential patterns of inter-industry knowledge spillovers.[4]

Our paper contributes to a literature that investigates the determinants of the cross-industry variations in innovation activity. This literature in particular has attempted to determine whether the ultimate driver of innovation across industries is the incentive pull of the size of demand or the push force of technological possibilities. A number of early and groundbreaking studies documented an industry-level relationship between market size and innovation (Griliches and Schmookler, 1963; Scherer, 1982). However, these studies did not rely on a general equilibrium framework and were therefore marred with conceptual issues. For instance, the empirical exercise favored in this literature, which consisted of

[4]We also show that the predictions of our theory generalize to alternative formulations of the process of innovation and endogenous growth. While we present the theory in the main paper based on an expanding varieties model to simplify the exposition, Appendix B.3 shows that all of our main results also generalize to a Schumpeterian model of innovation and growth.

correlating patents (or other measures of innovation) with sectoral outputs, did not account for the potential reverse causality between innovation and market size. In particular, the size of output is shaped by sectoral prices, which in turn reflect the state of sectoral technologies. Despite some further work in the literature (Pakes and Schankerman, 1984; Cohen, 2010), we still lack a theoretical framework to study the interactions between technology and demand in shaping the sectoral direction of innovation in a general equilibrium setting. More recent theoretical work on growth has largely disregarded the empirical regularities documented by this literature. In particular, even though the endogenous growth theory (Romer, 1986, 1990; Grossman and Helpman, 1990) relies on Schmookler's concept of private, profit-driven R&D, but it has abandoned the question of sectoral bias of innovation in favor of single sector models that study aggregate outcomes.

We argue that the gap in the literature on the sectoral direction of innovation in part stems from the lack of demand systems that feature persistent heterogeneity in income elasticities. In this paper, we employ nonhomothetic CES preferences, introduced first by Hanoch (1975b) and Sato (1975b) and recently used by Comin *et al.* (2015), to address the difficulty in formulating such demand systems.[5] We focus on constant growth paths where persistent variations in income elasticities, in combination with income growth, imply large variations in market sizes across sectors.[6]

Our theory shares some of the core features of the theory of factor-biased technical change (Acemoglu, 2002, 2007), but is also distinct in important ways. First, Acemoglu (2002) focuses on the bias of innovation toward different factors of production, e.g., skilled versus non-skilled labor, in a two-factor and one-good model of growth. Instead, we focus on the direction of innovation across different industries in a single-factor and multiple-good model of growth. In Acemoglu's model, the underlying heterogeneity stems from the variations in

[5]We note another recent theory that has studied sectoral composition of innovation in a two-sector model for the production of goods and services (Boppart and Weiss, 2013).

[6]Relatedly, previous work in the field of industrial organizations (see, e.g., Cohen *et al.*, 1987) has used variations in the income elasticity of demand to account for variations in R&D intensity of firms.

the stocks (endowments) of factors of production.[7] In contrast, the underlying heterogeneity in our model is driven by the (plausibly exogenous) differences in the income elasticity of goods and services produced by different sectors. There is no a priori reason to believe that factor intensities across different sectors necessarily produce biases similar to those we derive here based on income elasticities.[8] Even if such correlations exist, the mechanism identified in our paper would operate distinctly from that of a theory of factor-biased technical change.

On the empirical side, very few papers have attempted to carefully identify the response of innovation to market size. The few exceptions are Acemoglu and Linn (2004), who report in a study of pharmaceutical industry a response in terms of the number of drugs developed, but not for patents (see also Cerda, 2007; Budish *et al.*, 2015). More recently, Jaravel (2017) provides evidence that product innovations are disproportionately biased toward goods consumed by high-income households, using barcode-level scanner data from the US retail sector.

The paper is organized as follows. Section 2.2 discusses our empirical exercise. Section 2.3 presents the model, and Section 2.4 concludes the paper. Most proofs are contained in Appendix B.1.

2.2 Motivating Evidence

Let us begin by revisiting the cross-sectoral relationship between innovation and market size. As we mentioned in the introduction, an earlier literature has documented strong correlations between innovation activity and the size of output across sectors, but we cannot readily interpret such relations as evidence for the effect of latter on the former. Here,

[7] We note that, importantly, Acemoglu (2002) abstracts away from factor accumulation in his theory of factor-biased technical change, assuming exogenous and constant factor endowment across the economy. It is likely that accounting for factor accumulation may modify some of the predictions of the benchmark theory of factor-biased technical change.

[8] We note that a recent paper by Caron *et al.* (2014) in fact documents a correlation between skill intensity and income elasticity of demand for outputs across sectors, based on trade flows data.

we instead aim to document the relationship between innovation activity and plausibly exogenous measures of the income elasticity of sectoral outputs. This allows us to examine the extent to which the variations in the dynamics of innovation outputs and inputs may be driven by the dynamics of sectoral demand differences as income grows.

As for the measures of income elasticities for different industries, we rely on the micro estimates of Aguiar and Bils (2015a) based on the Consumption Expenditure Survey (CEX) for the period of 1980-2010. Aguiar and Bils (2015a) provide estimates for the income elasticity of 20 different categories of consumption expenditures. Their choices of these 20 categories are driven by the availability of consistent observations in the CEX. We first map these estimates to 69 standard Personal Consumption Expenditure (PCE) product types (BEA table 2.4.5). We then use the crosswalk between PCE categories and NAICS industry codes at the 2 and 3 levels of classifications. We associate all industry codes corresponding to the same PCE commodity type with the same elasticity.

We use two different proxies for innovation activity at the industry level: total number of patents and total R&D expenditures of private firms. Our patent data includes all patents issued by the USPTO between 1976 to 2015. We employ the following strategy to assign patents to different industries. We focus attention on the set of patents that are assigned to companies included in the compustat dataset. We then associate each patent to the industry class (NAICS or SIC) of the company based on the compustat data. This procedure leaves us with a total of 778,558 patents. For the R&D expenditure, we rely on the confidential Business R&D and Innovation Survey (BRDIS) and its predecessor Survey of Industrial Research and Development (SIRD) administered by NSF, matched to the LBD data. We aggregate the R&D expenditure of the firms that belong to the same industry code.

Our measures of income elasticity vary at the industry level, but the relation between our proxies of innovation activity and the underlying innovation may also vary across industries. To deal with this problem, we focus on the *rates of growth* of our proxies as the variables of interest. To the extent that the mapping between our proxies and the underlying innovation activity is linear and invariant over time, using the rates of growth as the outcome variable

Table 2.1: *Regressions of annual patent growth on the income elasticity parameters from Aguiar and Bils (2015a). The three columns on the left include the total number of patents, while the three columns on the right weight patents by their number of citations.*

	Raw Patents			Patent Weighted by Citations		
	(1)	(2)	(3)	(4)	(5)	(6)
Elasticity	0.024	.024***	.016***	.025	.025***	.016***
	(.020)	(.007)	(.007)	(.021)	(.006)	(.008)
Year FE	No	Yes	Yes	No	Yes	Yes
Broad Ind. FE	No	No	Yes	No	No	Yes
R^2	.0004	.91	.91	.0005	.90	.90

Obs. 3002, s.e.: robust, clustered at year and NAICS 1, respectively.

allows us to compare the differences in the dynamics of the underlying activity across sectors.[9] Accordingly, we run regressions of the form

$$InnovGrowth_{it} = \alpha + \beta\, \varepsilon_i + FE_t + FE_I + error_{it},$$

where $InnovGrowth_{it}$ is growth in R&D and patents in sector i at time t, ε_i is the proxy for the income elasticity of goods and services produced by industry i, FE_t is a time fixed effect, FE_I is a fixed effect for a broad industry code (NAICS 1-digit).

Table **??** shows the relation between the growth in the number of patents and the income elasticity across 3 digit NAICS codes. We find relatively robust and sizable correlations between the growth in patenting and our measures of income elasticity. An estimated value of 0.24 implies that the rate of patenting grows 1.5% faster in an industry at the 95 percentile of income elasticity ($\varepsilon_i = 1.48$) relative to one at the 5 percentile ($\varepsilon_i = 0.82$). As we can see, the estimates are fairly robust to the inclusion of fixed effects for time and broad industry classes (NAICS 1 digit). Note, however, that some of the variations across broad industries are also driven by the differences in the income elasticity of the outputs of these sectors, e.g., the difference between manufacturing and agriculture. Nevertheless, it is reassuring that we can find similar estimates when relying only on the variations of income elasticities within

[9] In Section 2.3 below, we discuss how these proxies relate to the predictions of the model.

Table 2.2: *Regressions of the rate of R&D expenditure growth on the income elasticity parameters from Aguiar and Bils (2015a).*

	(1)	(2)	(3)
Elasticity	0.001	.136***	.496***
	(.069)	(.06)	(.127)
Year FE	No	Yes	Yes
Broad Ind. FE	No	No	Yes
R-squared	.004	.257	.349

Number obs. is 1120. Robust standard errors in parenthesis. Weighted regression by number of obs. by industry.

broader sectors (NAICS 1 digit industries).

For further robustness, the three columns on the right of the same table present the regressions weighted by the patents' the numbers of citations, to account for the differences in the quality of the underlying innovation. We find that the estimates do not show sensitivity to the choice of weighting of patents.

Table **??** shows the relation between the growth of R&D expenditure, as a proxy for innovation inputs, and the income elasticity across industries using the BRDIS and SIRD data. In this case, we need to include the fixed effects to uncover the relationship. However, the estimates appear at least an order of magnitude larger than those obtained when we use patents as the proxy for innovation outputs.

We conclude from these results that micro-level estimates of income elasticity of industry outputs, drawn independently based on consumer surveys, are correlated with the growth in different proxies of inputs and outputs of innovation across industries. In the next section, we offer a theoretical framework to understand and interpret the empirical relationships uncovered in this section.

2.3 Model

Consider an economy endowed by one factor of production, labor, that is inelastically supplied by a mass H of households. Households consume goods produced by a fixed set

of I distinct industries, which we may interchangeably refer to as sectors. Goods in each industry are produced by perfectly competitive producers who use as their inputs a set of industry-specific intermediate goods. Innovators invent new varieties of intermediate inputs for each industry.

The innovation side of the economy has a simple Romer-style expanding varieties structure. R&D firms hire workers and direct their research and development efforts to a given industry and develop novel intermediate goods intended for production in that specific industry. Upon success in creating a new intermediate product in a given sector, the R&D firm earns perpetual monopoly rights on the invented blueprint. Equity shares in firms that produce the resulting intermediate inputs are the only investment instrument in this economy.

This simple set up allows us to focus attention on our main channel of interest, which is how the differential market size effect implied by preference nonhomotheticity determines the direction of firms' R&D efforts. In Appendix 2.2 we show that all of our results also generalize to a standard Schumpeterian structure for the innovation side of the model.

Notation We denote logarithms of variables with lower case letters, for instance, $p_i(t)$, $q_i(t)$, and $y_i(t)$ denote logarithms of $P_i(t)$, $Q_i(t)$, and $Y_i(t)$, respectively. We denote vectors and matrices with bold face notation, for instance $\boldsymbol{p}(t)$ is a vector with elements $[p_i(t)]_{i=1}^{I}$. The only exception to this rule is the real interest rate, which we denote by $r(t)$ to maintain the notation most familiar for the readers.

2.3.1 Demand and Income Elasticities

There is a mass H of identical households with preferences characterized by the standard intertemporal utility

$$\int_0^\infty e^{-\rho t} \frac{C(t)^{1-\theta} - 1}{1-\theta} dt, \tag{2.1}$$

where $C(t)$ aggregates a bundle of sectoral consumption goods $\boldsymbol{C}(t) \equiv \{C_i(t)\}_{i=1}^{I}$, according to the implicitly defined function

$$\sum_{i=1}^{I} \Xi_i^{\frac{1}{\sigma}} \left(\frac{C_i(t)}{C^{\epsilon_i}(t)} \right)^{\frac{\sigma-1}{\sigma}} = 1. \tag{2.2}$$

In Equation (2.2), $\epsilon_i > 0$ is a parameter specifying the income elasticity of demand for sector i, while $\Xi_i > 0$ is a sector specific taste parameter and σ is the elasticity of substitution.

The aggregator in Equation (2.2) belongs to the family of nonhomothetic CES preferences, introduced first by Hanoch (1975b) and Sato (1975b) and used recently by Comin *et al.* (2015) in a theory of structural change. This aggregator has the unique property that it allows for heterogeneity in income elasticities of sectoral goods for all levels of income, while maintaining the constancy of elasticity of substitution as in the standard CES preferences. Moreover, as we will see below, these preferences give rise to a log-linear demand system that closely parallels recent empirical work documenting robust variations in the income elasticity of demand across sectors (see Young, 2012b, 2013b; Aguiar and Bils, 2015a). Different features of nonhomothetic preferences have been extensively discussed in (Comin *et al.*, 2015).

Solving the expenditure minimization problem for the households, we find that sectoral demand is given by

$$\frac{C_i(t)}{C(t)} = \left(\frac{P_i(t)}{P(t)} \right)^{-\sigma} C(t)^{(1-\sigma)(\epsilon_i-1)}, \tag{2.3}$$

where $P(t)$ indicates the aggregate price index, which is a function of real consumption $C(t)$ and sectoral prices. This function is defined by

$$P(t) = \mathcal{P}(C(t), \boldsymbol{P}(t)) \equiv \left(\sum_{i=}^{I} \Xi_i \left(P_i(t) C(t)^{\epsilon_i-1} \right)^{1-\sigma} \right)^{\frac{1}{1-\sigma}}. \tag{2.4}$$

Equations (2.3) and (2.4) characterize the optimal allocation of consumption across goods from different sectors for aggregate real consumption $C(t)$ and vector of sectoral prices $\boldsymbol{P}(t)$. We normalize the household wage (labor income) to 1 in every period.

The next proposition characterizes the allocation of aggregate real consumption $C(t)$ over time for a given path of real interest rate $[r(\cdot)]_{t=0}^{\infty}$ and sectoral goods prices $[\boldsymbol{P}(\cdot)]_{t=0}^{\infty}$.

The proof of the proposition in included in Appendix B.1.

Proposition 2. *(Household Intertemporal Problem) Consider the problem of a household that chooses time paths of aggregate consumption and assets* $[C(\cdot), A(\cdot)]_{t=0}^{\infty}$ *maximizing (2.1) for given paths of sectoral prices and real interest rate* $[\boldsymbol{P}(\cdot), r(\cdot)]_{t=0}^{\infty}$*, where the per-capita stock of assets evolves according to*

$$\dot{A}(t) \leq 1 + r(t) A(t) - P(t) C(t), \tag{2.5}$$

where 1 is the wage of household, taken to be the numeraire, and $P(t)$ *is given by Equation (2.4). The household's allocation should further satisfy the No-Ponzi condition*

$$\lim_{t \to \infty} A(t) \exp\left(-\int_0^t r(t') dt'\right) \geq 0. \tag{2.6}$$

Assume an interior solution exists for a household starting with initial level of assets $A(0)$*. Then, the path of consumptions and assets should satisfy the following Euler equation*

$$\dot{c}(t) = \frac{\dot{C}(t)}{C(t)} = \frac{r(t) - \rho - \bar{\dot{p}}(t) \left[1 + (1-\sigma) Cov\left(\frac{\epsilon_i}{\bar{\epsilon}(t)}, \frac{\dot{p}_i(t)}{\bar{p}(t)}; t\right)\right]}{\theta + \bar{\epsilon}(t) \left[1 + (1-\sigma) Var\left(\frac{\epsilon_i}{\bar{\epsilon}(t)}; t\right)\right] - 1}, \tag{2.7}$$

as well as the transversality condition

$$\lim_{t \to \infty} e^{-\rho t} \frac{A(t)}{P(t)} C(t)^{-\theta} \frac{1}{\bar{\epsilon}(t)} = 0. \tag{2.8}$$

In the Euler equation above, $\overline{\dot{p}}(t)$ *and* $\overline{\epsilon}(t)$ *denote the economy-wide expenditure-weighted averages of sectoral price growth* $\dot{p}_i(t)$ *income elasticity parameters* ϵ_i *and the rate of growth of prices where the expenditure shares are given by* $\Omega_i(t) \equiv P_i(t) C_i(t) / P(t) C(t)$*. The variance and covariance terms are also similarly defined under the sectoral distribution defined by expenditure shares* $\boldsymbol{\Omega}(t)$*.*

Furthermore, assume that $\sigma \in (0, 1]$ *and* $\epsilon_i > 1 - \theta$ *for all* i*. Then, the household problem has a unique solution characterized by the Euler equation and the transversality condition above.*

Henceforth, we assume the conditions specified in the proposition to ensure the uniqueness of the solution are satisfied in our economy.

Proposition 2 and the static demand allocation Equation (2.3) jointly characterize the paths of sectoral consumption for consumers with preferences that feature long-run income

nonhomotheticity. First, we observe that in the special case where $\epsilon_i \equiv 1$ for all sectors i, the preferences reduce to the standard CES and the Euler equation (2.7) reduces to the familiar form of $\theta \cdot \dot{c} = r - \rho - \dot{p}$.[10]

In the more general case of nonhomothetic CES where ϵ_i's vary across sectors, the Equation (2.7) deviates from the standard Euler equations in two important ways. First, the term $\bar{\epsilon}(t)\left[1 + (1-\sigma)\, Var\left(\frac{\epsilon_i}{\bar{\epsilon}(t)}; t\right)\right] - 1$ in the denominator accounts for the fact that in presence of nonhomotheticity, the concavity of the instantaneous utility function and, correspondingly, the elasticity of intertemporal substitution both vary with time and real income. Second, the term $(1-\sigma)\, Cov\left(\frac{\epsilon_i}{\bar{\epsilon}(t)}, \frac{\dot{p}_i(t)}{\bar{p}(t)}; t\right)$ in the numerator shows that consumption grows faster to the extent that prices are falling relatively faster for more income-elastic goods, since households would be more inclined to substitute future consumption with current consumption.

The following corollary highlights a key implication of Proposition 2 that we will heavily use in our characterization of the general equilibrium.

Corollary 1. *Along an optimal consumption path for households, the growth rates of aggregate real consumption, aggregate consumption expenditure satisfy*

$$\overline{\epsilon_i}(t)\,\dot{c}(t) = \dot{e}(t) - \overline{\dot{p}_i}(t), \tag{2.9}$$

where $\dot{e}(t)$ denotes the growth rate of consumption expenditure $E(t) \equiv \mathcal{P}(C(t), \boldsymbol{P}(t)) \times C(t)$.

The growth in the share of a given sector i in consumption expenditure is given by:

$$\begin{aligned} \dot{\omega}_i(t) \equiv \frac{\dot{\Omega}_i(t)}{\Omega_i(t)} &= (1-\sigma)\left(\epsilon_i \dot{c}(t) + \dot{p}_i(t) - \dot{e}(t)\right), \\ &= (1-\sigma)\left[\left(\epsilon_i - \overline{\epsilon_i}(t)\right)\dot{c}(t) + \dot{p}_i(t) - \overline{\dot{p}_i}(t)\right]. \end{aligned}$$

This expression shows two distinct forces that shape the evolution of sectoral share of consumption expenditure: 1) the difference between the sector's income elasticity parameter

[10] Note that the equality $\dot{p}(t) = \sum_i \Omega_i(t)\,\dot{p}_i(t) = \overline{\dot{p}_i}(t)$, which we used to find $\theta \cdot \dot{c} = r - \rho - \dot{p}$, only holds in the special case of homothetic preferences where the changes in the aggregate price index are *only* driven by changes in prices.

and the share-weighted average of all sectors, and 2) the difference between the sector's rate of growth in prices and the share-weighted average rate among all sectors. The relation ensures that the identity $\sum_i \dot{\omega}_i(t) = 0$ is satisfied for all t.

2.3.2 Production

Labor is the only factor of production in the economy. Production in each sector involves two groups of producers. First, a continuum of monopolistically competitive firms produce sector-specific intermediate goods. Second, perfectly competitive final good producers at each sector combine sector-specific intermediate goods to produce output consumed by the households.

Let us assume that sector i at time t has a continuum of $N_i(t)$ varieties of intermediate goods. Second-stage final good producers in this sector produce the final sectoral output using the production function

$$Y_i(t) = \left(\int_0^{N_i} X_{iv}(t)^{\frac{\zeta}{\zeta+1}} di \right)^{\frac{\zeta+1}{\zeta}}, \tag{2.10}$$

where $X_{iv}(t)$ is the sector-specific intermediate input of type v. We assume $\zeta > 0$ so that the elasticity of substitution among different intermediate goods $\zeta + 1$ exceeds 1.

The producer of product v in sector i has an eternal monopoly over the market for that good. This monopolist firm produces intermediate goods at constant marginal cost Φ in units of labor, and sells it to the final producers of the sectoral good. Since the elasticity of demand faced by the monopolist for her particular variety is $1 + \zeta$, all monopolists charge the same price $P_{iv}(t) = \frac{\zeta+1}{\zeta}\Phi$. We normalize the price of intermediate goods to 1 by assuming that the value of the marginal cost satisfies $\Phi \equiv \zeta / (1 + \zeta)$.

Given that the markets for final sectoral goods are competitive, price $P_i(t)$ of final goods in sector i equals the marginal cost of production. Since all intermediate goods sell at the same unit price, sectoral good prices depend solely on the number of available intermediate good varieties $P_i(t) = N_i(t)^{-\frac{1}{\zeta}}$. Accordingly, we can write the profits $\Pi_{iv}(t)$ accrued to a

monopolist operating the variety v in sector i as

$$\begin{aligned}\Pi_{iv}(t) = \frac{1}{1+\zeta} X_{iv}(t) &= \frac{1}{1+\zeta}\left(\frac{P_i(t)}{P_{iv}(t)}\right)^{1+\zeta} Y_i(t), \\ &= \frac{1}{1+\zeta}\frac{1}{N_i(t)}\left(\frac{P_i(t)\,Y_i(t)}{Y(t)}\right) Y(t), \end{aligned} \tag{2.11}$$

where we have defined $Y(t)$ as the total value of output in the economy, that is, $Y(t) \equiv \sum_i P_i(t)\,Y_i(t)$. Note that the expression within the parentheses on the second line corresponds to the share of final production in sector i.

Equation (2.11) already contains the core mechanisms driving the direction of innovation in our model. In particular, we can identify two different components that influence the size of profits reaped by a monopolist in a given sector. First, there is the *price effect* $\left(\frac{P_i}{P_{iv}}\right)^{\frac{1}{\zeta}} = \frac{1}{N_i}$ that declines as the number of varieties rises and the technology advances in a given sector. Better technologies reduce the prices charged on a given item and reduce the potential profits for a given level of output. The second term $P_i(t)\,Y_i(t)\,/Y(t)$ is the share of industry i in the total output of the economy. With a given technology at the sectoral level, if the share of that sector in the economy rises, the profits of monopolists in that sector follows suit. This is the *market size* effect. As we will see below, in our model the income elasticities together with sectoral prices determine the market size for a given level of total output.

2.3.3 Innovation and Technology-Push

R&D firms hire workers to pursue research and develop new varieties. Since varieties are sector-specific, they have to decide the sector to which they direct their activities. If R&D firms hire Z_i workers to invent intermediate goods in sector i, new varieties arrive a flow rate

$$\dot{N}_i(t) = S_i(\boldsymbol{N}(t))\,Z_i(t), \tag{2.12}$$

where S_i is the efficiency of R&D workers in sector i. The efficiency $S_i(\cdot)$ of R&D workers depends on the state of technology in all sectors $\boldsymbol{N}(t)$ and specifies an *inter-sectoral innovation spillover function*. The idea is that innovators in sector i benefit from the available stock of knowledge in other sectors as well as that in their own sector, and therefore costs of inventing new varieties should vary as the number of varieties change over time. The function S_i characterizes the technology-push side of our model. Formally, we assume these functions satisfy the conditions stated in Assumption 1 below.

Assumption 1. *Let $\boldsymbol{S}: \mathbb{R}_+^I \to \mathbb{R}_+^I$ be a 2nd order differentiable and positive valued function such that $\partial S_i/\partial N_j \geq 0$ for all pairs i and j. We assume that each element S_i is homogenous of degree 1 in its arguments for all i. Moreover, we assume that the following limit always exists and is bounded away from zero*

$$\lim_{N_i \to \infty} \frac{S_i(\boldsymbol{N})}{N_i} > 0, \tag{2.13}$$

for all sectors i and for all vectors of technological states $\boldsymbol{N} \equiv (N_i, \boldsymbol{N}_{-i})$.

The first part of the assumption merely imposes the intuitive constraint that the rise in knowledge stock of a sector j *cannot* reduce the productivity of innovators in any other sector i. The homogeneity of degree 1 is required in order to allow for sustained growth in our economy, in which the scarce factor of labor is used in the innovation sector. The intuition behind the constraint (2.13) is also rather straightforward. From Equation (2.12), we can see that the fraction $\frac{S_i}{N_i}$ gives the technological growth generated in sector i per unit of R&D labor. The condition (2.13) then states that we can sustain growth in any given sector i by investing in R&D within that sector, regardless of the state of technology in all other sectors. This condition, in particular, rules out innovation spillover functions $S_i(\cdot)$ functions that are asymptotically Cobb-Douglas.[11]

[11] To see why, note that the asymptotic R&D employment in any given sector i has to satisfy

$$\lim_{t \to \infty} Z_i(t) = \lim_{t \to \infty} \frac{\dot{n}_i(t)}{S_i(t)/N_i(t)}. \tag{2.14}$$

If condition (2.13) is violated for a given sector, then sustaining asymptotically constant technological growth $\dot{n}_i(t)$ requires unbounded growth in R&D employment for that sector.

We generally attempt to provide characterizations of the behavior of our economy for general intersectoral innovation spillover functions $\boldsymbol{S}$ satisfying Assumption 1. However, in order to provide clear predictions from the model, we need to make more specific functional form assumptions. Below, we present one particular functional form, which we later employ for characterizing the equilibrium paths of our model.

Definition 3. Consider the following choice for the innovation spillover functions

$$S_i\left(\boldsymbol{N}\left(t\right)\right) = \frac{1}{\eta_i} N_i\left(t\right)^{\delta} S\left(\boldsymbol{N}\left(t\right)\right)^{1-\delta}, \tag{2.15}$$

where $\delta \in [0,1]$, and $S\left(\boldsymbol{N}\left(t\right)\right)$ is homogenous of degree 1 and monotonically increasing in its arguments. Note that in the special case of $\delta = 1$, we reach the case with no cross-industry innovation spillovers. Function S provides an index of the general-purpose stock of knowledge in the economy that is the main source of innovation spillovers to all other sectors. The spillover functions defined by Equations (2.15) allows for potential asymmetry across sectors in the way they *contribute* to the economy-wide source of spillovers. However, it assumes perfect symmetry across sectors in the way they *receive* spillovers from the economy-wide source of ideas S. To ensure condition (2.13), function S has to satisfy

$$\lim_{\boldsymbol{N}_{-i} \to \boldsymbol{o}} S\left(\boldsymbol{N}\right) \Big|_{N_i = 1} > 0,$$

where $\boldsymbol{N}_{-i} \in \mathbb{R}^{I-1}$ is the vector of all sectoral technologies with sector i removed. An example of such a function is given by

$$S\left(\boldsymbol{N}\right) \equiv \left(\sum_{i=1}^{I} \vartheta_i^{1-\varsigma} N_i^{\varsigma} \right)^{\frac{1}{\varsigma}}, \qquad \varsigma > 0, \tag{2.16}$$

where ς a parameter specifying the degree of substitutability between different types of sectoral knowledge in contributing to the general-purpose stock of knowledge.

2.3.4 Market Equilibrium

We first review the conditions that characterize equilibrium allocations of the economy and, along the way, identify the set of variables that together characterize an equilibrium

allocation in our economy.

Market Clearing Since we have assumed a closed economy, the output in each sector has to equal the total sectoral demand, $Y_i(t) = H C_i(t)$. Thus, the share of a given industry in total output value equals the household expenditure shares on products from that sector, i.e., $\frac{P_i Y_i}{Y} = \Omega_i$, which is in turn given by the demand equation (2.3). Substituting the share of each industry in total value in the expressions for the monopoly profits (2.11), we find the following relation for the relative profits of the monopolists in two different sectors i and j

$$\begin{aligned}\frac{\Pi_i}{\Pi_j} &= \underbrace{\frac{N_j}{N_i}}_{\text{price effect}} \times \underbrace{\left(\frac{N_j}{N_i}\right)^{\frac{1-\sigma}{\zeta}} C^{(1-\sigma)\left(\epsilon_i-\epsilon_j\right)}}_{\text{total market size effect}}, \\ &= \underbrace{\left(\frac{N_j}{N_i}\right)^{1+\frac{1-\sigma}{\zeta}}}_{\text{technology effect}} \times \underbrace{C^{(1-\sigma)\left(\epsilon_i-\epsilon_j\right)}}_{\text{income effect}}. \end{aligned} \tag{2.17}$$

The second equality above shows the decomposition of relative profits into the *total* effect of technology (directly through the price effect and indirectly through the market size effect) as well as the income effect. In particular, we see that a rise in the number of varieties of intermediates in sector i relative to sector j results in a greater than proportional fall in the relative profits of monopolists in that sector. As a result, the technology effect acts as a force bringing the technology levels in different sectors toward convergence: better technology in a given sector erodes the profits of producers in that sector, thereby reducing the incentives to invest in further innovation. The second term in Equation (2.17) distinguishes our model from previous models of directed technical change, by highlighting the force of income elasticities in shaping innovation incentives. If the demand for output of sector i is more income-elastic compared to sector j, the demand for the output of this sector grows relative to sector j as the households' aggregate consumption grows.

Clearing the goods markets further implies that the total consumption expenditure of households should equate the total value of output, that is,

$$
\begin{aligned}
H \cdot E(t) = Y(t) &= \sum_i P_i(t) Y_i(t), \\
&= \left(1 + \frac{1}{\zeta}\right) L(t), \qquad (2.18)
\end{aligned}
$$

where E stands for the expenditure of a household. In the second equality we have used the fact that the final goods markets are competitive, revenues are equal to the costs, which in turn equate the total revenues of intermediate goods monopolists, and the wage is normalized to unity. The total revenues of all monopolists are $(1+\zeta)/\zeta$ times their costs, which is equal to total wages paid, $L(t) \equiv \sum_i L_i(t)$.

Finally, labor markets clear when

$$
\sum_{i=1}^{I} (L_i(t) + Z_i(t)) = L(t) + Z(t) = H, \qquad (2.19)
$$

where we have defined total employment of R&D firms as $Z(t) = \sum_i Z_i(t)$. The economy's sole resource, total labor H, can be employed in R&D, in the form of $Z(t)$, or in production, in the form of $L(t)$. We assume that households maintain a balanced portfolio of equity shares in all intermediate goods producers at all times. Therefore, R&D employment is an effective instrument of investment. Equation (2.18) shows that the employment share of production $L(t)/H$ changes linearly with the per-capita nominal expenditure of households. Therefore, in the aggregate, we can think of the decomposition of H into Z and L to be reflective of the household's allocation of available resources between investment and consumption.

Free Entry Condition The value of owning the monopoly rights on different intermediate products is the same within a given sector i, given by the net present discounted value of all future profits

$$
V_i(t) \equiv \int_t^{\infty} e^{-\int_t^s r(t')dt'} \Pi_i\left(t + t'\right) dt'. \qquad (2.20)
$$

We can write the value $V_i(t)$ of owning the monopoly rights over an intermediate good in sector i as the solution to the Bellman equation

$$r(t) V_i(t) - \dot{V}_i(t) = \Pi_i(t). \tag{2.21}$$

Along any equilibrium path, the marginal product of labor in R&D firms in any sector i cannot exceed the unit costs of labor (wages)

$$1 \geq S_i(t) V_i(t), \tag{2.22}$$

where the expression is satisfied with equality for any sector i in which firms actively engage in R&D, that is, $Z_i(t) > 0$.

Equilibrium We define an *allocation* as a collection of the time paths of aggregate and sector consumptions of households $[C(t), \boldsymbol{C}(t)]_{t=0}^{\infty}$, employment in production and R&D in each sector $[\boldsymbol{L}(t), \boldsymbol{Z}(t)]_{t=0}^{\infty}$, masses of varieties in each sector $[\boldsymbol{N}(t)]_{t=0}^{\infty}$, and the price, quantity, and net present value of monopoly rights for each intermediate good in each sector $\left\{[P_{iv}(t), X_{iv}(t), V_{iv}(t)]_{v\in[0,N_i(t)],t=0}^{\infty}\right\}_{i=1}^{I}$.

An *equilibrium* is an allocation that corresponds to the combination of constraints imposed by household utility maximization, monopolist profit maximization, and the free entry condition everywhere along the time paths. The sectoral and aggregate consumption of households should satisfy the sectoral demand Equation (2.3) and the Euler Equations (2.7) and (2.8), where household assets satisfy $A_i(t) = \frac{1}{H}\sum_i \left(\int_0^{N_i(t)} V_{iv} dv\right)$ and aggregate and sectoral prices indices are given by Equations (2.4) and $P_i(t) = \left(\int_0^{N_i(t)} P_{iv}^{-\zeta} dv\right)^{-1/\zeta}$ for all i. Employment allocations satisfy $Z_i(t) = \dot{N}_i(t) / S_i(\boldsymbol{N}(t))$ and $L_i(t) = \frac{1}{\Phi}\int_0^{N_i(t)} X_{iv}(t)\, dv$ for all i, as well as the labor market clearing condition (2.19). Prices and quantities of intermediate goods satisfy $P_{iv}(t) = \left(1 + \frac{1}{\zeta}\right)\Phi$ and $X_{iv}(t) = \frac{1}{\zeta}\frac{P_i(t)C_i(t)}{N_i(t)}$. Finally, stocks of varieties and firm values satisfy the free entry condition (2.22).

2.3.5 Equilibrium Dynamics

In this section, we characterize the dynamics of the allocations along an equilibrium path, in terms of the vector of technological state variables $\boldsymbol{N}(t)$ and the aggregate consumption (control variable) $C(t)$. Throughout this section, we will assume an *interior* equilibrium path, along which R&D and innovation is carried on in all sectors, that is, $Z_i > 0$ for all i.

We begin by defining two sectoral distributions for the allocations of output and assets $\{\boldsymbol{\Omega}(t), \boldsymbol{\Lambda}(t)\}$. First, note that from the assets market clearing condition, the total value of assets in the economy is given by

$$H \cdot A(t) = \sum_{i=1}^{I} N_i(t) V_i(t) = \sum_{i=1}^{I} \frac{N_i(t)}{S_i(\boldsymbol{N}(t))}, \tag{2.23}$$

where in the second equality we have used the free entry condition (2.22). Now, define the share of equity values of all corporate assets held in sector i as

$$\begin{aligned} \Lambda_i(t) &\equiv \frac{N_i(t) V_i(t)}{\sum_{i'} N_{i'}(t) V_{i'}(t)}, \\ &= \frac{1}{H \cdot A(t)} \frac{N_i(t)}{S_i(\boldsymbol{N}(t))}, \end{aligned} \tag{2.24}$$

Equations (2.23) and (2.24) show how the free entry condition directly links the value of assets held in a sector to the vector of technological states across different sectors. Therefore, both the total value of assets and the sectoral distribution of assets are pinned down at time t by the vector of state of sectoral technology $\boldsymbol{N}(t)$.[12] The economic logic is the usual force of free entry: higher innovation spillovers to a sector reduces the cost of innovation in the sector, which in turn lowers the value of owning intellectual property rights on an innovation in that sector. Note that Assumption 1 ensures that this value always remains finite along any allocation path.

The second sectoral allocations are the shares of sectors in output, consumption expendi-

[12] As we show in Appendix 2.2 this relation also holds in a Schumpeterian version of the model.

ture, and production employment already defined by Equations (2.3) and (2.4) as

$$\begin{aligned}\Omega_i(t) = \frac{E_i(t)}{E(t)} &= \frac{L_i(t)}{L(t)}, \\ &= \Xi_i \left(\frac{N_i(t)^{-\frac{1}{\zeta}} C(t)^{\epsilon_i}}{E(t)} \right)^{1-\sigma}, \end{aligned} \tag{2.25}$$

where total consumption expenditure, or the total production employment from Equation (2.18), is given by

$$E(t) = \frac{\zeta+1}{\zeta} \frac{L(t)}{H} = \left(\sum_{i=1}^{I} \Xi_i \left(N_i(t)^{-\frac{1}{\zeta}} C(t)^{\epsilon_i} \right)^{1-\sigma} \right)^{\frac{1}{1-\sigma}}. \tag{2.26}$$

Both sets of allocations $\{\mathbf{\Omega}(t), \mathbf{\Lambda}(t)\}$ defined above are functions only of aggregate consumption and the vector of sectoral technologies $(C(t), \boldsymbol{N}(t))$. Moreover, we can also write the aggregate allocation of employment in the R&D sector as a function of the control and state variables $(C(t), \boldsymbol{N}(t))$, as follows

$$Z(t) = H - L(t) = H \cdot \left[1 - \frac{\zeta}{1+\zeta} \left(\sum_{i=1}^{I} \Xi_i \left(N_i(t)^{-\frac{1}{\zeta}} C(t)^{\epsilon_i} \right)^{1-\sigma} \right)^{\frac{1}{1-\sigma}} \right]. \tag{2.27}$$

Now, assuming $Z_i > 0$, let us combine the free entry condition (2.22) and the Bellman equation (2.21) to find

$$\begin{aligned} r(t) &= -\dot{s}_i(t) + \frac{1}{1+\zeta} \frac{P_i(t) Y_i(t)}{N_i(t) V_i(t)}, \\ &= -\dot{s}_i(t) + \frac{1}{\zeta A(t)} \frac{L(t)}{H} \frac{\Omega_i(t)}{\Lambda_i(t)}, \end{aligned} \tag{2.28}$$

where in the first equality, we have substituted for $\Pi_i(t)$ from Equation (2.11), and in the second equality we have used Equations (2.18) and (2.24). Equation (2.28) is an investment arbitrage condition: it states that the rate of return to investment in innovation across all sectors have to be equated with the interest rate $r(t)$. This equation is one of the two key conditions, along with the free entry condition (2.22), that together characterize the sectoral allocation of innovation in our model.

To make further progress, we need to assume a specific form for the spillover functions S_i to link the rate of growth of R&D productivity $\dot{s}_i$ to the rates of sectoral productivity growth. We assume the definition 3 and Equation (2.15) for the spillover function. With these definitions

$$A(t) = \frac{1}{H}\sum_i \eta_i \left(\frac{N_i(t)}{S(t)}\right)^{1-\delta},$$
$$\Lambda_i(t) = \frac{\eta_i N_i(t)^{1-\delta}}{\sum_{i'} \eta_{i'} N_{i'}(t)^{1-\delta}}, \tag{2.29}$$

where $S(t)$ is defined according to Equation (2.16). The following lemma characterizes sectoral R&D investment, rates of productivity growth, and the interest rate along the equilibrium path.

Lemma 2. *Anywhere along the equilibrium path where the interest rate is positive $r(t) > 0$ and there is R&D investment in all sectors, that is, $Z_i(t) > 0$ for all i, the allocations of R&D employment and the rates of sectoral growth are given by*

$$Z_i(t) = Z(t)\cdot\Lambda_i(t) + \frac{1}{\delta\zeta}L(t)\cdot[\Omega_i(t) - \Lambda_i(t)], \tag{2.30}$$
$$\dot{n}_i(t) = \frac{1}{A(t)}\left(\frac{Z(t)}{H} + \frac{1}{\delta\zeta}\frac{L(t)}{H}\cdot\left[\frac{\Omega_i(t)}{\Lambda_i(t)} - 1\right]\right). \tag{2.31}$$

and the interest rate is given by

$$r(t) = \frac{1}{A(t)}\left\{\frac{1}{\zeta\delta}\left[1 - (1-\delta)\overline{\overline{\left(\frac{\Omega_i}{\Lambda_i}\right)}}\right]\frac{L(t)}{H} - \frac{1}{A(t)}\frac{Z(t)}{H}\right\}, \tag{2.32}$$

where $\overline{\overline{\left(\frac{\Omega_i}{\Lambda_i}\right)}}$ is defined as the average of the relative output-to-assets sectoral shares, weighted by the shares of sectoral knowledge in the general-purpose economy-wide stock of knowledge

$$\overline{\overline{\left(\frac{\Omega_i}{\Lambda_i}\right)}} \equiv \frac{\sum_i \vartheta_i^{1-\varsigma} N_i^{\varsigma}(\Omega_i/\Lambda_i)}{\sum_i \vartheta_i^{1-\varsigma} N_i^{\varsigma}}.$$

Lemma 2 provides an intuitive characterization for the allocation of innovation across sectors. Equation (2.30) decomposes the allocation R&D employment (inputs) into two components: a first component that allocates total R&D employment proportionally to

the shares of sectors in total knowledge assets, and a second component that pulls R&D toward sectors whose market shares exceed their shares in assets. The pull-force of market size sharply manifests itself in the latter component, which is a direct consequence of the investment arbitrage condition (2.28). Since the returns to investment in sectoral innovation has to be equalized, and since the flow of monopoly profits depend directly on the market shares of sectors, innovation investments responds one-to-one to the changes in market size.

Equation (2.31) characterizes the implications for the rate of innovation and technical growth across sectors. To the extent that investments are allocated proportionally to the sectoral shares in knowledge assets, the implied rates of technical growth are the same across sectors (first term). However, the rates of technical growth across sectors vary to the extent that sectoral market shares deviate from their shares in knowledge assets. This implies potentially complex dynamics for the rates of innovation investments across sectors, depending on the evolution of the relative sectoral shares in output and assets. In addition, the proof of the lemma in Appendix B.1 shows that

$$\dot{n}_i(t) \propto \left[\frac{1}{1+\zeta}\frac{P_i(t)Y_i(t)}{Z_i(t)} - \delta\right]^{-1},$$

suggesting a one-to-one relationship between sectoral R&D intensity ($\frac{Z_i}{P_i Y_i}$) and the relative rates of technical growth across sectors.

We can now put the different results of this section together to characterize the aggregate dynamics of the economy. If the economy starts at a technological state $\boldsymbol{N}(0) = (N_1(0), \cdots, N_I(0))$. Assuming that the equilibrium allocation remains interior everywhere along the path, we can write the evolution of the entire economy as the following dynamical system:

$$\begin{aligned}
\frac{\dot{C}(t)}{C} &= \mathcal{F}(C(t), \boldsymbol{N}(t)), \\
\frac{\dot{N}_i(t)}{N_i(t)} &= \mathcal{G}_i(C(t), \boldsymbol{N}(t)), \qquad \text{for } 1 \leq i \leq I,
\end{aligned}$$

where $\mathcal{F}$ is given by the Euler equation (2.7) and the expression for the interest rate Equation (2.32) and $P_i(t) = N_i(t)^{-1/\zeta}$, and $\mathcal{G}_i$ is given by Equations (2.31), in which the expressions

for the aggregate employment in R&D and production $Z(t)$ and $L(t)$ are given by Equation (2.27).

2.3.6 Constant Growth Path

In this section, we focus attention to a class of equilibrium allocations that involve asymptotically constant rates of growth of consumption and sectoral technologies. Such equilibria closely parallel the balanced growth paths commonly studied in single sector growth models.

Definition. Constant Growth Path (CGP): An equilibrium path is CGP if along the allocation path aggregate consumption $C(t)$ and sectoral technologies asymptotically grow at constant rates. That is, if there exist constant nonnegative values $(g^*, \gamma_1, \cdots, \gamma_I)$ such that the following limits exist

$$\begin{aligned} \lim_{t\to\infty} \dot{c}(t) &= g^*, \\ \lim_{t\to\infty} \dot{n}_i(t) &= \gamma_i g^*, \qquad \text{for } 1 \leq i \leq I. \end{aligned}$$

Correspondingly, let us define the asymptotic levels of real per-capita consumption and states of sectoral technologies as

$$\begin{aligned} C^* &\equiv \lim_{t\to\infty} C(t)\, e^{-g^* t}, \\ N_i^* &\equiv \lim_{t\to\infty} N_i(t)\, e^{-\gamma_i g^* t}. \end{aligned}$$

We will now study how the demand side and the innovation technology side, each impose a distinct set of constraints on the allocation of employment and expenditure, and on the rates of technical growth across sectors along any CGP. We discuss these constraints on the rates of growth through two lemmas. Lemma 3 presents the constraints imposed by the demand side and characterizes the set of industries with nonnegligible asymptotic shares in production markets. Lemma 4 presents the constraints imposed by the innovation technology and characterizes the set of industries with nonnegligible asymptotic shares in asset markets. Each lemma is followed by a corollary that discusses the constraints on the asymptotic *levels* in the allocations of sectoral technical states and per-capita consumption.

Finally, Corollary 4 combines the two lemmas and shows that the two sets in fact have to coincide.

Lemma 3. *(Demand Side Constraints on CGPs) Along any CGP, the distribution of of sectoral consumption expenditure, employment, and output in our economy converges to a stationary* $\{\Omega_i^*\}_i$*, and total production employment converges to a constant, that is*

$$\lim_{t\to\infty} \Omega_i(t) = \Omega_i^*, \tag{2.33}$$

$$\lim_{t\to\infty} L(t) = L^* > 0. \tag{2.34}$$

Moreover, the asymptotic growth rates of technologies in different sectors (normalized by the growth rate of real consumption) satisfy

$$\zeta = \min_i \left\{ \frac{\gamma_i}{\epsilon_i} \right\} = \frac{\overline{\gamma}^*}{\overline{\epsilon}^*}, \tag{2.35}$$

where $\overline{\gamma}^*$ *and* $\overline{\epsilon}^*$ *denote average (normalized) rates of technical growth and income elasticity parameters of different sectors under distribution* $\{\Omega_i^*\}_i$*. Let* $\mathcal{I}^*$ *denote the set of industries that achieve the minimum in Equation (2.35). The production shares of different sectors asymptotically fall at the rate*

$$\lim_{t\to\infty} \dot{\omega}_i(t) = (1-\sigma)\left(\epsilon_i - \frac{\gamma_i}{\zeta}\right) g^* \leq 0, \tag{2.36}$$

with the inequality being strict for $i \neq \mathcal{I}^*$*.*

Proof. See Appendix B.1. □

Lemma 3 characterizes the set $\mathcal{I}^*$ of industries that asymptotically constitute a non-negligible share of consumption expenditure and production employment, for a given γ of relative rates of sectoral technical growth. The set $\mathcal{I}^*$ is comprised of sectors with the lowest ratio of technical growth to income elasticity parameter, as in Equation (2.35). Asymptotically, the shares of all other sectors in output converges to zero.[13] As with the model of Ngai and Pissarides (2007b), when $\epsilon_i = 1$ for all sectors i and the preferences are homothetic, $\mathcal{I}^*$ includes only the sectors with the slowest rate of technical growth. However,

[13]This result also holds along an equilibrium path for the economy described in the model of (Comin *et al.*, 2015), in which the rates of sectoral technical growth are exogenous.

in our setting the combination of supply and demand (income elasticity) forces together determine the asymptotic sectoral composition of the economy.

Another key implication of Lemma 3 is that, asymptotically, a constant and finite share $\frac{L^*}{H}$ of labor is employed in production and the remainder of labor is employed in the R&D sector. From Equation (2.34) and market clearing conditions (2.19) and (2.18), total consumption expenditure $E(t)$ and total R&D employment $Z(t)$ converge to constant values E^* and Z^* asymptotically. The Corollary below summarizes the constraints on the asymptotic levels implied by Lemma 3.

Corollary 2. *The asymptotic shares of consumption expenditure, employment, and output in sector i is given by*

$$\Omega_i^* = \begin{cases} \left(\frac{(C^*)^{\epsilon_i}}{E^* \left(N_i^*\right)^{1/\zeta}} \right)^{1-\sigma}, & i \in \mathcal{I}^*, \\ 0, & i \notin \mathcal{I}^*, \end{cases} \tag{2.37}$$

where the total consumption expenditure of households and the total production employment are asymptotically given by

$$E^* = \left(1 + \frac{1}{\zeta}\right) \frac{L^*}{H} = \left[\sum_{i \in \mathcal{I}^*} \Xi_i \left(\frac{(C^*)^{\epsilon_i}}{\left(N_i^*\right)^{1/\zeta}} \right)^{1-\sigma} \right]^{\frac{1}{1-\sigma}} \in \left(0, 1 + \frac{1}{\zeta}\right), \tag{2.38}$$

as functions of $(C^*, \boldsymbol{N}^*)$.

While Lemma 3 characterizes the sectoral composition of the production side of employment based on the demand-side forces, Lemma 4 below characterizes the sectoral composition of the R&D side based on the technological push forces. But before introducing the constraints imposed by the technological innovation side, let us introduce a further assumption regarding the nature of inter-industrial innovation spillovers.

Assumption 2. *Consider a time path $[\boldsymbol{N}(\cdot)]_{t=0}^{\infty}$ for the sectoral stocks of technological knowledge such that $N_i(t) \to N_i^* e^{\gamma_i g^* t}$ for all i. We assume that the technological spillover functions $\boldsymbol{S}$ has the property that along any such path, the following limits exist*

$$\Sigma_{ij}^*(g^*, \boldsymbol{\gamma}) \equiv \lim_{t \to \infty} \frac{\partial s_i}{\partial n_j}\bigg|_{\boldsymbol{n} = \boldsymbol{n}(t)}, \qquad \text{for all } i, j \in \{1, \cdots, I\}.$$

It is straightforward to check that the class of intersectoral innovation spillover functions introduced in Example 3 satisfy the condition stated in Assumption 2. This assumption implies that we can define the asymptotic rate of growth of innovation spillovers to sector i at $\gamma_i^S g^*$. Along a CGP, we have that

$$\gamma_i^S \equiv \lim_{t\to\infty} \frac{\dot{s}_i(t)}{g^*} = \sum_j \Sigma_{ij}^* \gamma_j.$$

Lemma 4. *(Innovation Side Constraints on CGPs) Along any CGP, our economy converges to a stationary distribution $\{\Lambda_i^*\}_i$ of sectoral shares of corporate assets. First, define the asymptotic rate of growth of innovation spillovers to sector i at $\gamma_i^S g^*$. Along a CGP, we have that*

$$\gamma_i^S \geq \gamma_i, \tag{2.39}$$

for all i. Let $\mathcal{I}^\dagger$ the set of sectors that satisfy the expression (2.39) with equality. The shares of sector i in R&D employment and total corporate assets fall at a rate

$$\lim_{t\to\infty} \dot{z}_i(t) = \lim_{t\to\infty} \dot{\lambda}_i(t) = \left(\gamma_i - \gamma_i^S\right) g^* \leq 0,$$

with the inequality being strict for $i \neq \mathcal{I}^\dagger$ (note that $\lambda \equiv \log \Lambda$).

Furthermore, the vector of asymptotic rates of technological growth for different sectors has the following relationship with the income elasticities of sectoral products

$$\gamma_i^S - \gamma_i = (1-\sigma)\left(\frac{\gamma_i}{\zeta} - \epsilon_i\right). \tag{2.40}$$

Proof. See Appendix B.1. □

Lemma 4 establishes conditions for a sector to asymptotically constitute a nonnegligible share of R&D employment and corporate assets in the economy. This result is the R&D parallel to Lemma 3, which provided the same analysis for production employment. The key result is fairly intuitive: the R&D resources required to sustain a constant rate of technical growth in a sector crucially depends on the rate of growth of productivity of R&D workers and, in turn, the spillovers to that sector. If spillovers grow faster than the state of technology in a given sector, asymptotical R&D investment in that sector falls over time. Because of the

free entry condition, the value of corporate assets in that sector inherit this shrinkage and the share of the sector in total value of firms also falls over time.

The lemma further connects the relative asymptotic rates of technical growth, R&D productivity growth, and income elasticity across different sectors in Equation (2.40). This result summarizes one of the key insights of the model: it shows how sectoral rates of technical growth will be determined by an interaction of the technology push forces, as captured by γ_i^S, and the demand-pull ones, as captured by demand elasticity parameters ϵ_i. Below, we will see how this condition helps characterize the asymptotic rates of productivity growth across sectors in the context of the specifications of intersectoral innovation spillover functions defined in Example 10.

In parallel to Corollary 2, which characterized the asymptotic composition of production employment and consumption in the economy, Corollary 3 below characterizes the asymptotic composition of R&D employment and corporate assets.

Corollary 3. *The asymptotic shares of different sectors in corporate assets across are given by*

$$\Lambda_i^* \equiv \lim_{t\to\infty} \Lambda_i(t) = \frac{1}{\gamma_i g^* A^*} \frac{Z_i^*}{H} = \begin{cases} 0, & i \notin \mathcal{I}^\dagger, \\ \frac{N_i^*}{S_i^* \cdot H \cdot A^*}, & i \in \mathcal{I}^\dagger, \end{cases}$$

where $S_i^* \equiv \lim_{t\to\infty} S_i(t) e^{-\gamma_i g^* t}$ *and we further have*

$$Z^* = H - L^* = g^* A^* H \left(\sum_{i\in\mathcal{I}^\dagger} \Lambda_i^* \gamma_i \right). \tag{2.41}$$

In light of the results of Lemmas 3 and 4, we can now combine the two sets of constraints imposed by the demand side and the innovation sides of the model for any CGP. Corollary 4 below states the key implication for the rates of technological growth across sectors.

Corollary 4. *Along any CGP, the set of sectors with asymptotically nonnegligible shares of production and R&D employment are identical, that is,* $\mathcal{I}^* = \mathcal{I}^\dagger$. *For all such sectors* $i \in \mathcal{I}^*$, *we have*

$$\gamma_i = \gamma_i^S = \zeta \epsilon_i.$$

Proof. Consider a sector i in $\mathcal{I}^\dagger$. From (2.39), we have that $\gamma_i^S = \gamma_i$ and then from Equation

(2.40) we find $\gamma_i = \zeta\epsilon_i$, which from (2.36) implies that $i \in \mathcal{I}^*$. The reverse argument is identical. □

Corollary 4 is a powerful result that holds for any intersectoral innovation spillover functions that are compatible with a constant growth path. It states that the asymptotic rate of technological growth in sectors that constitute significant shares of economic activity, in production and R&D, is increasing in the income elasticity of the goods produced by those sectors. In other words, *technological growth is faster in sectors with higher income elasticities*. This result rests on two main forces: the free entry condition and nonhomotheticity. To gain a better intuition about this result, let us combine the free entry condition with the R&D investment arbitrage condition to find the following condition for any technologically growing sector:

$$1 = \underbrace{S_i(t)}_{\text{Spillovers}} \times \underbrace{\frac{1}{N_i(t)}}_{\text{Price Effect}} \times \underbrace{\Omega_i(t)}_{\text{Market Size Effect}} \times \frac{E^*}{r_i^*},$$

where r_i^* is a constant denoting the sector-specific discount rate. Furthermore, we know that the market size effect implies

$$\Omega_i(t) \propto N_i(t)^{-\frac{1-\sigma}{\zeta}} \times C(t)^{(1-\sigma)\epsilon_i}.$$

The argument behind Corollary 4 is now transparent. Sectors that require nonnegligible R&D employment are those for which S_i and N_i grow at the same rate, which then implies that the market size should be asymptotically constant. In order for the market size to be asymptotically constant we need $\gamma_i = \zeta\epsilon_i$.

If we are further willing to make functional form assumptions about the intersectoral innovation spillover functions, we can further characterize which specific sectors belong to $\mathcal{I}^*$, what happens to other sectors, and what the overall rate of consumption growth in this economy is.

Proposition 4. *The asymptotic rate of technological growth for each sector i along any SGP is given*

by

$$\gamma_i = \zeta\left(\varrho\epsilon_i + (1-\varrho)\,\epsilon_{max}\right), \qquad \varrho \equiv \frac{1-\sigma}{\zeta\,(1-\delta)+1-\sigma} \in (0,1)\,. \tag{2.42}$$

Assume that there is a unique $i^* \in \mathcal{I}$, such that $\epsilon_{i^*} = \epsilon_{max} > \epsilon_i$ for all $i \neq i^* \in \mathcal{I}$. We have that $\mathcal{I}^* = \mathcal{I}^\dagger = \{i^*\}$, and the total value of assets asymptotically converges to

$$H \cdot A^* = \eta_{i^*}\vartheta_{i^*}^{(1-\delta)\left(1-\frac{1}{\zeta}\right)}.$$

If the discount rate ρ satisfies

$$\frac{1}{\zeta A^*}\left(1 - \frac{\theta - 1 + \zeta\,\epsilon_{max}}{1 + (\zeta+1)\,\epsilon_{max}}\right) < \rho < \frac{1}{\zeta A^*}, \tag{2.43}$$

the CGP exists and is unique. In this CGP, the asymptotic real interest rate r^* is given by

$$r^* \;=\; \frac{1}{\zeta A^*} - \frac{(1+\zeta)\,\epsilon_{max}}{\theta - 1 + \zeta\,\epsilon_{max}}\left(\frac{1}{\zeta A^*} - \rho\right).$$

The asymptotic rate of growth in aggregate consumption is given by

$$g^* = \frac{1/\zeta A^* - \rho}{\theta - 1 + \zeta\,\epsilon_{max}},$$

and the asymptotic share of employment in production by

$$\frac{L^*}{H} = \frac{\rho\zeta A^*\epsilon_{max} + \theta + (1+\zeta)\,\epsilon_{max} - 1}{\theta - 1 + \zeta\,\epsilon_{max}}.$$

Proof. See Appendix B.1. □

The proposition characterizes the long-run rates of technical growth across different sectors. Equation (2.42) suggests that, asymptotically, relative rates of technical growth are linear in the income elasticity parameters. This result allows us to explain the empirical patterns uncovered in Section 2.2. Let us assume that our proxy for innovation outputs, i.e., the number of patents, is a linear function of the number of new varieties, that is, $Patents_i(t) = k_{1,i} \cdot \dot{N}_i(t)$, and that the R&D expenditure is a linear function of the wages paid to R&D workers in the model, that is, $RDX_i(t) = k_{2,i} \cdot Z_i(t)$, for two sector-level constants. Asymptotically, along a CGP we have that $\dot{n}_i(t) = \dot{N}_i(t)/N_i(t) \to \gamma_i g^*$, which

implies

$$PatentGrowth_i(t) = \frac{\ddot{N}_i(t)}{\dot{N}_i(t)} \to \gamma_i g^* = \zeta g^* \left(\varrho \epsilon_i + (1-\varrho)\epsilon_{max}\right).$$

Moreover, from Equation (2.14), we find

$$\begin{aligned} RDXGrowth(t) = \frac{\dot{Z}_i(t)}{Z_i(t)} &= \frac{\dot{N}_i(t)}{N_i(t)} - \frac{\dot{S}_i(t)}{S_i(t)} + \frac{\ddot{n}_i(t)}{\dot{n}_i(t)}, \\ &\to \left(\gamma_i - \gamma_i^S\right) g^* = \zeta \varrho (1-\delta) g^* \left(\epsilon_i - \epsilon_{max}\right). \end{aligned}$$

The two results above provides theoretical explanations for the correlations between the growth rates of our innovation proxies and our measures of income elasticity.

2.4 Conclusion

In this paper, we construct the first theory that endogenously determines the direction of innovation across sectors that produce goods with robustly heterogenous income elasticities. Our theory provides a framework to study the general equilibrium interactions between demand-pull and technology-push forces in determining the equilibrium rates of innovation and productivity growth across industries. We provide simple and intuitive characterizations for the evolution of R&D intensity across sectors. In the model, the asymptotic rates of innovation and productivity growth are pinned down by the income elasticities of sectoral outputs. We show that the rates of growth of R&D expenditure and patents in the US show sizable correlations with the income elasticities of their outputs, estimated using micro-level household consumption surveys.

Chapter 3

Innovation, Knowledge Diffusion, and Selection

3.1 Introduction

Firm innovation is widely believed to be the main engine of productivity growth. Governments employ a variety of different policies such as R&D tax incentives and research grants to promote and support innovation among private businesses.[1] These programs, in part, aim to internalize the externalities that our canonical theories of growth attribute to the process of innovation. When each firm innovates, it creates positive *knowledge spillovers* and negative *competition effects* (business stealing) on other firms. Therefore, to design effective innovation policy, we need to understand the sizes of these firm-level externalities and their precise role in driving aggregate productivity growth.

One key step for understanding the link between firm-level innovation and aggregate productivity growth is to account for firm heterogeneity (Syverson, 2011). When firms vary in productivity, the reallocation of market shares from less to more productive firms may

[1]For broad examples of these programs across developed and developing countries, see OECD (2012) and World Bank (2010). For more information on the design and scope of R&D tax incentives, see OECD (2010) and OECD (2011).

also contribute to aggregate productivity growth.[2] Studies have found that, among these reallocations, the margin of firm entry and exit constitutes a particularly large component of productivity growth across different industries (e.g., Foster *et al.*, 2001, 2006; Bartelsman and Haltiwanger, 2009). Inspired by these findings, several authors have recently begun to revisit the design of innovation policy in the presence of such reallocations (Acemoglu *et al.*, 2013; Lentz and Mortensen, 2014; Atkeson and Burstein, 2015). Following the workhorse model of innovation proposed by Klette and Kortum (2004), they typically formalize these reallocations through the process of creative destruction: entrants build on the knowledge of an incumbent, steal its business, and push it out of the industry. However, this framework intrinsically ties together the two margins of innovation externalities, knowledge spillovers and competition; stealing one's idea means stealing one's business.

In this paper, I take an alternative approach and construct a theory of firm innovation and heterogeneity that disentangles the two externalities of knowledge spillovers and competition. Here, entrants can imitate the ideas of productive firms to steal business from unproductive ones. In other words, knowledge spillovers stem from the *diffusion* of productive ideas to new entrants, and competition results in the *selection* of unproductive ideas out of the industry. I show that this simple modification unleashes a host of novel, economically meaningful, and sizable externalities that matter for the design of innovation policy. I use the model to derive policies that internalize these innovation externalities, qualitatively and quantitatively examine their properties, and compare them across two broad sectors of the US economy, manufacturing and retail trade.

The novel externalities arise from the interplay between diffusion and selection. Selection implies that the ideas of firms gradually become obsolete as better ideas emerge elsewhere in the industry. Of course, more productive ideas take longer before becoming obsolete. They therefore generate greater knowledge spillovers as, over the course of their longer lifetimes, more entrants have the opportunity to adopt them. Therefore, knowledge spillovers vary

[2]In this paper, I use the term productivity to refer to a broad measure of firm performance that determines its profits and market share. Foster *et al.* (2008) document that the differences in firm performance may stem from variations in both quantity and quality efficiency across firms.

with the productivity of the ideas of firms. The dependence of knowledge spillovers on productivity constitutes the primary source of heterogeneity in externalities. In addition, when a firm innovates, the competition effect reduces the expected lifetimes of other firms and adversely affects their incentives for innovation. With each investment, a firm then reallocates innovation investments across other firms and toward itself. Since the spillover externalities are heterogeneous, this reallocation produces another external effect on aggregate growth.

An example may help clarify the intuition behind these new forces. Walmart, the largest and most productive firm in the US retail sector, has been responsible for many transformative technological and managerial innovations in this sector.[3] These innovations have fueled the expansion of Walmart, and it has replaced many less productive local retailers over time. But, in doing so, it has not necessarily built on the knowledge that goes into the latter's production processes. In addition, new retail chains may also adopt the new ideas generated by Walmart, enter the market, and replace other small local retailers. To the extent that Walmart innovations furnish such opportunities for new firms, Walmart innovations create positive externalities.[4] To the extent that they crowd out the innovations of smaller firms that might have benefited other entrants, Walmart innovations create negative externalities. In designing policies that subsidize business investments in innovation, the government has to account for both these separate forces and their potential variations across firms.

The structure of the model builds on the recent theories of knowledge diffusion and selection (Luttmer, 2007, 2012; Sampson, 2016) and extends them to include firm innovation (Ericson and Pakes, 1995; Atkeson and Burstein, 2010). Entrants receive knowledge spillovers from incumbent firms. Heterogeneous incumbents invest in innovation to raise their

[3]For an overview of these innovations, see (Basker, 2007).

[4]Case studies suggest that these forces are more than mere conceptual possibilities. For instance, in a study of the impact of Walmart on the US productivity growth in the late 90s, McKinsey Global Institute (2001) finds that the firm's innovations "directly and indirectly caused the bulk of the productivity acceleration through ongoing managerial innovation that increased *competition intensity* and drove the *diffusion* of best practice (both managerial and technological) [emphasis added]."

productivity and may endogenously exit due to fixed costs. Entry, exit, innovation, and the distribution of productivity in the industry are all endogenous. Three sets of structural features characterize these endogenous outcomes: costs of entry and innovation, volatility of firm productivity, and the strength of knowledge spillovers. I specify the strength of knowledge spillovers through two different characteristics of the process of entry and adoption: first, the imperfection of adoption, in the form of decreasing returns to adopting high productivity ideas, and second, directedness of adoption, in the sense of the probability of adopting the frontier idea in the industry.

The equilibrium of the model parsimoniously explains many stylized facts about firm heterogeneity and dispersion.[5] Through the selection margin, the model readily predicts the robustly documented negative correlation between firm productivity and likelihood of exit (e.g., Foster *et al.*, 2001). Moreover, volatility and selection together allow the model to endogenously generate a Pareto-tailed distribution of firm productivity and size, as is widely reported in empirical work on firm size distribution and industry concentration (Axtell, 2001; Luttmer, 2010). This feature is particularly important since key endogenous variables such as innovation investments and their external values all vary with firm productivity. Therefore, in order to explain the behavior of the industry both at the micro and at the aggregate levels, the model's prediction should fit the tail properties of the distribution of productivity.

One important analytical result is a decomposition of the rate of productivity growth into the direct contribution of incumbent innovation investments and the contribution of entry and exit (selection). The former is given by the market-share-weighted average of the investments of different incumbents. The latter is given by the product of the rate of firm turnover and the gap in the productivity of average entering and exiting firms. Due to spillovers to entrants and competition effects on exiting firms, incumbent innovations indirectly affect the contribution of selection to growth. In particular, the gap in the

[5] In Appendix C.1, I provide an overview of the main facts about the firm dispersion and dynamics that the current model accounts for.

productivity of industry entrants and exiting incumbents becomes endogenous in this model. This constitutes a new margin for the impact of firm innovation on aggregate growth.

I provide a partial equilibrium decomposition of innovation externalities into the positive knowledge spillover effects and the negative competition effects. Typically, both knowledge spillover externalities and competition effects grow in the productivity of the innovator. When they innovate, more productive firms generate better ideas that last longer in the industry and may be more likely to be adopted by entrants. However, they also take a heavier toll on the innovation incentives of other firms. If the adoption of ideas is imperfect (involves decreasing returns), the gaps between social and private returns to innovation may become a nonmonotonic function of productivity. In this case, for highly productive firms the knowledge spillover effects eventually become dominated by the competition effect, since the former grows more slowly than the market share of the innovator while the latter remains proportional to it.

Since externalities are heterogeneous under the market equilibrium, some firms overinvest and some underinvest in innovation. The social planner should therefore reallocate innovation investments across firms. I characterize the socially optimal allocation of innovation and production, and provide a method to compute it. Moreover, I derive the policy that implements this optimal allocation using taxes and subsidies to production and innovation. Under the assumption that we can observe the productivity ranking of firms, the derived policy uses a combination of a subsidy to startup costs for entrants, a lump-sum tax on the operation of firms, and a nonlinear subsidy to innovation investments that depends on the ranking of their productivity in the industry. The nonlinear innovation subsidy is the main instrument for the reallocation of innovation investments.

I examine the quantitative importance of these new forces by calibrating the model based on data from two broad sectors of the US economy, manufacturing and retail trade. I rely on the Business Dynamics Statistics (BDS) data that tabulates moments of the life cycle dynamics of firm employment for each sector. I calibrate a total of 7 parameters, two of

which are directly calibrated to the rates of productivity and employment growth in the sector. For the remaining five parameters, I use a simulated moment matching strategy. The identification of costs of innovation and volatility comes from the rates of employment growth and reallocation among mature firms. To identify the parameters of the process of entry, i.e., the imperfection and the directedness of adoption, I rely on the core logic of the model linking the quality of firm ideas to their expected lifetimes in the industry. According to this logic, any distribution on the initial productivity of a cohort of entrants translates into a distribution on the timing of their exit from the industry, conditional on the other model parameters. Therefore, I can use the empirically observed relationship between firm age and the hazard of exit to calibrate the parameters of the process of adoption. The results suggest that, despite its parsimony, the model provides a reasonable fit for the key moments in the data.

I further investigate the productivity schedule of innovation subsidies that implement the optimal allocations. I find that for all but the very top firms we can approximate the policy by a subsidy rate that is an affine function of the productivity ranking of the firm. Incumbent firms receive a fixed subsidy rate for their innovation investments and a varying subsidy rate that is linear in their position relative to other firms in the distribution of productivity (with the rank normalized to be between 0 and 1). The dependence of the subsidy on the productivity ranking is weaker in retail compared to manufacturing (the approximate slope is about twice as large in manufacturing). Moreover, as productivity rises the pattern eventually reverses and the subsidy rate decreases among the most productive firms. This pattern is more pronounced in retail trade, due to the stronger decreasing returns to adoption implied by the calibrated parameters in this sector. For large firms in retail, their negative competition effect grows faster than their positive knowledge spillovers.

Together, these results suggest that the details of firm heterogeneity make a first-order contribution to the optimal design of innovation policy in both sectors. I further compare the government's spending on innovation subsidies under two different policies: a productivity-dependent policy that implements the optimal allocation, and the best among all one-size-

fits-all, uniform subsidies that achieve the same rate of aggregate growth. I find that the productivity-dependent policy results in sizable savings in the spending of the government in both retail and manufacturing. Therefore, under this calibration, ignoring the details of firm heterogeneity appears to be fairly costly for the government.

In comparing retail and manufacturing, I have been inspired by the salient differences in the patterns of productivity growth that prior work has documented between the two sectors. For instance, case studies in the US retail sector often attribute the transformative innovations that have fueled the fast growth of productivity in this sector to mature and productive incumbents, particularly big-box stores such as Walmart and Target (e.g., Foster *et al.*, 2016; Basker, 2016). In this sector, firm productivity is highly correlated with measures of innovation such as IT investments (Doms *et al.*, 2004). This pattern stands in contrast to manufacturing, where we typically associate transformative innovations with newcomers.[6] In addition, firm productivity is highly correlated with measures of innovation such as IT intensity in retail,[7] and reallocations play a relatively larger contribution to aggregate productivity growth in retail compared to manufacturing (Doms *et al.*, 2004; Foster *et al.*, 2006). In the model, retail's distinctive growth experience is compatible with an environment that features high innovation costs, high business volatility, and weak knowledge spillovers. These structural differences in turn reflect in the design of innovation policy, as we saw above.

To my knowledge, this paper is the first to study potential inter-industry variations in the mechanics of productivity growth through the lens of the theories of firm dynamics. The empirical evidence suggests that the prominent role of entry and exit in retail may be representative of a broader pattern across service-based industries (Foster *et al.*, 2001).

[6]Following Schumpeter's view of creative destruction, in our theories young firms have a comparative advantage for generating transformative innovations. New firms face larger incentives for innovation since they get to steal business from incumbents, whose incentives are partly curbed due to Arrow's replacement effect. New firms also face lower costs for performing disruptive innovation simply because they are more flexible and open to change (Acemoglu *et al.*, 2014). In contrast, in the current model productive firms have an advantage for process innovation because of their longer expected lifetime.

[7]See Appendix C.1 for a brief discussion of the evidence on the relationship between firm innovation and productivity.

The growing role of service-based industries in employment both in the US and across the world requires us to understand the potential policy implications of these differences. By enriching the account of volatility and spillovers, this paper takes the first steps toward understanding the structural features that determine both the design of innovation policy and the response of growth to other interventions.

Prior Literature This paper unites theories of firm heterogeneity and dynamics with the literature on firm innovation and growth. The former class of theories has highlighted the role of firm-level dispersion and volatility as important determinants of aggregate productivity (Jovanovic, 1982; Hopenhayn, 1992; Ericson and Pakes, 1995; Restuccia and Rogerson, 2008). This framework has found many applications in the empirical studies that link micro-level firm dynamics with macro-level outcomes.[8] However, it has primarily focused on understanding the *level* of aggregate productivity rather than its *rate of growth.* Instead, I follow the literature on endogenous growth theory in taking a stance on knowledge spillovers as drivers of long-run productivity growth (Romer, 1990; Aghion and Howitt, 1992; Grossman and Helpman, 1991b). A number of other authors have similarly studied growth within this framework, including Luttmer (2007, 2012), Lucas and Moll (2014), Perla and Tonetti (2014); Benhabib *et al.* (2014); Perla *et al.* (2015), and Sampson (2016). However, the current paper is the first to simultaneously endogenize entry, exit, and the innovation of firms, and to therefore provide a framework for policy analysis.

As mentioned earlier, the core results of the paper contribute to a burgeoning literature that studies the normative implications of firm heterogeneity for the design of innovation policy. In particular, it complements the results of the recent paper by Atkeson and Burstein (2015), who emphasize the key role of the "social rate of depreciation" of innovation investments in the implications of any theory of growth for innovation policy. This rate

[8]This formulation of firm innovation behavior goes back to the seminal work of Griliches (1979) and has been widely used in theoretical and empirical work in the fields of Industrial Organization and International Trade. Using this formulation, Olley and Pakes (1996) built an empirical model of industry evolution that provided a rigorous framework for the estimation of the dynamics of productivity across firms. A large literature has used this framework to link micro-level firm outcomes to the aggregate growth of productivity (e.g., Xu, 2008; Aw *et al.*, 2011; Doraszelski and Jaumandreu, 2013; De Loecker, 2013).

captures our assumptions about the expected lifetime of ideas in the industry and, for instance, is zero in the workhorse model of innovation by Klette and Kortum (2004). Atkeson and Burstein show that the assumption our models make about the rate of social depreciation shapes the response of industry growth to the changes in the "aggregate" level of innovation investments. The current paper shows that the rate of social depreciation also shapes the response of growth to changes in the *distribution* of innovation investments across firms, if depreciation is due to the endogenous obsolescence of low-productivity ideas.

The competition effect identified in the current paper is similar in spirit to the heterogeneity in externalities found in the model of Lentz and Mortensen (2014). They examine innovation policy in an extension of the Klette and Kortum (2004) theory with heterogeneous innovative types for firms. In their theory, firms are able to innovate in as many product lines as they produce in. When low-innovative type firms innovate, they might randomly destroy the innovation option of a high-innovative type. This generates a negative externality for the former's innovation. However, this effect only applies to uninnovative (and typically unproductive) firms in their paper, whereas the competition effect in the current paper can also grow with the productivity and size of the innovator.

Finally, the current paper focuses on process innovation and takes a firm's productivity as the measure of its stock of knowledge. This contrasts with the workhorse innovation model of Klette and Kortum (2004), in which all innovations are product innovations and the firm's number of products is the measure of its stock of knowledge. The distinction is perhaps best manifested in the results of the closely related paper by Acemoglu *et al.* (2013). They include endogenous exit and selection as an independent channel in an extension of the Klette and Kortum (2004) theory with heterogeneous innovative types. As with the current paper, they also find that a lump-sum tax on the operation of firms combined with subsidies to innovation and entry can substantially raise growth. Quantitatively, the two papers find roughly similar extents of overall inefficiency in growth. However, the main interactions between selection and innovation that are the focus of this paper do not arise in their paper. Following the workhorse model, they also assume that all innovations are

product innovations.[9] Therefore, the productivity of the innovator is independent of the expected lifetime of the ideas. In contrast, both types of externalities identified in the current paper stem from the link between firm productivity and the expected lifetime of the ideas it generates. As a result, the design and details of optimal innovation policy in the two papers are different.[10]

The remainder of the paper is organized as follows. In Section 3.2, I present the details of the model and characterize its long-run market equilibrium. In Section 3.3, I characterize the long-run optimal allocation of innovation and production in the model and present the results on the decomposition of externalities. In Section 3.4, I present the details and the results of the calibration exercise.

3.2 Model

3.2.1 Households and Demand

A mass $N_t = N_0 e^{g_N t}$ of households populate the economy and live in an infinite horizon. They choose the time paths of their consumption to maximize dynastic utility defined according to $U_o = \int_0^\infty e^{-(r-g_N)t} \log q_t \, dt$, where r and g_N stand for the discount rate of households and the rate of population growth, respectively. Per-capita consumption q_t of the household at time t is a CES composite of consumption over a set Ω_t of products or services

$$q_t^{\frac{\rho}{1+\rho}} = \int_{\omega \in \Omega_t} q_t(\omega)^{\frac{\rho}{1+\rho}} \, d\omega,$$

where $\rho + 1$ is the elasticity of substitution. Normalizing household per-capita expenditure to unity at all times, we find the standard demand system $q_t(\omega) = P_t^\rho \, p_t(\omega)^{-(1+\rho)}$, where

[9] This undirected account of innovation in the workhorse model also implies a thick-tailed distribution for the growth rates of firm employment. In a recent empirical investigation, Garcia-Macia *et al.* (2015) find that models with directed innovation, such that firms generate new know-hows that are similar to their existing ones, may better match the observed dynamics of firm employment in the US data.

[10] For completeness, I note that the growth model of Peters (2011) also generates endogenous market equilibrium misallocations across firms, as does the current paper. However, the misallocations in that paper stem from variations in markups. Here, I assume that markups are constant across firms.

the price index P_t is given by

$$P_t^{-\rho} = \int_{\omega \in \Omega_t} p_t(\omega)^{-\rho} \, d\omega. \tag{3.1}$$

Households inelastically supply one unit flow of labor per person. Labor is homogeneous and earns wage rate w_t at time t. Households can also hold equity stakes in a balanced portfolio of all entrepreneurial activity and profit-making firms. Let a_t denote the per-capita assets and r_t be the interest rate at time t. The household problem involves the maximization of the dynastic utility subject to the flow budget constraint $\dot{a}_t = w_t - 1 + \mathsf{r}_t a_t - g_N a_t$ and a No-Ponzi condition. The Euler equation $\mathsf{r}_t = r$, household's initial asset endowment a_0, and the transversality condition $\lim_{t\to\infty} a_t e^{-(r-g_N)t} = 0$ together determine the time paths of household consumption.

3.2.2 Firms and Innovation

A continuum mass of producers of final goods or services form the industry. Each firm produces a single differentiated product ω and monopolistically competes against other firms active in the industry. Labor is the only factor of production and firms have to pay a flow cost of ψ_f (in units of labor) to continue operation. Firms are heterogeneous in terms of performance. I assume that all of this heterogeneity is captured by a single measure θ, which specifies the labor productivity of firms. Accordingly, the output of a firm θ is given by $q = \theta l$ where l is the number of its production workers.

Innovation Production Function Firm labor productivity θ_t is also a measure of the quality of its ideas at time t. I assume that the productivity stochastically evolves according to a stochastic process

$$d\theta_t = \Gamma_t \, dt + \sigma \, \theta_t \, d\mathcal{W}_t, \tag{3.2}$$

where $\mathcal{W}_t$ denotes a standard Wiener process and σ is the volatility parameter of the idiosyncratic productivity shocks, which arrive at the firm level. The drift of the process Γ_t is the outcome of the firm's instantaneous innovation investments, which raise its expected

productivity in the future (Ericson and Pakes, 1995; Atkeson and Burstein, 2010). The innovation production function of the firm

$$\Gamma_t = G\left(i_t, \theta_t; l_t\right),$$

takes as inputs its innovation investments i_t (in units of labor) and its current ideas θ_t. In addition, it also depends on the firm's number of production workers l_t. Correspondingly, we can define an innovation cost function[11]

$$i_t = \Phi\left(\Gamma_t, \theta_t; l_t\right) = l_t \, \varphi\left(\frac{\Gamma_t}{\theta_t}\right), \qquad \text{for } \; l_t > 0, \tag{3.3}$$

where $\varphi\left(\cdot\right)$ is a convex and monotonically increasing function, and satisfies $\varphi\left(0\right) = 0$ as well as a few other properties to be specified below. I assume that a firm that does not employ any workers cannot grow its productivity, i.e., $G\left(i_t, \theta_t; 0\right) = 0$.

Equation (3.3) assumes that innovation costs depend on the number of production workers. The idea is that in order to improve production processes in the firm, production workers at some level have to be involved in the implementation, experimentation, or training stages. Therefore, a firm that hires a larger number of employees has to pay a higher cost to modify its existing processes or adopt new technologies or managerial practices. Equation (3.3) also assumes that the current stock of firm ideas θ_t proportionally lowers the costs of improving productivity, which is a common assumption in theories of firm growth (e.g., Klette and Kortum, 2004). Together these two assumptions deliver Gibrat's law, which states that firm growth is independent of firm size and sales for the largest firms.

Production Decisions The simplest way to characterize the production decisions of firms is to consider a firm that has decided on a rate of productivity growth specified by $\gamma_t \equiv \Gamma_t / \theta_t$.

[11] Atkeson and Burstein (2010) assume a slightly different specification that also delivers Gibrat's law. However, in contrast to their specification, Equation (3.3) does not feature market size effects, since innovation costs scale with both firm employment and sales. Another important difference is that, unlike in their model, the formulation of innovation costs here does not depend on the elasticity of substitution parameter ρ.

Henceforth, I will refer to γ as the firm's innovation investment to simplify the exposition. Conditional on the innovation decision, we can specify the firm's employment, output, and pricing decisions. The firm maximizes its flow profits

$$\pi_t(\theta, \gamma) \equiv \max_q p_t(q) \cdot q - w_t \frac{q}{\theta}(1 + \varphi(\gamma)) - w_t \psi_f, \tag{3.4}$$

where I have substituted for firm production input $l \equiv q/\theta$. As we already saw, $p_t(q) \equiv P_t^{\frac{\rho}{1+\rho}} q^{-\frac{1}{1+\rho}}$. In choosing its scale of production, the firm takes the costs of innovation into account as part of its labor costs. Accordingly, the firm prices its products according to $p_t = \frac{1+\rho}{\rho} w_t (1 + \varphi(\gamma)) / \theta_t$.

Next, we can derive the relation between firm productivity and its market share. With the assumptions laid out above, it follows that the sales of a firm with productivity θ_t at time t are proportional to $s_t(\theta, \gamma) = N_t ((1+\rho) P_t / \rho w_t)^\rho \theta^\rho J(\gamma)$ where I have defined the function[12]

$$J(\gamma) \equiv (1 + \varphi(\gamma))^{-\rho}. \tag{3.6}$$

Function $J(\cdot)$ characterizes the effective costs of innovation for firms: investment in innovation raises effective production costs and therefore lowers instantaneous sales.[13]

Let us now define the aggregate-level measures of the industry. Let M_t denote the industry measure defined in the space of productivity θ, and assume that along an equilibrium path a firm with productivity θ chooses innovation investments $\gamma_t(\theta)$ at time t. We can use the expression for the industry price index (3.1) to find the sales, profits, and employment of firms. Define the total mass of firms $\overline{M}_t \equiv \int_{\Omega_t} d\omega = \int dM_t(\theta)$, and the aggregate (average)

[12] I assume that there exists some $\overline{\gamma}$ such that $\lim_{\gamma \to \overline{\gamma}} \varphi(\gamma) = \infty$, and that the function j defined here is strictly concave, which requires that, for all $\gamma \in (0, \overline{\gamma})$

$$\varphi''(\gamma) > (1+\rho) \frac{(\varphi'(\gamma))^2}{1 + \varphi(\gamma)}. \tag{3.5}$$

[13] Equation (3.6) implies that stronger (static) product market competition, as proxied by the substitutability of firm products ρ, raises innovation costs. But note that it also dynamically raises the future value of competition, as reflected in the marginal value of firm productivity.

productivity $\overline{\theta}_t$ through

$$\overline{\theta}_t^{\rho} \equiv \frac{1}{\overline{M}_t} \int \theta^{\rho} J\left(\gamma\left(\theta\right)\right) \, dM_t\left(\theta\right). \tag{3.7}$$

Then, we can write firm sales as

$$s_t\left(\theta, \gamma\right) = \frac{N_t}{\overline{M}_t} \left(\frac{\theta}{\overline{\theta}_t}\right)^{\rho} J\left(\gamma\right),$$

and firm profits and wage bills are given by

$$\pi_t\left(\theta, \gamma\right) = \frac{1}{\rho + 1} s_t\left(\theta, \gamma\right) - \psi_f w_t, \tag{3.8}$$

$$w_t l_t\left(\theta, \gamma\right)\left(1 + \varphi\left(\gamma\right)\right) = \frac{\rho}{\rho + 1} s_t\left(\theta, \gamma\right). \tag{3.9}$$

The total firm employment is the sum of firm production, innovation, and overhead workers given by $l\left(1 + \varphi\right) + \psi_f$. Firm innovation intensity, defined commonly as the innovation spending to sales ratio, is given by $\varphi\rho / \left(1 + \rho\right)$.

The formulation provided here considers firm innovation as *productivity-enhancing ideas embodied in the firm* (similar to the approach taken by Ericson and Pakes, 1995; Atkeson and Burstein, 2010). Conceptually, it shares core features with the organizational capital approach, which views firms as "storehouses of information" (Prescott and Visscher, 1980). Novel ideas improve a firm's knowledge of demand and technological environment, information on employee characteristics, managerial practices, organizational structure and culture, or any other competency that enables it to operate at a higher level of productivity.[14] This formulation contrasts with an alternative approach that considers innovation as embodied in the blueprints of new products developed by the firm (e.g., Klette and Kortum, 2004). When new ideas are embodied in the firm, as in the current model, the expected lifetime of the firm will matter for its incentives to invest in innovation. As I introduce the process of

[14] Syverson (2011) provides a broad classification of the factors that may influence firm productivity, including, firm-specific know-hows, management practices, and organizational structure. See Prescott and Visscher (1980) for examples of the first group, and Bloom and Reenen (2010) for a survey of the literature that establishes managerial practices as a key to the productivity and documents its heterogeneity across firms. A large body of work in the field of organizational economics has studied the impact of de jure processes and hierarchies, de facto norms, and firm leadership on the efficiency of a firm's operations (for recent reviews of the area, see the relevant chapters in the most recent handbook, e.g., Hermalin, 2012).

entry of firms in the next section, I will discuss another conceptual difference between the two approaches regarding the imperfection of transfer of knowledge across firms.

3.2.3 Entry

Entrepreneurs pay a flow ψ_e of costs (in units of labor) per unit flow of new firms. The initial productivity of a new firm is determined by the process of knowledge diffusion: an entrant builds on the knowledge available in the current pool of ideas in the industry (Luttmer, 2007, 2012; Sampson, 2016). I assume a general specification for this process that allows variations along two dimensions: (1) which incumbents an entrant is likely to adopt from, and (2) how much of the incumbent productivity can be transferred through adoption. Accordingly, once a new firm is born it receives the opportunity to adopt an idea from the pool of ideas currently available in the industry. The adoption distribution specifies the likelihood that this adopted idea may come from each of the firms in different tiers of the distribution of productivity. Once a firm adopts the idea of a given incumbent firm, the transfer of knowledge is characterized by the imperfection of adoption, as described below.

Adoption Distribution I assume that the likelihood of adopting the idea from any given incumbent firm depends on its current tier of productivity. We can think of $M_t/\overline{M}_t$ as a rank function mapping each tier of productivity to its rank in the current distribution of productivity. In particular, the exiting firm always has the zeroth rank $M_t(\theta_{o,t}) = 0$ and the rank converges to 1 as the productivity goes to infinity. I assume that, for entrants born at time t, the probability distribution of the adopted productivity θ_a is given by

$$\mathbb{P}(\theta_a \leq \theta) = F_a\left(\frac{M_t(\theta)}{\overline{M}_t}\right), \tag{3.10}$$

where F_a is a cumulative distribution function defined over the unit $[0,1]$ interval. Distribution F_a accounts for the structural features of the industry that relate to the accessibility of different tiers of productivity for a potential entrant. In some industries, most entrants may be able to draw upon the ideas of firms close to the frontier, while in others they may be

equally likely to draw upon lower-tier ideas.

As an example for the distribution of adoption, consider one in which the distribution is directed toward the more productive ideas.[15] Let F_a be defined over $x \in [0,1]$ as

$$F_a(x) = x^{\mu}, \qquad \mu \geq 1. \tag{3.11}$$

When $\mu > 1$, this distribution has a bias toward $x = 1$, since the mean of the distribution for the ranks is given by $\mu/(1+\mu)$. The corresponding probability density function f_a on the productivity ranks is monotonically increasing; we have $f_a(0) = 0$ and $f_a(1) = \mu$.[16]

If we assume $\mu = 1$, we find a uniform distribution F_a and we recover the common assumption that entrants uniformly adopt ideas from the current distribution of productivities:

$$\mathbb{P}(\theta_a \leq \theta) = \frac{M_t(\theta)}{\overline{M}_t}. \tag{3.12}$$

For its simplicity, this particular form has been a popular choice in prior work (see, e.g., Perla *et al.*, 2015; Atkeson and Burstein, 2015).[17]

Knowledge Transfer Once an entrant adopts the idea θ_a, the productivity θ_e with which it begins its production is given by

$$\theta_e = e^{\gamma_e} \widetilde{\theta}_t^{\eta} \, \theta_a^{1-\eta}, \tag{3.13}$$

where $\eta \in [0,1]$ is the degree of decreasing returns to adoption, γ_e is the innovation leap of entrants, and $\widetilde{\theta}_t$ is an industry-wide index of the quality of firm ideas. In what follows, I

[15] In the online appendix, I consider an alternative example, an atomic adoption distribution where all the weight is concentrated in firms with an exogenously given rank.

[16] We can think of distribution (3.11) as a special case of a beta distribution, which can produce a range of different means and dispersions over the unit interval. Here, I consider a subset of the parameter space by requiring that the likelihood of adoption from the frontier be nonzero, that is, $f_a(1) > 0$, which yields Equation (3.11) for the CDF of the distribution.

[17] The assumption of uniform entry does not match empirical facts on the small size of new entrants and their high likelihood of exit relative to the average firm in the industry (Dunne *et al.*, 1988; Arkolakis, 2016), unless if we include decreasing returns to adoption as in Equation (3.13). Nevertheless, it provides a reasonable theoretical benchmark against which we can compare other possibilities.

assume that $\widetilde{\theta}_t$ is the average productivity of the industry.[18]

When ideas are embodied in the structure and organization of firms, their transfer from one firm to another is likely to be imperfect. In this way, the specification above broadens the scope of innovation from those fully codifiable and transferrable, e.g., through patents, to ones that involve tacit and organizational knowledge.[19]

The formulation of entry in Equations (3.13) and (3.10) nests several previous models. If we assume $\eta = 0$, we find the case common in standard models of growth, in which knowledge perfectly transfers from incumbents to entrants (e.g., Luttmer, 2007; Perla *et al.*, 2015). For instance, assuming an atomic adoption distribution with all the weight concentrated on the zeroth rank, we find the specification of Luttmer (2012). If we consider the case of $\eta = 1$ and further assume a stochastic distribution on γ_e, we find the formulation of entry of Sampson (2016) (when $\widetilde{\theta}_t$ is the average industry productivity), or the no-spillover formulation of Atkeson and Burstein (2010) (when $\widetilde{\theta}_t$ is a constant). The intermediate cases with $\eta \in (0, 1)$ characterize situations where the entrants only imperfectly adopt the ideas of incumbents.

3.2.4 Stationary Constant Growth Paths

In this section, I will characterize the asymptotic behavior of the industry when aggregate productivity grows at a constant rate and the distribution of firm size is stationary. Since the focus of the paper is on the asymptotics, I will defer many of the details to the appendix and only present the structure required for the definition of the stationary constant growth paths. Appendix C.2.1 sets up the dynamic problem of an incumbent firm choosing its innovation γ and exit threshold, as well as a cohort of entrants choosing their rate of entry.

[18] In the Appendix, I generalize this to an aggregator function that combines the current distribution of ideas with varying degrees of substitutability between the ideas of firms.

[19] This account of innovation closely matches with the concept of investments in intangible capital. Intangible investments are those costs that firms incur in product design, marketing, and organizational development (Corrado *et al.*, 2009; Aizcorbe, 2009; Corrado and Hulten, 2010).

Combining the two, it defines the industry equilibrium path and the general equilibrium of the economy.

The next definition formally introduces the concept of a stationary constant growth path. We can think of this as the heterogenous-firm counterpart to the steady state of standard growth models. Among the potential equilibrium paths, it limits attention to the set that is of empirical interest.[20]

Definition 5. *(Stationary Constant Growth Path Equilibrium)* An equilibrium path along which a nonzero flow of entrants continuously arrive in the industry, and,

1. The mass and average productivity of firms asymptotically grow at constant rates, that is, $\lim_{t\to\infty} \dot{\overline{M}}_t/\overline{M}_t = g_M$ and $\lim_{t\to\infty} \dot{\overline{\theta}}_t/\overline{\theta}_t = g_\theta$,

2. The distribution of firm size converges to a stationary distribution $\lim_{t\to\infty} \mathbb{P}\left\{l_t\left(\theta\right) \leq l\right\} = \mathbb{P}_l\left(l\right)$ with finite mean.

To characterize the stationary constant growth path, I need to transform the productivity states into a space where their distribution remains stationary. Since by definition the distribution of firm employment is stationary along such an equilibrium path, this transformed space has to be a one-one function of firm employment. Define the firm *efficiency* state z_t such that is satisfies[21]

$$e^{z_t(\theta)} \equiv \frac{1}{(\rho+1)\,\psi_f\,w_t}\frac{N_t}{\overline{M}_t}\left(\frac{\theta}{\overline{\theta}_t}\right)^{\rho} \tag{3.14}$$

This state is the ratio of the gross profits to fixed costs for a θ-productivity firm that chooses *not* to invest in productivity growth (hence, $J = 1$). It captures the position of the firm

[20]Other potential paths are those that feature, e.g., no entry (Luttmer, 2012, characterizes such equilibria in a setting similar to the one here), or nonstationary distributions of firm size with explosive dispersion. Empirical work has typically found stable distributions of firm size across a wide range of industries (Luttmer, 2010), and broad industries such as those that are the focus of this paper almost always exhibit continuous entry of new firms. I leave a full theoretical characterization of all potential equilibria in the model for future work.

[21]It is straightforward to show that under the condition $\lim_{t\to\infty} \dot{\overline{M}}_t/\overline{M}_t = g_N$, the distribution of any monotonic function of $\theta/\overline{\theta}_t$ will be stationary along a stationary constant growth path, and therefore all such functions provide alternative transformations. The particular choice in definition (3.14) is similar to the one used by Luttmer (2012) and helps simplify the form of the differential equation characterizing the value of function of firms along a stationary constant growth path.

relative to the other firms in the industry. A firm with efficiency $z = 0$ earns gross profits that exactly compensate for its fixed costs.

Along a constant growth path, the wage rate converges to a constant, since per-capita expenditure is normalized to unity. We can show that a necessary condition for the existence of a stationary constant growth path is that the rate of growth of the total mass of firms is the same as the rate of growth of population $g_M = g_N$ (see Appendix C.2.1). Since the average firm employment remains constant and since firm employment and profits are proportional in a stationary distribution, the growth in aggregate sales has to be absorbed by the rising mass of firms.

Since $g_M = g_N$ and the wage rate w_t is asymptotically constant, the evolution of firm efficiency states along a stationary constant growth path is given by the brownian motion process

$$dz_t = \rho \left(\gamma - g_\theta - \frac{1}{2}\sigma^2 \right) dt + \rho \, \sigma \, d\mathcal{W}_t. \tag{3.15}$$

This expression shows that we can define the *rate of profit erosion* $g \equiv \rho \left(g_\theta + \frac{1}{2}\sigma^2 \right)$ as the rate at which growth in the aggregate productivity erodes the profits of any firm active in the industry. Alternatively, we can think of this as the rate of depreciation of ideas in the industry. As we might expect, this rate depends on the rate of aggregate productivity growth g_θ, volatility σ, and the index of market competition ρ.

Along a stationary constant growth path, we can write the evolution of the industry measure as

$$M_t(\theta) = \overline{M} \, e^{g_M t} \times H(z_t(\theta)), \tag{3.16}$$

where H is a cumulative distribution function. The measure asymptotically exhibits the so-called *traveling-wave* behavior, since it is a stationary function of $\log \theta - g_\theta t$ (Lucas and Moll, 2014). The total mass of firms grows at a constant rate, and their productivity distribution travels at a constant rate g_θ in the log-productivity space. The last condition in the definition of the stationary constant growth path requires that the moment $\mathbb{E}_H [e^z J]$ should be finite, ensuring that a well-defined average size of firm employment exists (this is the only empirically relevant case, see Luttmer, 2010).

The remainder of this section studies the behavior of firms and the aggregate industry along a stationary constant growth path. I will first characterize the firm value function and optimal policies, and then characterize the distribution of firm productivity H. Finally, I combine the two and examine how the rate of profit erosion g is determined through the zero entry condition.

3.2.5 Incumbent Problem

Along a stationary constant growth path, we can write the value of a firm with efficiency state z as $\psi_f w_t V(z)$. So long as the firm continues to operate, standard arguments imply that the value function $V(z)$ satisfies the stationary Hamilton-Jacobi-Bellman (HJB) equation

$$rV(z) = \max_{\gamma} e^z J(\gamma) - 1 + \rho \left(\gamma - g_\theta - \frac{1}{2}\sigma^2\right) V'(z) + \frac{\rho^2}{2}\sigma^2 V''(z), \tag{3.17}$$

for $z \geq z_o$ where z_o is the exit threshold and the function J is defined by Equation (3.6). The following proposition characterizes the solution to this problem. The derivation of the stationary HJB equation and the proof of the proposition in a more general setting are included in Appendix C.2.1.

Proposition 6. *Assume that the innovation cost function $\varphi(\cdot)$ satisfies the conditions stated in Equation (12), and that the rate of productivity growth g_θ is large enough so that*

$$r + \rho\, g_\theta - \frac{\rho}{2}(\rho - 1)\sigma^2 > \rho\overline{\gamma} - \frac{J(\overline{\gamma})}{J'(\overline{\gamma})}. \tag{3.18}$$

Then the HJB equation (3.17) has a unique, continuous, convex, and monotonically increasing value function $V(\cdot)$ that converges to a stable, asymptotic function $\lim_{z\to\infty} V(z) = u^* e^z$. Moreover, the corresponding policy function $\gamma^*(z)$ is unique, continuous, and monotonically increasing. It satisfies $\gamma^*(z_o) = 0$ for exiting firms, and an asymptotic Gibrat's law, in the sense of $\lim_{z\to\infty} \gamma^*(z) = \gamma^{**}$, for the largest firms. The asymptotic rate of productivity growth of the largest firms γ^{**} satisfies the maximum in the following problem

$$u^* = \max_{\gamma} \frac{(1 + \varphi(\gamma))^{-\rho}}{r + \rho\left(g_\theta - \gamma - \frac{1}{2}(\rho - 1)\sigma^2\right)}. \tag{3.19}$$

Proposition 6 establishes the conditions that ensure the existence of a uniquely stable equilibrium for the growth strategies of incumbent firms. The key requirement involves a lower bound on the effective discount rate of firms. This result is intuitive: since productivity gains are permanent, the discount rate has to be large enough to contain firms from infinite investments in growth.

The convexity of the value function follows from the fact that firms face different expected lifetimes of activity in the industry. Productive firms lie farther from the exit cutoff and expect to live longer in the industry. A marginal increase in the productivity of a firm raises its value along two margins: it increases the flow of profits as well as the expected lifetime of the firm. This implies that the marginal net present value of productivity growth is larger for larger firms. Due to their shorter expected lifetimes, unproductive firms have less to gain from investing to permanently raise their efficiency. Facing higher returns to investment, large, high-efficiency firms invest more than small, low-efficiency firms, giving rise to an increasing investment policy function $\gamma^*(\cdot)$. The proposition further shows that the value function becomes linear in e^z in the limit of very large firms. As a result, the positive relationship between productivity (size) and innovation investments declines for very large firms and we find that Gibrat's law holds in the limit.

As expected, factors that contribute to higher discounting of the net present value of incumbent profits have a negative impact on the firms' value and innovation investments. In particular, the rate of aggregate productivity growth g_θ negatively affects the firms' incentives to grow.[22] To see this, consider a firm that is forced to remain active at all productivity levels. The value function of this firm is simply given by $J(\gamma^{**})\, e^z/r^* - 1/r$ where γ^{**} is the firm's productivity growth and $r^* = r + \rho\left(g_\theta - \gamma^{**} - \frac{1}{2}(\rho - 1)\sigma^2\right)$ is its effective discount rate. Increasing r^*, say, through an increase in the rate of productivity growth g_θ, erodes the future value and therefore the rate of return to innovation investments.

[22]The same negative relationship between the rate of creative destruction and the incentives to grow exists in canonical models of innovation and growth that feature both entry and incumbent innovation (see, e.g., Klette and Kortum, 2004; Acemoglu and Cao, 2015). The mechanism in both cases is the erosion of future profits along the growth path, which lowers the expected private lifetime of benefiting from innovation. The main difference is that here the effect is stronger on low-productivity firms, whereas it is uniform under the canonical model.

For high-productivity firms that have large efficiency z, the option to exit the industry is not relatively valuable. Therefore, this analysis provides a good approximation for the effect of faster growth on their incentives to innovate.

For less productive firms that lie close to the exit cutoff, faster growth creates additional considerations since it affects their valuation of the option to exit. Of course, the option to exit the industry is relatively more valuable for these firms. A rise in the rate of productivity growth g_θ changes the option value of exit since it affects first, how fast the relative position of a firm in the industry erodes, and second, at which efficiency level it ceases operation (the exit cutoff). The latter effect only emerges when the extent of volatility is high and low-productivity firms account for the option value of continuation in their exit decisions. Together, these forces imply that the dynamic competition effect of faster growth creates heterogeneous responses in the innovation investments of firms.

3.2.6 Industry Problem

As we saw in Section 3.2.7, the mass of active firms at time t along a stationary constant growth path takes the form $M_t(z) = \overline{M} e^{g_M t} H(z)$, where $H(\cdot)$ is the cumulative distribution of firm efficiencies. This stationary distribution emerges as the result of the innovation and exit decisions of all firms, and characterizes the long-run state of the industry. In this section, I characterize this distribution and provide conditions required for its existence.

Defining the industry-wide index of firm efficiency through

$$e^{\tilde{z}} \equiv \left(\int e^{z/\rho} h(z) \, dz \right)^{\rho},$$

we can transform the knowledge transfer Equation (3.13) to the space of firm efficiencies as

$$z_e = \eta \tilde{z} + (1 - \eta) z_a + \rho \gamma_e.$$

The distribution of the efficiency with which entrants attempt entry is then given by $F_e(z) \equiv H\left(\frac{z - \eta \tilde{z} - \rho \gamma_e}{1 - \eta}\right)$.

Along a stationary equilibrium path, the exit cutoff for firms corresponds to a cutoff

in the space of efficiency. Let z_o denote this cutoff, and $\underline{z}_a$ denote the lowest level of incumbent efficiency that an entrant can adopt to successfully enter the industry, satisfying $\eta\tilde{z} + (1-\eta)\,\underline{z}_a + \rho\gamma_e = z_o$. If $\rho\gamma_e < -\eta\,(\tilde{z} - z_o)$, this level of efficiency corresponds to the one of an active incumbent firm; therefore, a flow of firms do not succeed in their attempt to enter the industry and immediately exit. In this case, the gross rate of entry is larger than the rate of successful entry by a factor of $[1 - H\,(\underline{z}_a)]^{-1}$. Otherwise, all entry attempts are successful. We can accordingly define a distribution of entry efficiency F with a cumulative distribution function

$$F\,(z) = \begin{cases} \frac{F_e(z) - F_e(z_o)}{1 - F_e(z_o)}, & \rho\gamma_e < -\eta\,(\tilde{z} - z_o)\,, \\ F_e\,(z)\,, & \text{otherwise.} \end{cases}$$

Let $h\,(z) \equiv H'\,(z)$ denote the probability density function for the stationary distribution, and λ_e and λ_o be the rates of (successful) entry and exit. The distribution H then has to satisfy the following stationary Kolmogorov Forward Equation (KFE) for $z > z_o$

$$g_M\,H\,(z) = \lambda_e\,F\,(z) - \lambda_o - \rho\left(\gamma^*\,(z) - g_\theta - \frac{1}{2}\sigma^2\right)h\,(z) + \frac{\rho^2}{2}\sigma^2\,h'\,(z)\,, \tag{3.20}$$

along with boundary conditions $h\,(z_o) = 0$ and $\lim_{z\to\infty} H\,(z) = 1$, where $\gamma^*\,(\cdot)$ is the incumbent investment policy solving problem (3.17). As z goes to infinity (zero), the cumulative and probability distribution functions H and h converge to 0 and 1 (0 and 0), respectively. Equation (3.20) then implies a rate of exit given by

$$\lambda_o = \frac{1}{2}\rho^2\sigma^2 h'_+\,(z_o)\,, \tag{3.21}$$

and a rate of (successful) entry λ_e given by the conservation of mass condition $\lambda_e = \lambda_o + g_M$

.

The empirically relevant equilibria are those in which the industry features both a stable and stationary distribution of firm size and a constant stream of entrants and exiting firms. Therefore, it is natural to search for industry equilibria that feature this property. The following proposition gives us the necessary and sufficient conditions required to establish

the existence of a stationary distribution of firm efficiency and size along with a positive rate of entry. It also establishes that the distribution will always have a right tail that exhibits Pareto behavior.

Proposition 7. *(Stationary Distribution) Assume that at least one of the two following conditions holds: either adoption is imperfect $\eta > 0$, or the likelihood of adoption from the frontier is zero $f_a(1) = 0$.*[23] *Then, for a stationary distribution of firm size with a positive rate of entry λ_e and finite mean to exist, the rate of growth in the mass of firms should be large enough, in the sense that $g_M > 0$ and $g_M > \rho(\gamma^{**} - g_\theta) + \frac{1}{2}\rho(\rho - 1)\sigma^2$. The rate of entry for any other potential stationary distribution will necessarily be zero (if such a distribution exists). The right tail of the distribution exhibits a Pareto behavior with a tail index given by*[24]

$$\zeta \equiv \lim_{z\to\infty} \frac{h(z)}{1 - H(z)} = \frac{1}{\rho}\left[\sqrt{\left(\frac{g_\theta - \gamma^{**}}{\sigma^2} + 1\right)^2 + \frac{2g_M}{\sigma^2}} + \frac{g_\theta - \gamma^{**}}{\sigma^2} + 1\right]. \tag{3.22}$$

Proposition 7 imposes a lower bound on g_M in order to establish the existence of the stationary distribution. The intuition is also similar to the condition we found in Proposition 6: we require a flow of entrants that is large enough to prevent the most productive firms from indefinitely growing their share of employment (Luttmer, 2012). Furthermore, it requires that no entrants receive spillovers directly from the frontier technology, or if they do, then the .

3.2.7 Equilibrium Industry Growth

Combining Propositions 6 and 7, we can now close the model through a free entry condition that pins down the rate of productivity growth in the long-run equilibrium.

I assume that the rate of growth of population (aggregate sales) is positive $g_N > 0$.

[23] In the case where adoption is perfect $\eta = 0$ and $f_a(1) > 0$, a necessary condition for the existence of a stationary equilibrium is that $\gamma_e < 0$. In addition, the rate of growth in the mass of firms g_M has to satisfy a more stringent condition to ensure the existence of an equilibrium. The characterization of this case is available upon request.

[24] Note that $-\zeta$ is the negative root of the quadratic characteristic equation $\mathsf{Q}(x) = \frac{1}{2}\rho^2\sigma^2x^2 + \rho(\gamma^{**} - g)x - g_M$, corresponding to the KFE characterizing the stationary distribution.

As discussed earlier, the mass of firms grows at the rate $g_M = g_N$ and therefore the first condition for the existence of the stationary distribution in Proposition 7 is satisfied. Defining the relative entry to fixed costs $\psi \equiv \psi_e/\psi_f$, we can write the free entry condition as

$$\psi = \int V\left(\eta\tilde{z} + (1-\eta)\, z + \rho\gamma_e\right) f\left(H\left(z\right)\right) h\left(z\right) dz = \left(1 - F_e\left(z_o\right)\right) \mathbb{E}_F\left[V\left(z\right)\right], \tag{3.23}$$

where the pre- and post-entry distribution of efficiencies F_e and F are defined as in the previous section. Equation (3.23) pins down the rate of productivity growth g_θ, which is ceteris paribus decreasing in the ratio of entry to fixed costs ψ.

All the conditions required for the existence of the value function and the stationary distribution together translate into a constraint on the relative costs of entry, in the form of $\psi \in \left(\underline{\psi}, \overline{\psi}\right)$. Such a condition ensures, for any relative cost ψ within the interval, the existence of a unique stationary distribution H of efficiencies with nonzero rate entry λ_e, such that

$$\lambda_e = g_N + \frac{1}{2}\rho^2\sigma^2\, h'_+\left(z_o\right).$$

The relative market share of a firm in state z in the aggregate sales of the industry is given by

$$\nu\left(z\right) = \frac{e^z J\left(\gamma^*\left(z\right)\right)}{Z}, \tag{3.24}$$

where I have defined an index of average firm size $Z \equiv \mathbb{E}_H\left[e^z J\left(\gamma^*\left(z\right)\right)\right]$. Variable ν compares the sales of a firm at efficiency level z with the average sales of the industry; therefore, it averages to one across the industry , $\mathbb{E}_H\left[\nu\left(z\right)\right] = 1$. Therefore, market shares $\nu\left(\cdot\right)$ define a distribution, which I also denote by ν_s. The expectation of this distribution over any firm-level outcome $X\left(\cdot\right)$ in the efficiency space defines a market-share-weighted average of that outcome given by $\mathbb{E}_\nu\left[X\left(z\right)\right] \equiv \mathbb{E}_H\left[e^z J\left(\gamma^*\left(z\right)\right) X\left(z\right)\right]/Z$.

The following proposition provides a decomposition of the rate of average productivity (and hence the rate of profit erosion) to the components stemming from incumbent investment as well as the margin of entry and exit.

Proposition 8. *(Growth Decomposition) The overall rate of profit erosion of the industry can be*

decomposed into

$$\begin{aligned}\mathbb{E}_\nu\left[1+\varepsilon_{J,1}\right]\cdot g \;&=\; \rho\,\mathbb{E}_\nu\left[\gamma^*(z)\left(1+\varepsilon_{J,1}\right)\right]+\frac{\rho^2}{2}\sigma^2\left(\mathbb{E}_\nu\left[\left(1+\varepsilon_{J,1}\right)^2\right]+\mathbb{E}_\nu\left[\varepsilon_{J,2}\right]\right)\\&\quad+(\nu_e-\nu_o)\cdot\lambda_o-(1-\nu_e)\cdot g_M \qquad (3.25)\end{aligned}$$

where ν_e and ν_o denote the market shares of entrants and exiting firms, given by $\mathbb{E}_F\left[e^z J\right]/Z$ and e^{z_o}/Z, respectively, and ε_1^J and ε_2^J denote first and second order semi-elasticities of the innovation costs with respect to efficiency, given by $\varepsilon_{J,1}\equiv\frac{\partial}{\partial z}\log J\left(\gamma^*(z)\right)$ and $\varepsilon_{J,2}\equiv\frac{\partial^2}{\partial z^2}\log J\left(\gamma^*(z)\right)$.

The decomposition in Equation (3.25) has a straightforward interpretation. The first term on the right hand side gives the share-weighted growth rates of firms at different efficiency states. This term is driven by the endogenous investments of incumbent firms. The second term captures the contribution of volatility to growth. Idiosyncratic shocks may also contribute to aggregate growth as market shares reallocate toward firms that receive better shocks. We can interpret this term as the contribution of involuntary experimentation of firms in the industry. Together, the two expressions on the first line summarizes the contribution of incumbent firms to productivity growth.

The third and fourth terms, on the second line of the equation, summarize the contribution of entry and exit. The third term captures the contribution of selection (entry and exit). New firms replace firms at the exit threshold at a rate given by the exit (turnover) rate λ_o. Each new entrant improves the efficiency of an exiting firm by a factor $\left(\mathbb{E}_F\left[e^z J\right]-e^{z_o}\right)/Z$, that is, the difference between the average size of entrants and exiting firms, relative to the average firm size. This expression in turn depends on the gap in the productivity of the entrants and exiting firms.

Finally, the fourth term accounts for another distinct contribution of entry, stemming from a potential gap between the productivity of entrants and the average incumbent. The net rate of entry of firms is given by the rate of growth in the mass of firms g_M. Each new firm faces a deficiency of size $1-\mathbb{E}_F\left[e^z\right]/Z$ relative to an average incumbent. Therefore, a faster rate of growth in the number of firms can reduce the overall rate of growth of productivity in the industry.

To more clearly see the intuition behind the decomposition, let us simplify it by assuming that $\gamma^*(z)$ is approximately constant across two high- and low-innovation regions of the efficiency space. Therefore, the terms involving the elasticities of the firm investment policies drop out and we find

$$g_\theta \approx \underbrace{\mathbb{E}_\nu\left[\gamma^*(z)\right] + \frac{1}{2}(\rho-1)\sigma^2}_{\text{Incumbents}} + \underbrace{\left(\frac{\nu_e - \nu_o}{\rho}\right)\cdot\lambda_o}_{\text{Turnover}} - \underbrace{\left(\frac{1-\nu_e}{\rho}\right)\cdot g_M}_{\text{Net Entry}}, \tag{3.26}$$

where $\mathbb{E}_\nu\left[\gamma^*(z)\right]$ is the market-share-weighted average of incumbent innovations. Innovation investments of incumbent firms influence the rate of productivity growth through two distinct channels: 1) directly, through the first term in Equation (3.26), and 2) indirectly, through their effects on the contribution of turnover. The second effect happens both because incumbent innovation raises the rate of turnover, and because it influences the gap between entering and exiting firms. Despite the fact that firms invest in improving their own processes, they still destory the bunsiness of unproductive firms and push them out of the industry. Here, creative destruction operates through the selection margin.

Typically, any shock to the environment or any policy intervention shifts the composition of growth in the industry between the different components above. Consider again a small rise in the ratio of entry to fixed costs ψ. From the free entry condition, we expect the aggregate rate of profit erosion g to respond negatively to this shock. Equation (3.25) allows us to further quantify the resulting shift in the composition of growth. Based on what we saw from the behavior of incumbent investments, they will rise in response to this fall in the rate of profit erosion g. We also expect the rate of entry $\lambda_o + g_M$ to fall as a result of the rise in relative entry costs. Therefore, the role of incumbent investment in overall industry growth rises in response to this shock.

3.2.8 Solving for the Model

As with most models of investment under uncertainty, we cannot analytically characterize the solution to either the incumbent or the industry problem.[25] Instead, I employ the following scheme to compute numerical solutions to both probelms.

In order to solve the Bellman equation in Proposition (6), I first transform the problem into a first-order system of differential equations $\boldsymbol{u}'(z) = \mathcal{Q}(\boldsymbol{u}(z))$ defined in a two-dimensional state space $\boldsymbol{u} \in \mathbb{R}^2_+$ where

$$u_1(z) = \left(V(z) + \frac{1}{r}\right) e^{-z}, \tag{3.27}$$

$$u_2(z) = V'(z) e^{-z}. \tag{3.28}$$

We have a Boundary Value Problem (BVP) with the following two boundary conditions: 1) the smooth pasting condition implies $u_2(z_o) = 0$, and 2) Proposition 6 implies $\lim_{z\to\infty} u_1(z) = u^*$ where u^* is given by equation (3.19). Figure 3.1 gives an example of how the solution typically behaves in the $\boldsymbol{u}$-space. The standard smooth-pasting condition implies that the path has to begin somewhere on the u_1-axis $\boldsymbol{u}(z_o) = (u_1(z_o), 0)^t$, which is located to the southwest of the equilibrium point such that $u_1' \leq 0$ and $u_2' \geq 0$. From the standard value-matching condition $V(z_o) = 0$ and Equation (3.27), this point pins down the exit threshold

$$e^{z_o} = \frac{1}{r\, u_1(z_o)}.$$

For the infinite boundary, the solution has to converge to the point $\boldsymbol{u}^* = (u^*, u^*)$, which is an *equilibrium point* for the system of differential equations in the sense that $\boldsymbol{Q}(\boldsymbol{u}^*) = 0$. Condition (3.18) both ensures that such an equilibrium point exists, and that it is locally saddle path stable. In other words, the BVP has a unique solution, characterized by a path

[25] Nevertheless, we can use approximations to transform the two problems from differential equations into algebraic equations. In the online appendix I showcase different properties of the incumbent and the industry problems under a specific innovation cost function and an atomic adoption distributoin through both numerical solutions and analytical approximations. However, even in this case, or eve cases where the value function can be analytically computed, the resulting expressions involve nonlinearities that do not give rise to transparent comparative statics. More generally, the difficulty in analytical study of comparative statics is a well-known feature of models of investment under uncertainty (Dixit and Pindyck, 1994). Dixit (1991) suggests analytical approximations based on Taylor expansions as an alternative approach.

of $\boldsymbol{u}(z)$ that converges to $(u^*, u^*)^t$. Figure 3.1 shows both the stable and unstable paths around the equilibrium point.

I derive a similar scheme to solve for the stationary distribution of efficiencies $H(\cdot)$ for any given specification of the model. I transform the problem into a first-order BVP $\boldsymbol{h}'(z) = \boldsymbol{\mathcal{P}}(\boldsymbol{h}(z))$ defined in a two-dimensional state space of $\boldsymbol{h}(z) \equiv (H(z), h(z)) \in \mathbb{R}^2_+$. The problem has the following boundary conditions: $h(z_o) = H(z_o) = 0$ and $\lim_{z\to\infty} H(z) = 1$. Figure 3.2 gives an example of how the solution typically behaves in the $\boldsymbol{h}$-space. In the limit of large z, the path of the solution has to converge to the point $\boldsymbol{h}^* = (1, 0)^t$ that forms an equilibrium point for the system of differential equations in the sense that $\boldsymbol{\mathcal{P}}(\boldsymbol{h}^*) = 0$ (another equilibrium point $(0, 0)^t$ is also on the path). The conditions in Proposition 7 ensure that this equilibrium point is locally saddle path stable. Figure 3.2 further shows both the stable and unstable paths around the equilibrium point $\boldsymbol{h}^* = (1, 0)^t$.

I use different generalizations of the method above to solve for the variations of the Bellman equation encountered in the cases involving distortions and the social planner's problems.

Figure 3.1: *The solution to the Bellman equation in the two-dimensional space of* $(u_1(z), u_2(z))$ *for a given choice of model parameters (details in the online appendix). The relative sizes of arrows on the path of* $u(z)$ *indicate the relative values of the derivative of* u *with respect to* z *along the path. The equilibrium point* u^**, which lies on the* 45° *line, is shown with a red dot. The stable arm around the saddle-path stable equilibrium point is also shown in light gray. For very large firms, as* $z \to \infty$*, the path of* $u(z)$ *asymptotically converges to this arm.*

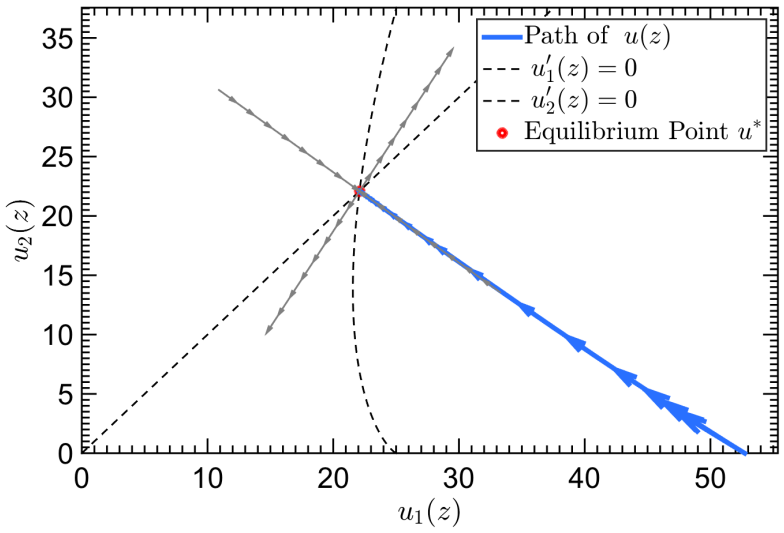
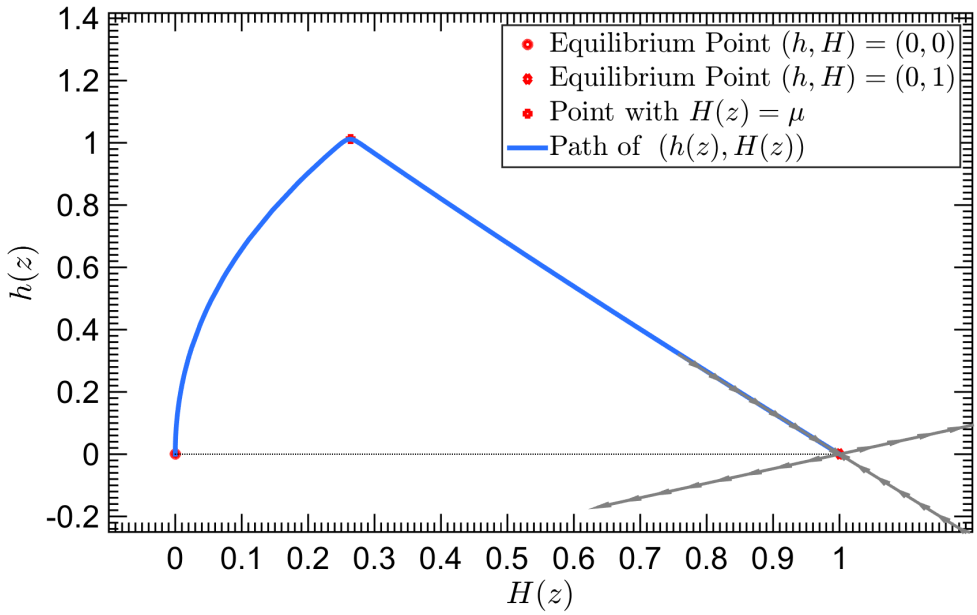


Figure 3.2: *The solution to the differential equation characterizing the stationary distribution in the space of* $(h(z), H(z))$ *for a given choice of model parameters and an atomic distribution of adoption* F_a *(details in the online appendix). The two equilibrium points and the point of entry are shown in red. The stable arm around the saddle-path stable equilibrium point* $(H, h) = (1, 0)$ *is shown in light gray. For very large firms, as* $z \to \infty$*, the path of* (H, h) *asymptotically converges to the stable arm.*



3.3 Optimal Allocations

Let us now examine the allocative efficiency of the market equilibrium by comparing it against the allocations advised by a social planner. As mentioned in the Introduction, we find that the potential gaps between the social and private returns to innovation are substantially different in this model from those in standard models that do not include selection and knowledge diffusion. In particular, the externalities created by firm-level innovations on the growth of industry-level productivity vary with firm productivity.

Section 3.3.1 begins by characterizing the socially optimal allocation of innovation and production. Section 3.3.2 discusses the different channels through which innovation investments generate externalities on industry-level productivity growth in this model. Section 3.3.3 then briefly discusses the numerical scheme used for solving for the optimal allocation. Finally, Section 3.3.4 presents two different schemes that the social planner may

employ to implement the optimal allocation, under the assumption that she can observe the productivity ranking of firms. The implementation allows us to study the sizes of gaps between social and private returns to innovation investments.

3.3.1 Socially Optimal Allocations

In the environment of this model, for incumbent firms at any given level of efficiency the social planner has to decide on 1) the extensive margin of production (selection), i.e., whether the firms should be selected to continue operation, 2) the intensive margin of production, i.e., how much output and labor input to allocate to them, and 3) innovation investments. In addition, the social planner decides on the rate of entry into the industry.

Conditional on the extensive margin of production and the choices of firm innovation investments, the first-order conditions corresponding to the static allocations of output and employment coincide between the market equilibrium and the socially optimal allocations. This result directly follows from the assumption of CES demand and constant markup pricing. Therefore, the intensive margin of production decisions (output and labor input) is efficient, conditional on the extensive margin (exit) and innovation investments. However, the extensive margin, whether the firm continues to operate or exits, may vary between the market equilibrium and the socially optimal allocation.[26] In addition, differences between the social and market allocations may arise in the innovation investment decisions of firms. I will denote the socially optimal efficiency cutoff for production as $z_{o,s}$, and the socially optimal innovation investments of firms as $\gamma_s(\cdot)$ for $z \geq z_{o,s}$.

The following proposition presents the key result that characterizes the socially optimal allocations.[27]

[26] In the static or symmetric setting, it is well known that with CES preferences business stealing and appropriability effects exactly cancel out and the market entry (and exit) of firms and products becomes optimal (Grossman and Helpman, 1991a, p. 82). The framework offered here is flexible enough to allow for distortions due to potential variations in markups that exist in more general demand specifications (e.g., Dhingra and Morrow, 2012).

[27] The proposition provides the necessary conditions for a constant growth path to characterize a socially optimal allocation in the long-run. Characterizing sufficient conditions for the existence of this equilibrium remains for future work.

Proposition 9. *(Socially Optimal Allocations) Assume that the socially optimal allocation has a stationary constant growth path characterized by a rate of productivity growth $g_{\theta,s}$, the social value function $V_s(\cdot)$, the CDF and pdf of the stationary distribution of efficiencies $H_s(\cdot)$ and $h_s(\cdot)$, respectively, and the rate of entry $\lambda_{e,s}$. Social value function V_s satisfies a social HJB equation*

$$\begin{aligned} rV_s(z) &= \max_{\gamma} e^z J(\gamma) - 1 + \rho\left(\gamma - g_{\theta,s} - \frac{1}{2}\sigma^2\right) V_s'(z) + \frac{\rho^2}{2}\sigma^2 V_s''(z) \\ &\quad + \lambda_{e,s} \cdot (U_s(z) - \Psi_s), \end{aligned} \tag{3.29}$$

for $z \in (z_{o,s}, \infty)$ and the boundary conditions $V_s(z_{o,s}) = V_s'(z_{o,s}) = 0$, where the social crowding cost Ψ_s and the spillover function $U_s(\cdot)$ are defined as follows. The social crowding and spillover are each composed of two separate contributions due to knowledge diffusion and to the industry-wide average stock of knowledge given by

$$\begin{aligned} U_s(z) &= \eta\left(e^{\frac{z-z_0}{\rho}} - 1\right)\tilde{u}_s + (1-\eta)\, u_s(z), \\ \Psi_s &= \eta\left(e^{\frac{\tilde{z}-z_0}{\rho}} - 1\right)\tilde{u}_s + (1-\eta)\,\psi_s, \end{aligned}$$

where the knowledge diffusion terms are given by

$$\psi_s \equiv \int_{z_{o,s}}^{\infty} V_s'(\eta\tilde{z} + (1-\eta)x + \rho\gamma_e)\, f_a(H_s(x))\,(1 - H_s(x))\, dx, \tag{3.30}$$

$$u_s(z) \equiv \int_{z_{o,s}}^{z} V_s'(\eta\tilde{z} + (1-\eta)x + \rho\gamma_e)\, f_a(H_s(x))\, dx, \tag{3.31}$$

and the term corresponding to the contribution to the industry-wide knowledge stock is given by

$$\tilde{u}_s \equiv \rho e^{-\frac{\tilde{z}-z_0}{\rho}} \int_{z_{o,s}}^{\infty} V_s'(\eta\tilde{z} + (1-\eta)x + \rho\gamma_e)\, f_a(H_s(x))\, h_s(x)\, dx. \tag{3.32}$$

In addition, H_s satisfies a corresponding KFE and, if $\lambda_{e,s} > 0$, the socially optimal value function and efficiency distribution satisfy a social free entry condition

$$\psi = \int_{z_{o,s}}^{\infty} V_s(\eta\tilde{z} + (1-\eta)z + \rho\gamma_e)\, f_a(H_s(z))\, h_s(z)\, dz, \tag{3.33}$$

where, as before, we have $\psi \equiv \psi_e/\psi_f$.

When the long-run equilibrium involves entry, the proposition shows that the distinction

between the social and private value of firms appears in a constant Ψ_s , and an efficiency-dependent knowledge spillover term $U_s(\cdot)$ in the social planner's HJB equation. The terms corresponding to the stock of knowledge are standard spillover terms: the external contribution of firms to the industry-wide knowledge stock is proportional to their productivity (that is, $e^{z/\rho}$). Raising this stock then generates social value by increasing the efficiency of all entrants, creating social value given by Equation (3.32).

For the remainder of this section, I focus on understanding expressions (3.30) and (3.31), corresponding to the contribution of knowledge diffusion. The two expressions capture simple intuitions about the workings of knowledge diffusion in this model. To better see these intuitions, let us consider a perturbation to the socially optimal (or the market equilibrium) allocation whereby we add a small mass Δm of firms with efficiency z to the long-run stationary distribution, removing a mass of the same size at random from the distribution. This small perturbation keeps the total mass of firms intact, but influences the ranking of firms and, accordingly, modifies the distribution of entrants. We can derive the social value of the knowledge spillovers created by this perturbation, that is, expressions (3.30) and (3.31), in three steps.

Step (I): Improving or Crowding of the Pool of Ideas First, consider an entrepreneur E who has received the draw of $\mu_E \in (0,1)$ as the ranking of the idea she adopts. The corresponding efficiency z_a then satisfies $\mu_E = H(z_a)$. The effect of the perturbation on her efficiency z_a depends on whether the efficiency z of the added firms is better or worse than her pre-perturbation adopted idea z_a. If the new idea is better $z > z_a$, this will improve the pool of ideas for entrepreneur E. This will change the rankings of ideas from the pre-perturbation $H(z_a)$ by a small amount

$$\Delta H \approx H(z_a) \times \frac{1}{1+\Delta m} - H(z_a) \approx -H(z_a)\,\Delta m.$$

Since her draw μ_E should remain constant, the perturbation pushes up her adopted efficiency by $\Delta z_a = H(z_a)/h(z_a)\,\Delta m$. On the other hand, if the new firms z are worse than her pre-

perturbation idea $z < z_a$, this will deteriorate the pre-perturbation ranking by

$$\Delta H \approx H(z_a) \times \frac{1}{1+\Delta m} + \frac{\Delta m}{1+\Delta m} - H(z_a) \approx (1 - H(z_a))\, \Delta m.$$

Therefore, the resulting change in her position will be $\Delta z_a = -(1 - H(z_a)) / h(z_a)\, \Delta m$. We can summarize the change in the two cases using the following two expressions: a constant negative term and a positive term that only operates for entrepreneurs E with ideas originally worse than z, as

$$\Delta z_a = \left[\frac{1}{h(z_a)} \mathbb{I}(z > z_a) - \frac{1 - H(z_a)}{h(z_a)} \right] \Delta m,$$

where $\mathbb{I}$ stands for an indicator function. We can think of the first term inside the bracket as whether or not z improves the pool of ideas for any given entrepreneur. The second term then is a constant crowding term that is identical for all firms z. Intuitively, adding any new firm makes it less likely for entrepreneurs to find their original ideas, and in this way the likelihood of adopting this new idea produces an opportunity (crowding) cost.

Step (II): Accounting for the Expected Lifetime of Ideas The calculation above accounted for the quality of knowledge spillovers generated by different firms. Next, we have to account for the social values associated with these spillovers. The social values of ideas depend not only on their quality, but also on the time horizon over which they benefit the industry. From the social planner's perspective, each change Δz_a in the position of an entrant generates an external value $V'(z_e)\, \Delta z_a$, where the marginal value is calculated at the entry efficiency z_e which is related to the adopted efficiency z_a through $z_e \equiv \eta \tilde{z} + (1 - \eta) z_a + \rho \gamma_e$.

Step (III): Summing Up the Effect Across All Entrants The two terms in Equations (3.30) and (3.31) sum up all of these external values across firms at different tiers of adopted efficiency z_a, accounting for the fact that the rate of adoption in each tier is given by $\lambda_e \times h(z_a)\, f_a(H(z_a))$.

The next corollary immediately follows from the proposition.

Corollary 5. *The market equilibrium is socially efficient if there are no spillovers, i.e.* $\eta = 1$*, and/or no entry happens along the long-run equilibrium, i.e.,* $\lambda_e = 0$*.*

When the distribution of productivity of new entrants is determined exogenously, market equilibrium innovation and production decisions become optimal. In the absence of spillovers, the efficiency results that Dhingra and Morrow (2012) have derived for the static benchmark model of firm heterogeneity and selection (Melitz, 2003) generalize to the dynamic case with firm innovation, which corresponds to the innovation model of Atkeson and Burstein (2010). Similarly, when the equilibrium path does not feature entry, knowledge spillovers do not play a role and the innovation and exit decisions of firms become socially optimal.

In addition to the characterization of the social value function, the proposition also provides the social free entry condition. Equation (3.33) shows that the optimal entry condition for the social planner takes the same form as the free entry conditions under the market equilibrium. While both the optimal distribution H_s and the value function V_s diverge to those under the market equilibrium, the relative value of entry (in terms of fixed operation costs) has to be the same between the two allocations.

3.3.2 Unpacking Innovation Externalities

Proposition 9 characterizes the socially optimal allocation. However, it does not directly show us how the market equilibrium and the socially optimal and innovation decisions deviate from each other for a firm in efficiency state z. These two investments satisfy

$$e^z J'(\gamma(z)) = V'(z), \qquad e^z J'(\gamma_s(z)) = V_s'(z), \tag{3.34}$$

respectively. Investigating the market equilibrium and social HJB equations, we find that the differences between the marginal private and social values of a firm with efficiency z derive from two distinct sources:

1. Knowledge spillovers: characterized by the endogenous distributions (H, H_s) and the exogenous adoption process F_a,

2. Selection (expected lifetime): characterized by the endogenous rates of productivity growth $(g_\theta, g_{\theta,s})$ and the exogenous volatility parameter σ.

To unpack these different sources, let us again use a partial-equilibrium perturbation argument. This time, consider raising the innovation investments of all firms within a Δz neighborhood of the efficiency level z by a small amount $\Delta\gamma$. I will examine the external impacts of this change that firms within Δz do not account for when they make their innovation decisions. The key is to note that, due to the dynamic effect of business stealing, other firms in the long-run equilibrium will respond to this change by lowering their own innovation investments. However, these long-run effects are fairly complex, since the initial change sets in motion a chain reaction of responses across different types of firms that propagates across the industry and creates a new equilibrium.

To isolate the main channels, I consider a partial-equilibrium setup whereby I consider only the *direct* effect of the change above on a group of firms at another efficiency level, say, x. Here, I assume that no other groups of firms respond except firms within a small neighborhood of x. Of course, one can then propagate these first-level responses to a second-level response, and so on. But the direct effect of this perturbation allows us to examine the main logic of the externalities.

If no other firms respond to the change $\Delta\gamma$ among firms z, the long-run rate of profit erosion rises to reflect the change. From the decomposition (3.26), we find this effect to be given by[28]

$$g \longrightarrow g + \nu(z) \times \Delta z \Delta\gamma,$$

where $\nu(z) = h(z) e^z J(z) / Z$ denotes the market share of firms with efficiency state z.

Externality (I): Knowledge Improving When firms with efficiency z raise their investments, they raise their likelihood of moving up in the efficiency ladder. This raises the

[28] I emphasize that this expression only captures the direct effect of the innovation, under the assumption that other incumbents and entrants do not change their strategies immediately. It is certainly possible that the general equilibrium response of other firms results in the fall of the rate of profit erosion in response to the perturbation considered here.

density of firms immediately above them, and accordingly changes the knowledge spillovers that entrants receive. Using the same argument as the previous section, we can see that the size of this change depends on the marginal social value of the resulting improvement in knowledge spillovers, and is proportional to $\lambda_e \times U_s'(z) \times h(z) \Delta z \Delta \gamma$.

Externality (II): Innovation Shifting However, the rise in the innovation investments of firms z raises the strength of competition across all the other firms. Other firms then respond to this competition by cutting down on their own innovation investments. This is the *innovation shifting* effect that we briefly discussed in the Introduction and earlier in this section. This effect is always negative since the innovation investments are decreasing in the rate of profit erosion, $\partial \gamma / \partial g < 0$, and the spillover function is monotonically increasing $U_s'(z) > 0$. Due to the higher market shares of productive firms, this negative effect rises with efficiency z.

Externality (III): Stifling Effect on Startups Competition also affects the incentives for new firms to enter the market. The faster rate of growth changes the expected value of entry and therefore potential entrants respond by reducing the rate of entry. This results in a reduction in spillovers proportional to $\mathbb{E}_{F_s}[U_s(z)] - \Psi_s$, where F_s denotes the distribution fo entrant efficiency.

Externality (IV): Cleansing Effect on Unproductive Firms Finally, competition raises the rate of exit among the least productive firms. This stronger selection lowers the negative crowding costs on the industry, proportional to Ψ_s. This is a positive effect.

The net external effect of the rise in the innovation investment of firms z depends on all these components. However, the industry response is not limited to these; they are only the immediate impacts of the perturbation that clarify the channels. These effects in turn trigger further innovation and exit and entry responses among other firms. In the long-run the stationary distribution H also responds to the perturbation. The combination of the

responses of other firms and the stationary distribution may in fact give rise to a situation where the aggregate rate of productivity growth falls as the result of this perturbation.

Despite the complex and interconnected nature of the competition effects, we can still numerically compute the optimal allocation and compare it with the market equilibrium one. The next section discusses a scheme that allows us to solve for the optimal allocation.

3.3.3 Computing the Socially Optimal Allocation

To solve for the optimal allocation, I use a computational scheme that parallels the one used for solving the market equilibrium. However, the key difference here is that the industry and incumbent problems are fully entangled through the term involving the spillover function in Equation (3.29). Therefore, in order to solve for the value function, investment, and exit policy functions of firms, we need to know the stationary distribution of firm efficiencies.

I use an iterative scheme to solve the two problems jointly. Starting with an initial guess for the rate of profit erosion, I solve for the value function of the firm and the stationary distribution of firm efficiencies, under the market equilibrium. Next, I use these results to form a guess about the spillover function and solve for the social value function from Equation (3.29). Then, I continue to iteratively solve the social HJB and KFE until convergence. I use the difference between the resulting cost of entry and the actual cost of entry to update my guess about the optimal rate of profit erosion and repeat the same procedure. The final solution satisfies all the different conditions in Proposition 9.

3.3.4 Tax/Subsidy Implementation of Optimal Policies

In this section, I discuss potential implementations of the optimal allocation under the assumption of information symmetry between the government and private firms. Since the government can observe the productivity ranking of firms (efficiency z), we can reach the first-best through two different potential schemes. First, the government can simply offer an efficiency-dependent reward schedule $u_s(\cdot)$ and a lump-sum tax of ψ_s to firms. This scheme directly implements the socially optimal allocation and does not require any further

incentives for entry. We can think of this scheme as an *ex post* reward that corrects for the gap between social and private returns to innovation investments. Needless to say, this scheme is fairly costly, and involves large transfers between the government and private businesses.

Alternatively, the government can use a combination of targeted taxes/subsidies to implement the optimal allocations characterized in Section 3.3.1. In particular, the following set of instruments together achieve the optimal allocation: a flat tax on the operation of firms, a subsidy for firm startup costs, and two additional efficiency-dependent instruments, a subsidy on innovation investments and a tax on the labor inputs of incumbents. The lump-sum tax internalizes the negative crowding externality of firms and the entry subsidy internalizes the entrants' positive spillovers. The efficiency-dependent input taxes and subsidies together internalize innovation externalities while maintaining the static efficiency of allocations across firms.

To see how this works, consider a firm with efficiency z that invests the amount $\gamma_s(z)$ under the optimal allocation. The government offers tax $\tau_l(z)$ on production labor inputs and subsidy $\tau_i(z)$ to the innovation labor inputs of the firm. The purpose of the production input tax is simply to ensure the static efficiency of firm outputs: therefore the government chooses

$$\tau_l(z) = -\tau_i(z)\ \varphi(\gamma_s(z)), \tag{3.35}$$

to ensure that the total labor costs of the firm under the implementation scheme match those under the optimal allocation. In addition, because of this tax the government does not need to transfer a net amount to the firm.

Now, if we let τ_o denote the tax on the operation of firms, the value function of the firm under the implementation satisfies the following HJB equation

$$r\widehat{V}_s(z) = \max_{\gamma} e^z J_\tau(\gamma; \tau_i(z), z) - (1+\tau_o) + \rho(\gamma - g_s)\widehat{V}_s'(z) + \frac{1}{2}\sigma^2\widehat{V}_s''(z)$$

where I have defined the function

$$J_\tau(\gamma;\tau_i,z) \equiv \left(1+(1+\tau_i)\,\varphi(\gamma)-\tau_i\varphi(\gamma_s(z))\right)^{-\rho}.$$

It is easy to show that the derivative of this function with respect to γ, when evaluated at $\gamma_s(z)$, is given by $(1+\tau_i)\,J'(\gamma_s(z))$. Together with condition (3.34), this implies that if we set

$$\tau_i(z) = \frac{\widehat{V}_s'(z)}{V_s'(z)} - 1,$$

where the social value function V_s satisfies the social HJB (3.29), the choice of innovation of firm z under the implementation coincides with its choice under the optimal policy.

The following lemma then characterizes the set of instruments that implements the optimal allocation.

Lemma 5. *(Tax/Subsidy Implementation) The social planner can achieve the optimal allocation by imposing a flat tax τ_o on the operation of firms, a tax/subsidy $\tau_{e,s}$ on the costs of entry, and a size-dependent tax/subsidy $\tau_{i,s}(\cdot)$ on incumbent innovation investments. Let $\widehat{V}_s(\cdot)$ denote the value function of the firm under the market equilibrium featuring the additional taxes and subsidies. This value function and the operation tax τ_o satisfy the following differential equation*

$$r\,\widehat{V}_s(z) = e^z J(\gamma_s(z)) - (1+\tau_o) + \rho\left(\gamma_s(z)-g_s\right)\widehat{V}_s'(z) + \frac{\rho^2}{2}\sigma^2\,\widehat{V}_s''(z), \tag{3.36}$$

subject to the boundary conditions $\widehat{V}_s'(z_{o,s}) = \widehat{V}_s(z_{o,s}) = 0$, where $(\gamma_s(\cdot), z_{o,s})$ stand for the firm policies under the optimal allocation from Proposition 9. Given the value function $\widehat{V}_s(\cdot)$ and the social value function V_s , satisfying social HJB (3.29), the incumbent and entrant innovation taxes/subsidies that implement the optimal allocation are given by

$$\tau_{i,s}(z) = \frac{\widehat{V}_s'(z)}{V_s'(z)} - 1, \tag{3.37}$$

$$\tau_{e,s} = \frac{\mathbb{E}_{F_s}\left[\widehat{V}_s'(z)\right]}{\mathbb{E}_{F_s}\left[V_s(z)\right]} - 1. \tag{3.38}$$

Note that Equation (3.36) has a different form from the social planner's HJB equation (3.29). As a result, the social value function of the firms V_s and their private value under

the subsidy/tax implementation $\widehat{V}_s$ do not necessarily coincide. Therefore, the subsidies in Equation (3.37) and (3.38) will in general be nonzero.

3.4 Quantitative Results

In this section, I use data on the life cycle dynamics of firms to calibrate the parameters of the model and to quantitatively investigate its positive and normative predictions. The model is parsimonious and characterizes each sector of the economy by three key sets of structural characteristics relating to 1) process of innovation (volatility, and costs of innovation and entry), 2) demand (elasticity of substitution and the rate of growth of demand), and 2) the process of entry and adoption (directedness and the degree of decreasing returns).

I face a challenge in calibrating the parameters of the process of entry, since reliable measures of the productivity of entrants are rarely available. Employments, sales, or even estimates of multifactor productivity do not provide precise measures for the performance or efficiency of new firms. In the early stages of their life cycles, firms invest more heavily in tangible and intangible capital relative to mature firms and therefore the comparison between the two groups may not be justifiable. As I will explain below, I rely on the relationship between the rate of exit and the age of the firm to identify the quality of entering cohorts relative to other firms active in the industry.

3.4.1 Data

I use the Business Dynamics Statistics (BDS) tables, a dataset that has been specifically developed to account and track sources of job creation and destruction through entry and exit, disaggregated by firm size, age, and industry.[29] The data starts in 1977, but the truncation of the firm age variable is heavy for the first few years. I focus on the period 1987-2007 to allow for a broader coverage of the age variable and avoid the arguably different

[29]The underlying data for LBD and BDS is at the establishment level, but the Census Bureau matches establishments based on the commonality of ownership or control to create firm-level data (see Haltiwanger *et al.*, 2009). The dataset can be accessed online at http://www.census.gov/ces/dataproducts/bds/.

firm dynamics during the Great Recession.

As I will explain below, the empirical exercise further requires the rates of growth in the aggregate employment and labor productivity of firms in each sector. BDS data already provides information on the total employment of firms. For measures of sectoral productivity growth, I rely on the Bureau of Labor Statistics (BLS) National Accounts to obtain estimates of the average rate of labor productivity growth in each sector for the period under study.[30]

3.4.2 Model Specification

The model in Section 3.2 characterizes a single-sector general equilibrium economy, but I perform the calibration at the level of two broad sectors of the US economy: manufacturing (SIC code 20) and retail trade (SIC code 52). I sketch a two-sector generalization of the model in Section C.2.3 of the Appendix, and show that the model is compatible with an equilibrium path along which each sector grows in terms of average productivity and mass of firms with a sector-specific rate. This provides a theoretical ground for separately calibrating the model to these two sectors of the US economy.[31]

I assume the following form for the costs of innovation

$$\varphi(\gamma) = \chi \frac{\gamma^{1+\kappa}}{1+\kappa}, \tag{3.39}$$

defined in the interval $\gamma \in [0, \overline{\gamma}]$, where

$$\overline{\gamma} \equiv \left(\frac{\kappa}{\chi} \frac{1+\kappa}{\rho(1+\kappa)+1} \right)^{\frac{1}{1+\kappa}}. \tag{3.40}$$

[30] BLS estimates sectoral labor productivity based on measures of value added and hours worked provided in the NIPA accounts (for the details of the methods used, see Bureau of Labor Statistics, 2008). The dataset can be accessed online at https://www.bls.gov/lpc/.

[31] There is evidence in the data that the sectors in fact have been undergoing some restructuring and therefore may not be precisely evolving along their long-run stationary equilibrium paths. However, the focus of this work is in documenting the essential inter-sector differences in the mechanics of growth. Hence, I abstract away from characterizing the transitional dynamics, and assume that the two sectors observed in the data evolve along their respective stationary balanced growth paths and match the data moments with those implied by the model.

For the process of diffusion, I consider the adoption distribution (3.11) that allows for a potential bias of the distribution toward more productive firms. In addition, I assume the relation (3.13) between entry and adopted productivities, and let $\tilde{\theta}_t$ denote the avearge productivity of the industry at time t. Then, the tuple (μ, η, γ_e) characterizes the parameters of the adoption process.

3.4.3 Calibration Strategy

To calibrate the main parameters of the model, I find the values that match the predictions of the model with the data on the sectoral rates of growth and the life-cycle dynamics of firm employment.

For the parameters of demand, I assume a discount rate $r = 0.05$ and an elasticity of substitution of $\rho = 2$ for both sectors, both values common in the literature.[32] As I show in Section C.2.3 of the Appendix, the model allows me to separately choose the growth rates g_N, corresponding to both rates of growth in the number of firms and in total employment in the two sectors. I choose to match this value to the rate of growth of total sector employment based on the BDS data.

The remaining parameters of the model are (1) the primitives of the innovation process, i.e., relative costs of entry ψ, size and curvature of the innovation cost function (χ, κ), and volatility σ, and (2) the primitives of the diffusion process, i.e., the directedness μ, the decreasing returns parameter η and the shifter γ_e. I choose to search in the space of 5 parameters $(\chi, \sigma, \mu, \eta, \gamma_e)$, and to directly use two moments from the data and prior work to pin down the two parameters (ψ, κ).[33]

To find the two parameters (ψ, κ), I employ the following scheme. Given other model parameters, I find the curvature of the innovation cost function κ to match the values

[32] The choice of $\rho = 2$ implies an elasticity of substitution of 3, close to the average of the estimates that Broda and Weinstein (2006) find for the elasticity of substitution based on US import demand data. I choose this number, following previous calibration exercises that rely on CES industry-level aggregators to study within-industry firm heterogeneity (e.g., Hsieh and Klenow, 2009; Perla *et al.*, 2015).

[33] The fixed cost of operation ψ_f does not play a role in determining the dynamics of the model in the long run.

reported in the empirical literature for the user cost elasticity of firm-level R&D spending.[34] The literature typically relies on variations in the corporate tax code over time (Hall, 1993; Hall and Van Reenen, 2000), across countries (Bloom *et al.*, 2002), and across US states (Wilson, 2009) as sources of changes in user costs of R&D to estimate the response in the R&D expenditure of firms. I follow Akcigit and Kerr (2012) and Acemoglu *et al.* (2013) who conclude that an elasticity of around -1 well summarizes the findings of this literature. Therefore, I set κ such that the corresponding value in the model is -1. Finally, the zero entry condition in Equation (3.23) suggests that given other model parameters, the rate of productivity growth g_θ pins down the relative costs of entry ψ in each sector.

I search for the set of parameters $(\chi, \sigma, \mu, \eta, \gamma_e)$ that minimize the distance between the moments observed in the data and the ones predicted from the model (for previous examples of a similar approach, see Lentz and Mortensen, 2008; Acemoglu *et al.*, 2013). I target the following three sets of moments:

1. Rate of entry,
2. Rate of employment growth and reallocation for old (15-20 year) firms,
3. Rates of exit for mature (10-15 year) and old (15-20 year) firms.

The first three moments (1 and 2) help identify the costs of innovation χ , the extent of volatility σ, and the directedness of the adoption process μ. The third group of moments together provide us with information about the relationship between age and the rate of exit in the industry, which informs the identification of the decreasing returns parameter η. If entrants begin with low-quality ideas relative to incumbents, they exit heavily in their early stages of life since they are very likely to cross the exit threshold early on due to negative shocks. If entrants begin with high-quality ideas relative to incumbents, their rates of exit has a more gradual relationship with age, since enough negative shocks have to accumulate in order for them to cross the exit threshold. This allows us to infer the initial characteristics

[34] In Section C.2.3 of the Appendix, I derive the relation between the parameter κ and the observed user cost elasticity of innovation investments based on the model. Using this relation, the latter directly pins down the value of the former, given other parameters of demand and innovation environment, and the rate of sectoral productivity growth.

of firms based on their likelihood of exit as a function of age.

Let $\mathbf{p}$ denote the vector of model parameters and $\widehat{\mathbf{m}}$ a vector of targeted empirical moments observed in the data. Let $\mathbf{m}(\mathbf{p})$ denote the vector of the expected values for the same moments based on the model, when calibrated with the set of parameters $\mathbf{p}$. To find the set of parameters that best explain the data, I use the following criterion

$$\mathbf{p}^* \equiv \min_{\mathbf{p}:\, \mathbf{p}_{fix}=\mathbf{p}^*_{fix}} \|\mathbf{m}(\mathbf{p}) - \widehat{\mathbf{m}}\|_{\widehat{\Sigma}}, \tag{3.41}$$

where $\mathbf{p}_{fix}$ is the subset of parameters fixed to $\mathbf{p}^*_{fix}$ (see above), and $\widehat{\mathbf{\Sigma}}$ is a positive definite matrix computed based on the variance of the moments in the data over time. We cannot derive analytical expressions for moment functions $\mathbf{m}(\mathbf{p})$ under the model. Hence, I use simulation to compute the predictions of the model for each moment.[35] Finally, I employ a simple brute force grid search to minimize the objective function of the problem (3.41). I search in a subset of the parameter space that yields a feasible solution and find the parameters with the lowest value for the objective function (3.41).

3.4.4 Calibration Results

Table 3.1 presents the calibrated moments for the two sectors. The top three moments are each directly used in the calibration of one specific parameter. Therefore, the model and the data exactly match with regard to these moments. The case of the rate of employment growth in manufacturing poses an exception, since the data suggests a negative rate,but the model requires a positive rate of growth of employment to ensure the existence of a stationary long-run equilibrium.[36] Therefore, I instead assign a very small positive value for

[35]The logic of the approach is the same as that of the Method of Simulated Moments (MSM) for estimation (McFadden, 1989; Pakes and Pollard, 1989; Duffie and Singleton, 1993; Gourieroux and Monfort, 1993). The key challenge in computing the moments implied by the model stems from the continuous-time nature of the underlying stochastic process. A wide literature in empirical finance has studied the problem of estimating the parameters of continuous-time stochastic differential equations such as those used in this paper using simulation-based/analytical and ML/GMM-based methods (Pedersen, 1995; Aït-Sahalia, 2002; Durham and Gallant, 2002; Brandt and Santa-Clara, 2002). Here, I use a simple and fast simulation-based approach that relies on time discretization and importance sampling (Pedersen, 1995).

[36]The leading explanation for a negative trend in the data is that the sector is moving along an adjustment path toward a new stationary equilibrium with a smaller manufacturing sector. This could be due to the changes

g_N in manufacturing (equal to1×10^{-4}) to deliver the existence of a long-run equilibrium. The remaining five moments are used for the grid search according to the objective function in Equation (3.41).

Table 3.1: *Targeted moments and the calibrated parameters for retail trade and manufacturing. The first three targeted moments each directly pin down a corresponding parameter in the model: the rate of productivity growth pins down the relative cost of entry ψ, the rate of employment growth pins down the rate of growth in the number of firms g_M, and the cost elasticity pins down the curvature parameter of innovation costs κ. The 5 moments of the firm life-cycle dynamics together calibrate the five remaining parameters of the model, characterizing the volatility of firm outcomes σ, the costs of innovation χ, and the process of entry and adoption (μ, η, γ_e).*

Targeted Moment		**Retail Trade**		**Manufacturing**	
		Model	**Data**	**Model**	**Data**
Rate of productivity growth	g_θ	0.032	0.032	0.038	0.038
Rate of emp growth	g_N	0.017	0.017	0.000	-0.012
Cost elasticity of investment		1.0	1.0	1.0	1.0
Rate of entry		0.117	0.119	0.079	0.079
Rate of emp growth of old (15-20 yr)		-0.015	-0.007	-0.024	-0.017
Rate of reallocation of old (15-20 yr)		0.26	0.23	0.20	0.20
Rate of exit of mature (10-15 yr)		0.075	0.083	0.060	0.060
Rate of exit of old (15-20 yr)		0.065	0.071	0.054	0.052
Parameter					
Volatility	σ	0.19		0.15	
Innovation costs	χ	13360		8103	
Curvature of costs	κ	2.2		2.3	
Entry costs	ψ	6.2		11.3	
Directedness of adoption	μ	3		1	
Decreasing returns of adoption	η	0.6		0.3	
Entry Productivity Shifter	γ_e	-0.5		-0.5	

in the technological environment (e.g., skill-biased technical change) or the effects of globalization (rise of the Japanese manufacturers in the 70s and 80s and Chinese manufacturers in the 2000s).

The calibrated parameters show some differences between the two sectors. As one would expect, the relative costs of innovation are higher in retail compared to manufacturing. The volatility parameter is higher in retail due, among other things, to the higher rate of reallocation of employment for old firms. This is in line with the facts documented by previous empirical work (Foster *et al.*, 2006). The parameters of the process of entry and adoption (μ, η) also present clear contrasts between the two sectors. The decreasing returns parameter appears higher in retail trade. The calibrated parameters suggest that firms enter the market with an efficiency (log productivity) that is only 40 percent that of the firms they are trying to imitate (70 percent in manufacturing), and have to build up the remainder of their competency by investing in innovation through their life cycle.[37] Whereas the adoption seems fairly directed toward high-productivity firms in retail, uniform adoption provides the best fit in manufacturing. In other words, entrants are more likely to adopt the ideas of the most productive incumbents in retail. Together these calibrated parameters suggest that the adoption process may be more directed but less perfect in retail compared to manufacturing.[38]

Targeted Moments The model provides a good fit for the targeted moments, and in particular those that capture the relationship between age and the likelihood of exit. In the cases of the rates of growth and reallocation of old firms, the fit is not as close. The reason

[37]The model infers this mainly based on the fact that the rate of exit of mature and old firms relative to overall the rate of turnover, as proxied by the rate of entry, is higher in manufacturing than in retail. One way to interpret this is that the retail sector is better characterized by a so-called *revolving-door* feature, in which most entrants soon exit and rarely grow to penetrate into the small band of industry leaders. Audretsch (1995) introduces this concept and contrasts it with what he calls a *forest* industry where leading incumbents are frequently taken over by new entrants that slowly grow from beneath. Seen from the prism of the current model, the differences between the two evolutionary modes reflect the differences in the relative position of entrant ideas, relative to the incumbents they are trying to imitate.

[38]This finding is compatible with the view that an important component of the competency of a firm in retail trade corresponds to the types of intangible capital such as building supplier networks, branding, organizational capacity that the firm has to acquire throughout its life-cycle, and cannot simply imitate from others. For instance, a more detailed structural account of the differences between retail and manufacturing should take into account the process of market penetration and spatial expansion, which inevitably requires an extended catch-up process for entrants (for the example of the growth of Walmart, see Holmes, 2011).

is that the data in fact suggests a strong trend in the relationship between age and volatility, which the current simple model is abstracting from.[39] This means that the calibration has to find a compromise in fitting the patterns for old versus young firms. Nevertheless, the calibration preserves the ordering of the moments between retail and manufacturing for all the moments.

Growth Decomposition and Untargeted Moments Table 3.2 presents a number of different moments predicted by the model under the calibrated parameters for the two sectors (in the columns under the heading "Before").

Table 3.2: *The impact of a drastic fall in the costs of innovation (parameter χ) to 75% of its initial value on industry outcomes. For each sector, the leftmost column, under the heading "Before," shows the value of the variable under the calibrated parameters, the middle column, under the heading "After," shows the value after the change. The third column then presents the difference between the two values. IDR stands for Inter-Decile Range.*

Variable	Retail Trade			Manufacturing		
	Before	**After**	**Change**	**Before**	**After**	**Change**
Rate of Productivity Growth (%)	3.22	3.51	0.28	3.76	4.13	0.37
... contribution of Turnover	0.774	0.772	-0.002	0.538	0.547	0.009
... contribution of Net Entry	-0.776	-0.775	0.001	-0.005	-0.005	0.000
Rate of Entry	0.118	0.119	0.001	0.079	0.081	0.002
Entry/Exit Productivity Gap	0.519	0.524	0.005	0.795	0.798	0.004
Concentration (Pareto Index)	1.167	1.161	-0.006	1.199	1.196	-0.003
Dispersion of Investments (IDR)	0.377	0.395	0.018	0.361	0.380	0.019

The first three rows show the decomposition of Equation (3.26), separately showing the contributions of turnover (replacement of exiting firms with entrants) and net entry. First,

[39] Interestingly, the data suggests that the gap in the volatility of young and old firms is higher in manufacturing relative to retail, even though on average the volatility is lower in the former.

note that, relative to the overall sectoral rate of productivity growth, the contribution of turnover is considerably larger in retail compared to manufacturing. This is the case despite the fact that, as the fourth line in the table shows, the gap in the productivity of entering and exiting firms is smaller in retail compared to manufacturing. The faster rate of turnover in retail compensates for this smaller gap and results in a stronger contribution of selection to growth in this sector. However, the third row indicates another difference between the two sectors in the contribution of entry: faster growth in the number of retail firms means that the sector is constantly flooded by small and relatively unproductive entrants. The negative contribution of this component more than cancels out the contribution of selection.[40]

The fixth row shows that the Pareto tail indices in both sectors are very close to one, and therefore suggest the approximate emergence of the Zipf law (Gabaix, 2016). This contrasts with the workhorse models of innovation and growth (Klette and Kortum, 2004; Lentz and Mortensen, 2008) that predict counterfactually thin tails for the distribution of firm size (see Appendix C.1 for further discussion). Relatedly, the stationary distribution of efficiencies in manufacturing exhibits a larger dispersion of productivity compared to retail. The interdecile range (IDR) of log productivity is 2.6 in manufacturing and 2.0 in retail. This is compatible with the findings of Foster *et al.* (2006), who highlight the larger dispersion of productivity in manufacturing despite the higher volatility of businesses in retail.

Table 3.2 also compares the relative dispersion of innovation investments between the two sectors. As the measure of this dispersion, I compute the interdecile range (IDR) of the investments, normalized by the maximum level of investments in each sector. The interdecile range of the investments in retail is predicted to be slightly larger than that in manufacturing. The model makes this prediction based on the higher volatility of firm outcomes in retail

[40]The closest empirical counterparts to these results are the productivity decompositions reported by Foster *et al.* (2001) and Foster *et al.* (2006) in the two sectors. They find a far more pronounced difference in the contribution of entry and exit to growth in between the two sectors, with entry and exit explaining almost all of the growth in retail in the 90s while accounting for about 25% of growth in manufacturing within the same period. They also find a larger gap between the productivity of entering and exiting establishments in retail compared to manufacturing. However, note that Foster *et al.* (2006) focus mainly on entry and exit of establishment rather than firms. It is likely that their results are driven by the entry of establishments that belong to high-productivity incumbent firms, fast growing retail chains such as Walmart and Target, rather than new firms.

compared to manufacturing. In a volatile environment, highly productive firms have a stronger advantage in process innovation because they expect to reap the benefits of their investments over a longer horizon. High productivity shields these firms from the threat of imminent exit, which undermines the innovation incentives of unproductive firms in a volatile industry.

Counterfactual Experiment: Reduction in the Costs of Innovation Next, I use the model to explore the consequences of a large reduction in the costs of innovation in the two sectors. I study what happens when due to forces external to the sector, for instance, the emergence of novel General Purpose Technologies (GPTs), new opportunities for innovation become available to all firms. I study this question by examining the changes in the long-run equilibrium when the parameter of innovation costs χ falls to 75 percent of its original level in both sectors. I assume all other parameters, including the costs of entry, remain the same before and after this change.

Table 3.2 shows a number of different outcomes before and after the reduction in costs. The long-run rate of productivity growth in both sectors rises by about 0.3 percent in response to this change. The lion's share of this response is provided by the rising investments of incumbent firms as we find a very small response in the contribution of turnover. Interestingly, the impact on the contribution of turnover to growth is in reverse direction between the two sectors. The rate of entry mildly rises in response to lower costs in both sectors. This response is mainly driven by the responsiveness of the exit decisions of unproductive firms to the higher innovation investments of productive firms. Finally, the overall impact of these changes raises the concentration in both industries, as reflected in the reduction in the tail index of the long-run distribution. The sectors become more concentrated, since more productive firms disproportionately benefit from the lower innovation costs, while the costs of entry have remained intact. This is compatible with the evidence on the rise in the concentration of the retail sector (Foster *et al.*, 2006).

3.4.5 Optimal Allocations

Let us now examine the optimal allocation of innovation investments and production in the two sectors. These allocations solve the social planner's HJB equation (3.29), the corresponding KFE, and the social free entry condition (3.33).

Figure 3.3 compares the market equilibrium and socially optimal innovation investments in the two sectors. Figure 3.4 performs the same comparison, but for the long-run distribution of productivity in the two sectors. Finally, Figure 3.5 shows the difference between the market equilibrium and the socially optimal allocation for the two variables.

Figure 3.3: *Comparison between the optimal and market equilibrium allocations of innovation in (Right) manufacturing and (Left) retail, as a function of the relative efficiency of firms. RET and MAN stand for retail and manufacturing, respectively.*

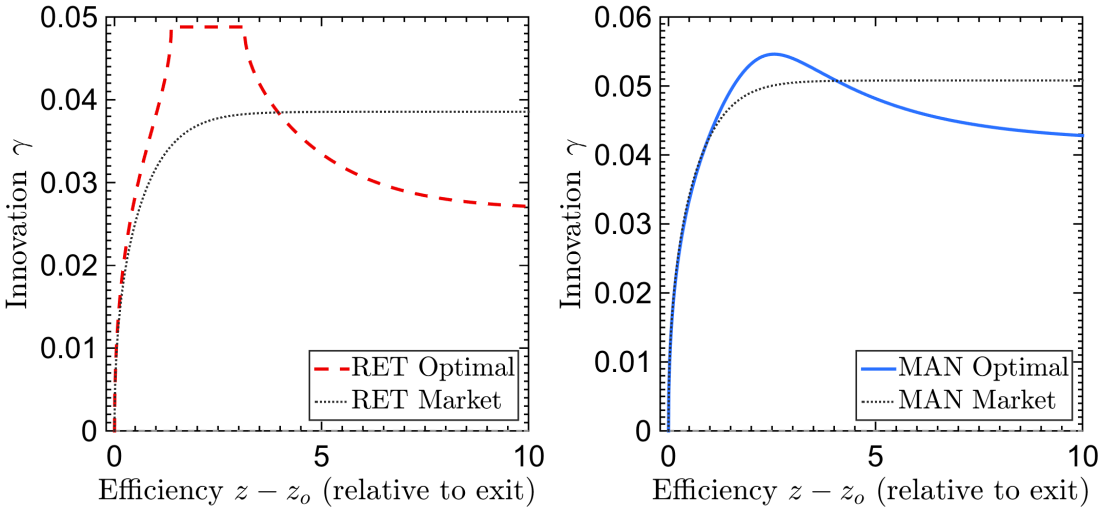
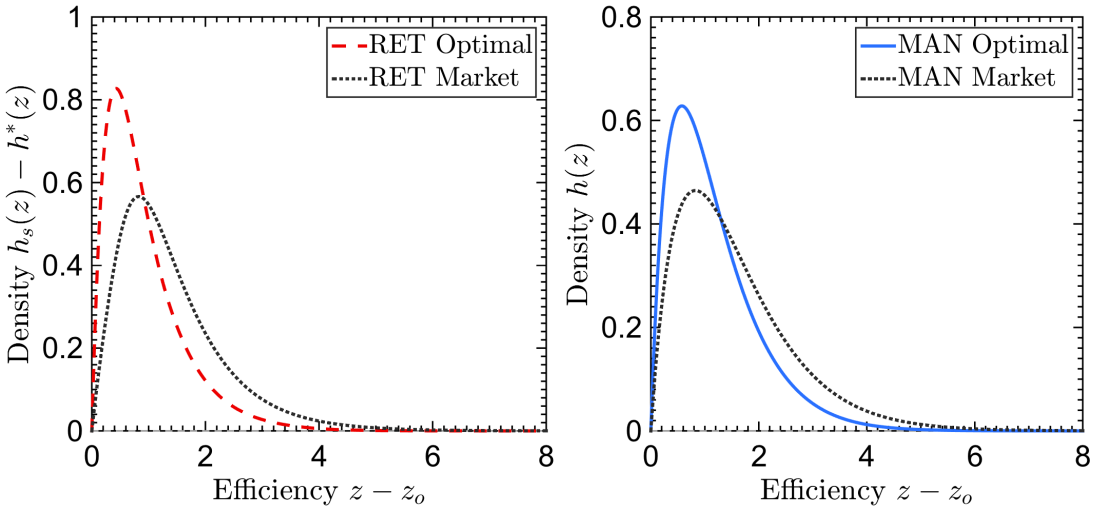
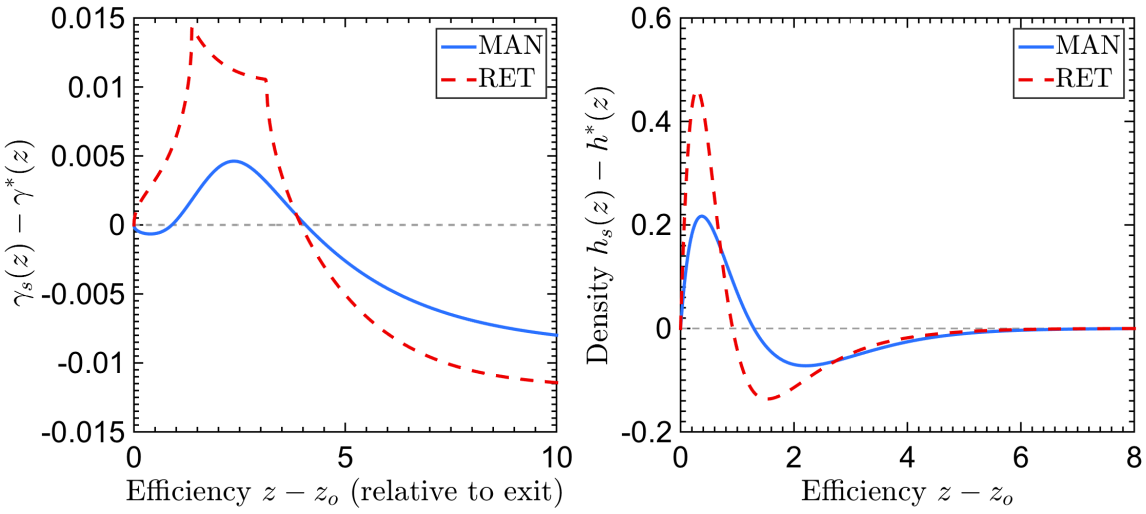


Figure 3.4: *Comparison between the optimal and market equilibrium long-run distributions of efficiency in (Right) manufacturing and (Left) retail. RET and MAN stand for retail and manufacturing, respectively.*



First, observe that in both sectors, large and productive firms overinvest in innovation. Relative to the innovation intensity of each sector, the extent of misallocation in the innovation investments appears somewhat larger in retail compared to manufacturing. Examining the differences in innovation investments Figure 3.5, we find that underinvestment of firms in the medium range of productivity is particularly severe in retail trade. For a range of medium efficiency firms in retail, the socially optimal innovation investments reaches the maximum value given by Equation (3.40) and is constant. This suggests a strong negative competition effect stemming from the innovation investments of high-efficiency firms in this sector.

Figure 3.5: *The difference between the optimal and market equilibrium allocations in (Right) innovation investments and (Left) long-run probability distribution function, as a function of the relative efficiency of firms. RET and MAN stand for retail and manufacturing, respectively.*



Under the socially optimal allocation, the long-run distributions of efficiency in both sectors shift to the left. This means that the concentration of both sectors under the market equilibrium is suboptimally high. The socially optimal allocation raises the rate of productivity growth and therefore the rate of dynamism of the industry and the obsolescence of ideas. As a result, it is harder for firms to grow very large under the optimal allocation.

Finally, note that Figure 3.5 presents the results as a function of efficiency relative to the cutofff under the corresponding allocation. The socially optimal exit cutoffs are larger than those of the market equilibrium because the social planner internalizes the negative crowding externalities of firms on entrants.

Implementation of the Optimal Allocation I use the scheme introduced in Lemma 5 to compute the set of subsidies and taxes that the social planner can employ to implement the optimal policy under the market equilibrium. These Pigouvian taxes serve the purpose of showing us the wedges that exist under the market equilibrium between the social and private returns to innovation investments. Table 3.3 compares this policy between the two sectors. The average subsidy rate that the scheme offers to incumbents and entrants is

higher in retail trade compared to manufacturing. Figure 3.6 shows that the profiles of the schedule of subsidies as a function of firm efficiency also appear fairly different between the two sectors.

Figure 3.6: *(Left) Comparison between the subsidy scheme that achieves the optimal innovation investments in manufacturing and retail, as a function of the relative efficiency of firms. (Right) The same comparison but zooming in on the right tail of the productivity distribution. Here, the x-axis shows the percentile ranking of the productivity of firms at the top of the distribution of productivities. RET and MAN stand for retail and manufacturing, respectively.*

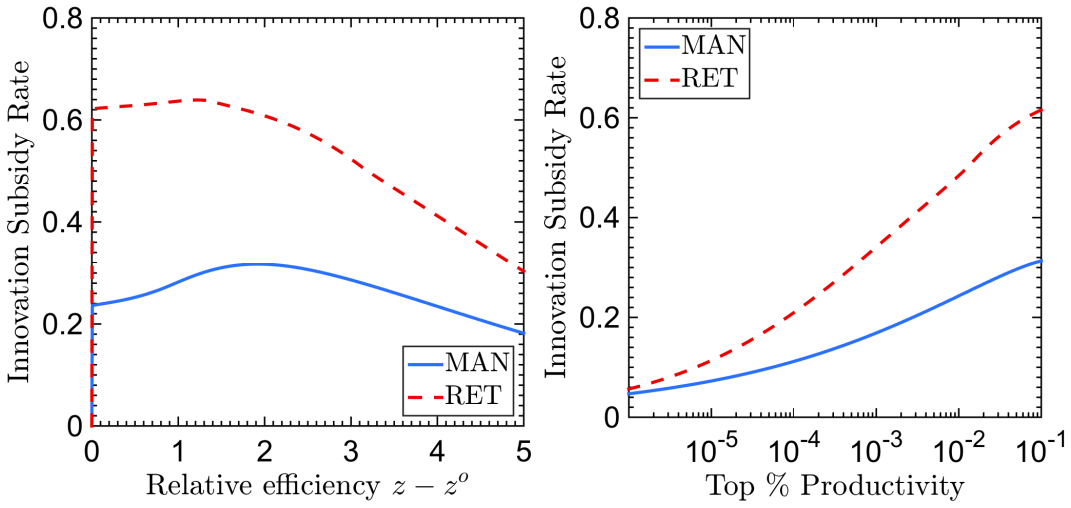
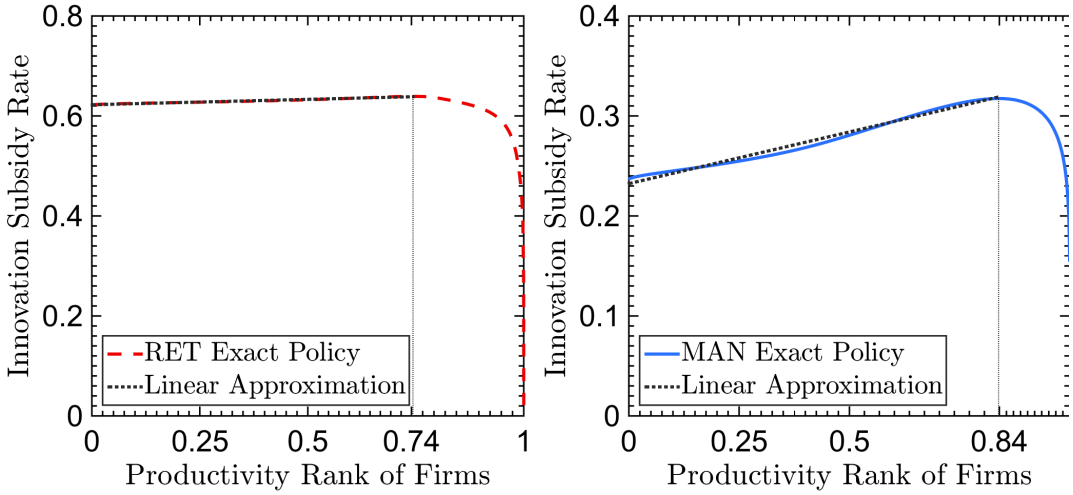


Figure 3.7: *Comparison between the exact and the linearly approximated optimal subsidy for (Left) retail and (Right) manufacturing. The gray vertical line indicates the rank corresponding to the maximum of the subsidy rate.*



First, note that the relationship between efficiency and the subsidy rate is nonmonotonic. The subsidy is increasing with efficiency among the lower tiers of productivity, but then declines among high-efficiency firms. However, the initial rise is much less pronounced in retail compared to manufacturing. Figure 3.7 shows that this increasing relationship can be well approximated with a linearly growing subsidy rate for a vast majority of firms (for firms up to the 74 percentile of productivity and size in retail, and 84 in manufacturing). In Table 3.3 (row 7), we can compare the slopes of the approximate schedule of subsidy rates between the two sectors. Among lower-productivity firms, the relationship between the subsidy rate and productivity is far weaker in retail compared to manufacturing. In contrast, among high-efficiency firms, the relation between productivity and the optimal subsidy rate is stronger in retail: the subsidy declines faster among high-effciency firms in this sector. This point relates to the observation we made earlier: the negative competition effect of high-productivity firms is relatively larger in retail compared to manufacturing.

Table 3.3: *The characteristics of the optimal allocations of innovation and production, as well as a subsidy scheme that implements them, in retail trade and manufacturing. See Section 3.4.5 for details.*

Variable		**Retail Trade**	**Manufacturing**
Optimal Rate of Growth	$g_{\theta,s}$	0.052	0.047
Average Incumbent Subsidy	$-\mathbb{E}_H[\tau_{i,s}(z)]$	0.62	0.28
Entrant Subsidy	$-\tau_{e,s}$	0.45	0.18
Operation Tax	$\tau_{o,s}$	4.90	2.95
Baseline Subsidy Rate		0.62	0.23
Slope of the Subsidy Rate		0.02	0.10
Optimal Schedule Spending		0.32	0.20
Best Uniform Spending		2.57	1.96

Comparison with Uniform Incumbent Subsidies How much do we gain by implementing the nonlinear subsidy scheme above, relative to a uniform one-size-fits-all subsidy rate to the innovation investments of all incumbents? To investigate this question, I take the following strategy. I compare the amount of spending by the government on subsidies to entrants, under the subsidy schedule that implements the optimal allocation and under a uniform subsidy scheme that minimizes spending. As with the optimal subsidy, I assume that the government in the case of uniform subsidies also imposes a tax on the wages of production workers, given by Equation (3.35), to cancel out the effect of subsidies on the production decisions of firms. As a result, the net spending of the government on subsidies is given by the amount paid to entrants.

For each constant rate of uniform subsidies, I find the corresponding subsidy rate to entrants that achieves the same rate of productivity growth as that of the optimal allocation. To control for the effect of lump-sum operation taxes, I perform the comparison with and without the additional lump-sum tax for the case of the uniform subsidies.[41]

The last two rows of Table 3.3 show considerable gaps between the spending on subsidies between the scheme that optimally accounts for heterogeneity and one that ignores it. I conclude that the model implies misallocations in the innovation investments of firms that are first-order to the design of policies that aim to internalize the externalities of innovation.

3.4.6 Effect of Static Competition

As we saw, the model implies sizable misallocations in innovation investments across heterogeneous firms. How does the strength of product market rivarly, as proxied by the degree of substitutability between firm products, affect these misallocations? To investigate this point, I perform the following exercise. I consider the effect of changing the elasticity of substitution parameter from $\rho = 2$ to $\rho_{new} = 3$. In each sector, I adjust the parameters of the costs of innovation (χ, κ) and entry ψ such that the cost elasticity of innovation, the rate

[41] This scheme effectively solves the constrained optimum subsidy problem, where the imposed constraint is the constancy of the subsidy rates for incumbent firms.

of growth of large firms, given by $\rho\left(\gamma^{**} - g_\theta + \frac{1}{2}(\rho-1)\sigma^2\right)$, and the rate of productivity growth remain the same. However, I keep the volatility parameter σ and the parameters of the process of knowledge diffusion (η, γ, γ_e) intact, so that the analysis only focuses on the effect of static product market rivarly.

Figure 3.9 shows the difference between the market and socially optimal investments in innovation for the two sets of parameters with the low and high substitutability. Figure 3.8 compares the optimal subsidy schemes between the two product market competition settings. Based on the discussions above, we should expect that stronger competition amplifies the negative externalities of large firms on smaller ones. As a result, with stronger competition, the peak of the gap in the market and optimal investments, and the subsidy schemes in both sectors shifts to the left.

Figure 3.8: *Comparison of the difference between the optimal and market equilibrium allocations with two different degrees of substitutability in firm outputs in (Right) retail and (Left) manufacturing, as a function of the relative efficiency of firms. RET and MAN stand for retail and manufacturing, respectively.*

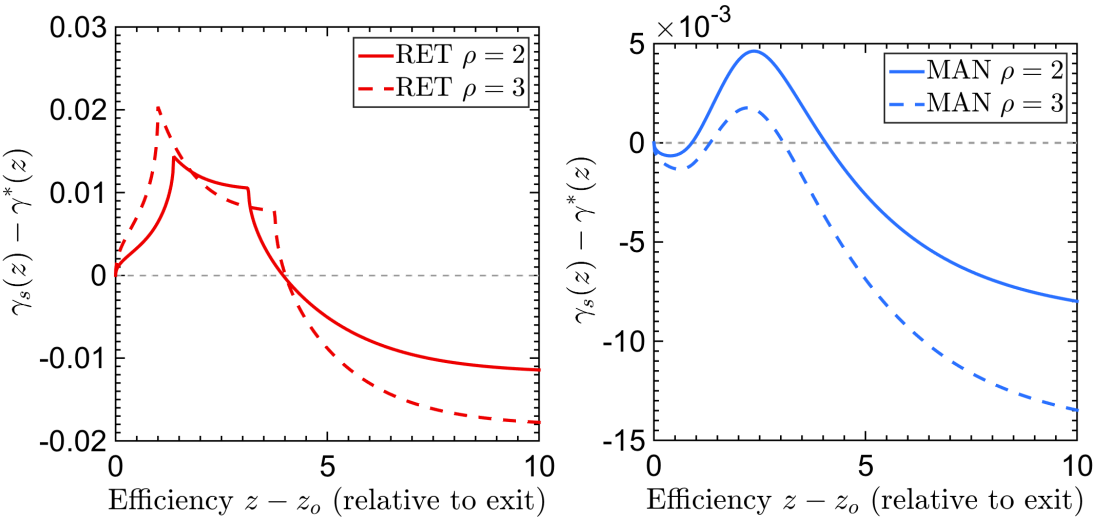
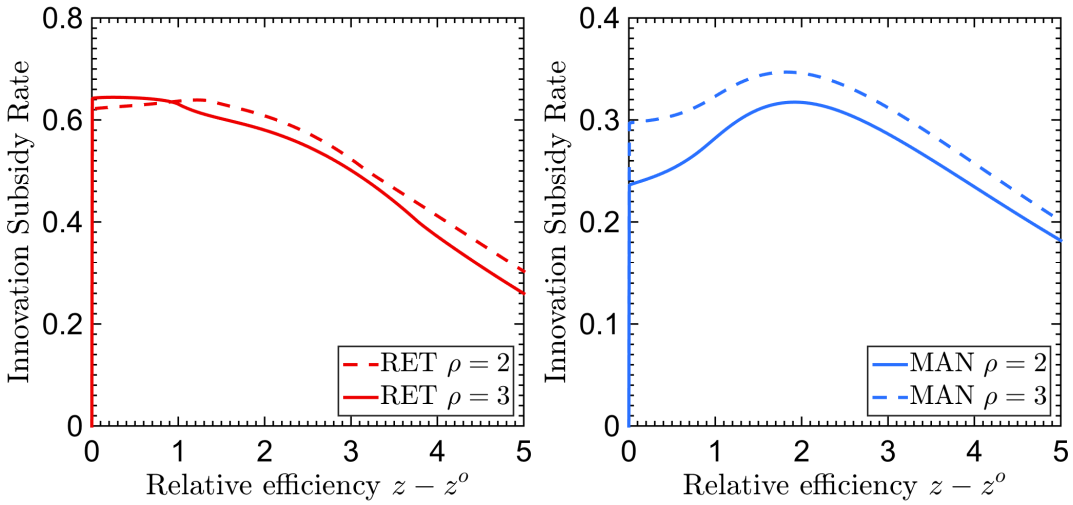


Figure 3.9: *Comparison between the subsidy scheme that achieves the optimal innovation investments with two different degrees of substitutability in firm outputs in (Left) retail and (Right) manufacturing, as a function of the relative efficiency of firms. RET and MAN stand for retail and manufacturing, respectively.*



3.5 Discussion and Future Work

Knowledge spillovers are the cornerstone of our modern theories of growth. In this paper, I study them in a standard model of firm heterogeneity and selection (Hopenhayn, 1992; Melitz, 2003) where firms actively invest in innovation (Ericson and Pakes, 1995; Atkeson and Burstein, 2010). I find that, under a fairly broad set of assumptions, markets misallocate innovation investment across firms. In a calibration based on data from two broad sectors of the US economy, I find that these misallocations are quantitatively large and removing them could result in a substantial rise in the aggregate rate of productivity growth.

The model offered here only focuses on the dynamic effects of competition that work through the expected lifetimes of firms. Specifically, the assumption of CES demand implies constant markups that do not respond to the innovation investments or changes in firm productivity. More importantly, under this assumption, the extent to which firms cannot fully appropriate the social value of innovation exactly equates their business stealing effect on other firms. However, extensions of this model can accommodate more general demand systems such as those with Variable Elasticities of Substitution (VES) in which

both markups and the gap between appropriability and business stealing effects vary across firms (Zhelobodko *et al.*, 2012; Dhingra and Morrow, 2012) .

In this paper, I mainly characterize the behavior of industry equilibria along the stationary constant growth paths. Therefore, I do not examine the regularities observed along the industry life cycle, which has been the focus of a separate line of theoretical and empirical work. Klepper and Graddy (1990) and Klepper and Simons (2000) document the fact that as a new industry emerges, the number of firms rises, only to fall again after a so-called shake-out phase. Among others, Jovanovic and MacDonald (1994b) and Klepper (1996) offer theoretical frameworks that explain these patterns. I note that some of these patterns may also be rationalized through the model offered in this paper along the transitional dynamics of the path of the industry. I leave the investigation of this point for future research.

References

(2013). Innovation pessimism: Has the ideas machine broken down? *The Economist*.

Acemoglu, D. (2002). Directed Technical Change. *Review of Economic Studies*, **69** (4), 781–809.

— (2007). Equilibrium bias of technology. *Econometrica*, **75** (5), 1371–1409.

— (2008). *Introduction to Modern Economic Growth*. Princeton University Press.

— (2009). *Introduction to Modern Economic Growth*. Princeton, NJ: Princeton University Press.

—, Akcigit, U., Bloom, N. and Kerr, W. (2013). Innovation, Reallocation and Growth.

—, — and Celik, M. A. (2014). Young , Restless and Creative: Openness to Disruption and Creative Innovations.

— and Cao, D. (2015). Innovation by Entrants and Incumbents. *Journal of Economic Theory*, **157**, 255–294.

— and Guerrieri, V. (2008). Capital Deepening and Nonbalanced Economic Growth. *Journal of Political Economy*, **116** (3), 467–498.

— and Linn, J. (2004). Market size in innovation: theory and evidence from the pharmaceutical industry. *Quarterly Journal of Economics*, **119** (3), 1049–1090.

Aghion, P. and Howitt, P. (1992). A model of growth through creative destruction.

Aguiar, M. and Bils, M. (2015a). Has Consumption Inequality Mirrored Income Inequality? *American Economic Review*, **105** (9), 2725–2756.

— and — (2015b). Has Consumption Inequality Mirrored Income Inequality. *American Economic Review*, **105** (9), 2725–56.

Aït-Sahalia, Y. (2002). Maximum Likelihood Estimation of Discretely Sampled Diffusions: A Closed-form Approximation Approach. *Econometrica*, **70** (1), 223–262.

Aizcorbe, A. (2009). Toward better measurement of innovation and intangibles. *Survey of Current Business*, (January), 10–24.

Akcigit, U. and Kerr, W. R. (2012). Growth through Heterogeneous Innovations.

Alvarez-Cuadrado, F. and Poschke, M. (2011). Structural Change Out of Agriculture: Labor Push versus Labor Pull. *American Economic Journal: Macroeconomics*, **3** (3), 127–58.

Aoyama, H., Yoshikawa, H., Iyetomi, H. and Fujiwara, Y. (2010). Productivity dispersion: Facts, theory, and implications. *Journal of Economic Interaction and Coordination*, **5** (1), 27–54.

Arkolakis, C. (2016). A Unified theory of firm selection and growth. *Quarterly Journal of Economics*.

Atkeson, A. and Burstein, A. T. (2010). Innovation, Firm Dynamics , and International Trade. *Journal of Political Economy*, **118** (3), 433–484.

— and — (2015). Aggregate implications of innovation policy.

Audretsch, D. B. (1995). *Innovation and Industry Evolution*. MIT Press.

Aw, B. Y., Roberts, M. J. and Xu, D. Y. (2011). R&D Investment, Exporting, and Productivity Dynamics. *American Economic Review*, **101** (June), 1312–1344.

Axtell, R. L. (2001). Zipf distribution of U.S. firm sizes. *Science*, **293** (5536), 1818–1820.

Bartelsman, E. and Haltiwanger, J. (2009). *Measuring and Analyzing Cross-Country Differences in Firm Dynamics*. January.

—, — and Scarpetta, S. (2013). Cross-Country Differences in Productivity: The Role of Allocation and Selection. *American Economic Review*, **103** (1), 305–334.

Basker, E. (2007). The causes and consequences of Wal-Mart's growth. *Journal of Economic Perspectivess*, **21** (3), 177–198.

— (2016). The evolution of technology in the retail sector. In *Handbook of the Economics of Retailing and Distribution*, Edward Elgar Publishing, pp. 38–53.

Baumol, W. J. (1967a). Macroeconomics of Unbalanced Growth: The Anatomy of Urban Crisis. *American Economic Review*, **57** (3), 415–426.

— (1967b). Macroeconomics of unbalanced growth: the anatomy of urban crisis. *American Economic Review*, **57** (3), 415–426.

Benhabib, J., Tonetti, C. and Perla, J. (2014). Reconciling Models of Diffusion and Innovation: A Theory of the Productivity Distribution and Technology Frontier.

Blackorby, C. and Russell, R. R. (1981). The morishima elasticity of substitution; symmetry, constancy, separability, and its relationship to the hicks and allen elasticities. *The Review of Economic Studies*, pp. 147–158.

Bloom, N., Griffith, R. and Van Reenen, J. (2002). Do R&D tax credits work? Evidence from a panel of countries 1979-1997. *Journal of Public Economics*, **85** (1), 1–31.

— and Reenen, J. V. (2010). Why do management practices differ across firms and countries? *Journal of Economic Perspectives*, **24** (1), 203–224.

Boppart, T. (2014a). Structural Change and the Kaldor Facts in a Growth Model With Relative Price Effects and Non-Gorman Preferences. *Econometrica*, **82**, 2167–2196.

— (2014b). Structural change and the Kaldor facts in a growth model with relative price effects and non-Gorman preferences. *Econometrica*, **82** (6), 2167–2196.

— and Weiss, F. J. (2013). Non-homothetic preferences and industry directed technical change. (June), 1–45.

Brandt, M. W. and Santa-Clara, P. (2002). Simulated likelihood estimation of diffusions with an application to exchange rate dynamics in incomplete markets. *Journal of Financial Economics*, **63** (2), 161–210.

Broda, C. and Weinstein, D. E. (2006). Globalization and the Gains from Variety. *Quarterly Journal of Economics*, **121** (2), 541–585.

Budish, E., Roin, B. N. and Williams, H. (2015). Do firms underinvest in long-term research? Evidence from cancer clinical trials. *The American economic review*, **105** (7), 2044–2085.

Buera, F. J. and Kaboski, J. P. (2009). Can Traditional Theories of Structural Change Fit The Data? *Journal of the European Economic Association*, **7** (2-3), 469–477.

— and — (2012a). Scale and the origins of structural change. *Journal of Economic Theory*, **147** (2), 684–712.

— and — (2012b). The Rise of the Service Economy. *American Economic Review*, **102** (6), 2540–69.

Bureau of Labor Statistics (2008). *Technical Information About the BLS Major Sector Productivity and Costs Measures*. Tech. rep.

Caron, J., Fally, T. and Markusen, J. R. (2014). Internationl trade puzzles: a solution linking production and preferences. *Quarterly Journal of Economics*, **129** (3), 1501–1552.

Carvalho, V. M. and Grassi, B. (2015). Large Firm Dynamics and the Business Cycle. pp. 1–64.

Cerda, R. A. (2007). Endogenous innovations in the pharmaceutical industry. *Journal of Evolutionary Economics*, **17** (4), 473–515.

Chaney, T. (2008). Distorted gravity: The intensive and extensive margins of international trade. *American Economic Review*, **98** (4), 1707–1721.

Clark, G. (2016). Winter is coming: Robert gordon and the future of economic growth. *American Economic Review*, **106** (5), 68–71.

Cohen, W. M. (2010). Fifty years of empirical studies of innovative activity and performance. In *Handbook of the Economics of Innovation*, vol. 1, 1st edn., Elsevier B.V., pp. 129–213.

— and Klepper, S. (1996). A Reprise of Size and R&D. *Economic Journal*, **106** (437), 925–951.

—, Levin, R. C. and Mowery, D. C. (1987). Firm size and R&D intensity: a re-examination. *Journal of Industrial Economics*, **34** (2), 217–223.

Comin, D., Lashkari, D. and Mestieri, M. (2015). Structural change with long-run income and price effects.

Corrado, C. A. and Hulten, C. R. (2010). How do you measure a "technological revolution"? *American Economic Review*, **100** (2), 99–104.

—, — and Sichel, D. (2009). Intangible Capital and U.S. Economic Growth. *Review of Income and Wealth*, **55** (3), 661–685.

Davis, S. J., Haltiwanger, J. and Schuh, S. (1996). Small business and job creation: Dissecting the myth and reassessing the facts. *Small Business Economics*, **8** (4), 297–315.

De Loecker, J. (2013). Detecting learning by exporting. *American Economic Journal: Microeconomics*, **5** (3), 1–21.

Deaton, A. and Muellbauer, J. (1980). An almost ideal demand system. *American Economic Review*, **70** (3), 312–26.

Decker, R., Haltiwanger, J., Jarmin, R. and Miranda, J. (2014). The Role of Entrepreneurship in US Job Creation and Economic Dynamism. *Journal of Economic Perspectives*, **28** (3), 3–24.

Dennis, B. N. and Iscan, T. B. (2009). Engel versus Baumol: Accounting for structural change using two centuries of U.S. data. *Explorations in Economic History*, **46** (2), 186–202.

Dhingra, S. and Morrow, J. (2012). The Impact of Integration on Productivity and Welfare Distortions Under Monopolistic Competition.

Diewert, W. E. (1976). Exact and superlative index numbers. *Journal of Econometrics*, **4** (2), 115–145.

Disney, R., Haskel, J., Heden, Y. and Haskelt, J. (2003). Entry, Exit and Establishment Survival in UK Manufacturing. *the Journal of Industrial Economics*, **51** (1), 91–112.

Dixit, A. (1991). Analytical Models in Approximations of Hysteresis. *Review of Economic Studies*, **58** (1), 141–151.

Dixit, A. K. and Pindyck, R. S. (1994). *Investment under Uncertainty*. Princeton, NJ: Princeton University Press.

Doms, M. E., Jarmin, R. S. and Klimek, S. D. (2004). Information technology investment and firm performance in US retail trade. *Economics of Innovation and New Technology*, **13** (7), 595–613.

Doraszelski, U. and Jaumandreu, J. (2013). R&D and Productivity: Estimating Endogenous Productivity. *The Review of Economic Studies*, **80** (4), 1338–1383.

Duarte, M. and Restuccia, D. (2010). The Role of the Structural Transformation in Aggregate Productivity. *The Quarterly Journal of Economics*, **125** (1), 129–173.

Duffie, D. and Singleton, K. J. (1993). Simulated Moments Estimation of Markov Models of Asset Prices. *Econometrica*, **61** (4), 929–952.

Dunne, T., Roberts, M. J., Samuelson, L. and Dunne, T. (1988). Patterns of firm entry and exit in U . S . manufacturing industries. *RAND Journal of Economics*, **19** (4), 495–515.

Durham, G. and Gallant, A. (2002). *Numerical techniques for maximum likelihood estimation of continuous-time diffusion processes*, vol. 20.

Echevarria, C. (1997). Changes in Sectoral Composition Associated with Economic Growth. *International Economic Review*, **38** (2), 431–52.

Ericson, R. and Pakes, A. (1995). Markov-Perfect Industry Dynamics: Framework for Work Empirical. *Review of Economic Studies*, **62** (1), 53–82.

Evans, D. S. (1987). The Relationship Between Firm Growth, Size, and Age: Estimates for 100 Manufacturing Industries. *Journal of Industrial Economics*, **35** (4), 567–581.

Foellmi, R., Wuergler, T. and Zweimüller, J. (2014). The macroeconomics of Model T. *Journal of Economic Theory*, **153** (C), 617–647.

— and Zweimüller, J. (2006). Income Distribution and Demand-Induced Innovations. *Review of Economic Studies*, **73** (4), 941–960.

— and — (2008). Structural change, Engel's consumption cycles and Kaldor's facts of economic growth. *Journal of Monetary Economics*, **55** (7), 1317–1328.

Foster, L., Haltiwanger, J. and Krizan, C. J. (2001). *Aggregate Productivity Growth: Lessons from Microeconomic Evidence*. January.

—, — and — (2006). Market Selection, Reallocation, and Restructuring in the U.S. Retail Trade Sector in the 1990s. *Review of Economics and Statistics*, **88** (4), 748–758.

—, — and Syverson, C. (2008). Reallocation, Firm Turnover, and Efficiency : Selection on Productivity or Profitability? *American Economic Review*, **98** (1), 394–425.

—, Haltiwanger, J. C., Klimek, S. D., Krizan, C. J. and Ohlmacher, S. (2016). The evolution of national retail chains: how we got here. In E. Basker (ed.), *Handbook of the Economics of Retailing and Distribution*, Edward Elgar Publishing.

Gabaix, X. (2011). The Granular Origins of Aggregate Fluctuations. *Econometrica*, **79** (3), 733–772.

— (2016). Power Laws in Economics: An Introduction. *Journal of Economic Perspectives*, **30** (1), 185–206.

Garcia-Macia, D., Hsieh and Klenow, P. J. (2015). How Destructive is Innovation?

Geerolf, F. (2014). A Static and Microfounded Theory of Zipf' s Law for Firms and of the Top Labor Income Distribution.

Gibrat, R. (1931). *Les inégalités économiques*. Paris: Recueil Sirey.

Gollin, D., Parente, S. and Rogerson, R. (2002). The Role of Agriculture in Development. *American Economic Review*, **92** (2), 160–164.

Gordon, R. J. (2016). *The rise and fall of American growth: The US standard of living since the civil war*. Princeton University Press.

Gorman, W. (1965). Production functions in which the elasticities of substitution stand in fixed proportions to each other. *Review of Economic Studies*, **32**, 217–224.

Gourieroux, C. and Monfort, A. (1993). Simulation-based inference. A survey with special reference to panel data models. *Journal of Econometrics*, **59** (1-2), 5–33.

Griliches, Z. (1979). Issues in assessing the contribution of research and development to productivity growth. *The Bell Journal of Economics*, **10** (1), 92–116.

— (2000). *R&D, Education, and Productivity: A Retrospective*. Harvard University Press.

— and Schmookler, J. (1963). Inventing and maximizing. *American Economic Review*, **57** (3), 415–426.

Grossman, G. M. and Helpman, E. (1990). Trade, innovation, and growth. *American Economic Review*, **151** (3712), 867–8.

— and — (1991a). *Innovation and Growth in the Global Economy*. Massachusetts Intitute of Technology.

— and — (1991b). Quality ladders in the theory of growth. *Review of Economic Studies*, **58** (1), 43—-61.

Hall, B. and Van Reenen, J. (2000). How effective are fiscal incentives for R&D? A review of the evidence. *Research Policy*, **29** (4-5), 449–469.

Hall, B. H. (1987). The relationship between firm size and firm growth in the US manufacturing sector. *Journal of Industrial Economics*, **34** (2), 217–223.

— (1993). R&D tax policy during the 1980s: success or failure? In J. Poterba (ed.), *Tax Policy and the Economy, Vol 7*, vol. 7, MIT Press, pp. 1–36.

— (1996). The private and social returns to research and development. In B. L. R. Smith and C. E. Barfield (eds.), *Technology, R&D, and the Economy*, Brookings Institute.

Haltiwanger, J. (2014). Job Creation, Job Destruction, and Productivity Growth: The Role of Young Businesses. *Annual Review of Economics*, **7** (1), 150504160918000.

—, Jarmin, R. and Miranda, J. (2009). *Business Dynamics Statistics: An Overview*. Tech. Rep. January.

Haltiwanger, J. C., Jarmin, R. S. and Miranda, J. (2013). Who creates jobs? Small vs. large vs. young. *Review of Economics and Statistics*, **95** (2), 347–361.

Hamilton, J. D. (1994). *Time series analysis*. Princenton University Press.

Hanoch, G. (1975a). Production and demand models with direct or indirect implicit additivity. *Econometrica*, **43** (3), 395–419.

— (1975b). Production and Demand Models with Direct or Indirect Implicit Additivity. *Econometrica*, **43** (3), 395–419.

Hart, P. E. and Prais, S. J. (1956). The analysis of business concentration: a statistical approach. *Journal of the Royal Statistical Society. Series A (General)*, **53** (2), 365–375.

Heathcote, J., Perri, F. and Violante, G. L. (2010). Unequal We Stand: An Empirical Analysis of Economic Inequality in the United States: 1967-2006. *Review of Economic Dynamics*, **13** (1), 15–51.

Helpman, E., Melitz, M. J. and Yeaple, S. R. (2004). Export versus FDI with Heterogeneous Firms. *American Economic Review*, **94** (1), 300–316.

Hermalin, B. E. (2012). Leadership and corporate culture. In R. Gibbons and J. Roberts (eds.), *Handbook of Organizational Economics*, Princeton University Press, pp. 432—-478.

Herrendorf, B., Rogerson, R. and Valentinyi, A. (2013). Two Perspectives on Preferences and Structural Transformation. *American Economic Review*, **103** (7), 2752–89.

—, — and — (2014). Growth and Structural Transformation. In *Handbook of Economic Growth*, *Handbook of Economic Growth*, vol. 2, *6*, Elsevier, pp. 855–941.

Holmes, T. J. (2011). The Diffusion of Wal-Mart and Economies of Density. *Econometrica*, **79** (1), 253–302.

Hopenhayn, H. A. (1992). Entry, Exit, and firm Dynamics in Long Run Equilibrium. *Econometrica*, **60** (5), 1127–1150.

Hsieh, C.-T. and Klenow, P. J. (2009). Misallocation and manufacturing TFP in China and India. *Quarterly Journal of Economics*, **124** (4), 14031–448.

Huber, P., Oberhofer, H. and Pfaffermayr, M. (2013). Who Creates Jobs ? Estimating Job Creation Rates at the Firm Level. *WIFO working paper*.

Jaravel, X. (2017). The unequal gains from product innovations: evidence from the US retail sector.

Johnson, D. S., Parker, J. A. and Souleles, N. S. (2006). Household Expenditure and the Income Tax Rebates of 2001. *American Economic Review*, **96** (5), 1589–1610.

Jorgenson, D. W. and Timmer, M. P. (2011). Structural Change in Advanced Nations: A New Set of Stylised Facts. *Scandinavian Journal of Economics*, **113** (1), 1–29.

Jovanovic, B. (1982). Selection and the Evolution of Industry. *Econometrica*, **50** (3), 649–670.

— and MacDonald, G. M. (1994a). Competitive diffusion. *Journal of Political Economy*, **102** (3), 547–565.

— and — (1994b). The Life Cycle of a Competitive Industry. *Journal of Political Economy*, **102** (2), 322–347.

Klepper, S. (1996). Entry, exit, growth, and innovation over the product life cycle. *American Economic Review*, **86** (3), 562–583.

— and Graddy, E. (1990). The evolution of new industries and the determinants of market structure. *The RAND Journal of Economics*, **21** (1), 27–44.

— and Simons, K. L. (2000). The Making of an Oligopoly: Firm Survival and Technological Change in the Evolution of the U.S. Tire Industry. *Journal of Political Economy*, **108** (4), 728–760.

Klette, T. J. and Kortum, S. (2004). Innovating Firms and Aggregate Innovation. *Journal of Political Economy*, **112** (5), 986–1018.

Kongsamut, P., Rebelo, S. and Xie, D. (2001). Beyond Balanced Growth. *Review of Economic Studies*, **68** (4), 869–82.

Krueger, D. and Perri, F. (2006). Does Income Inequality Lead to Consumption Inequality? Evidence and Theory. *Review of Economic Studies*, **73** (1), 163–193.

Kuznets, S. (1973). Modern Economic Growth: Findings and Reflections. *American Economic Review*, **63** (3), 247–58.

Lashkari, D. (2011). *In search of functional specificity in the brain: generative models for group fMRI data*. Ph.D. thesis, Massachusetts Institute of Technology.

Lentz, R. and Mortensen, D. T. (2008). An Empirical Model of Growth Through Product Innovation. *Econometrica*, **76** (6), 1317–1373.

— and — (2014). Optimal Growth Through Product Innovation.

Lucas, R. E. and Moll, B. (2014). Knowledge growth and the allocation of time. *Journal of Political Economy*, **122** (1).

Luttmer, E. G. (2007). Selection, growth, and the size distribution of firms. *Quarterly Journal of Economics*, **122** (3), 1103–1144.

— (2010). Models of Growth and Firm Heterogeneity. *Annual Review of Economics*, **2** (1), 547–576.

Luttmer, E. G. J. (2011). On the mechanics of firm growth. *Review of Economic Studies*, **78** (3), 1042–1068.

— (2012). Technology diffusion and growth. *Journal of Economic Theory*, **147** (2), 602–622.

Maddison, A. (1980). Economic Growth and Structural Change in Advanced Countries. In *Western Economies in Transition*, Croom Helm, pp. 41–60.

Matsuyama, K. (1992). Agricultural Productivity, Comparative Advantage, and Economic Growth. *Journal of Economic Theory*, **58** (2), 317–334.

— (2015). The Home Market Effect and the Patterns of Trade between Rich and Poor Countries, Unpublished Manuscript.

McFadden, D. (1989). A Method of Simulated Moments for Estimation of Discrete Response Models Without Numerical Integration. *Econometrica*, **57** (5), 995–1026.

McKinsey Global Institute (2001). *U.S. Productivity Growth 1995-2000*. Tech. rep., McKinsey Global Institute, Washington, DC.

Melitz, M. J. (2003). The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity. *Econometrica*, **71** (6), 1695–1725.

Mizuno, T., Ishikawa, A., Fujimoto, S. and Watanabe, T. (2012). Power laws in firm productivity. *Progress of Theoretical Physics Supplement*, (194), 122–134.

Neumark, D., Wall, B. and Zhang, J. (2011). Do Small Businesses Create More Jobs? New Evidence for the United States from the National Establishment Time Series. *Review of Economics and Statistics*, **93** (1), 16–29.

Ngai, L. R. and Pissarides, C. A. (2007a). Structural Change in a Multisector Model of Growth. *American Economic Review*, **97** (1), 429–443.

— and — (2007b). Structural Model of Growth Change in a Multisector. *American Economic Review*, **97** (1), 429–443.

OECD (2010). *R&D tax incentives: rationale, design, evaluation*. Tech. rep., Organization for Economic Cooperation and Development Innovation Platform.

— (2011). International Experience with R&D Tax Incentives.

— (2012). *OECD Science, Technology and Industry Outlook 2012*. Tech. rep.

Olken, B. (2010). Discussion of The African Growth Miracle, mimeo.

Olley, G. S. and Pakes, A. (1996). The Dynamics of Productivity in the Telecommunications Equipment Industry. *Econometrica*, **64** (6), 1263–1297.

Pakes, A. and Pollard, D. (1989). Simulation and the asymptotics of optimization estimators. *Econometrica*, **57** (5), 1027–1057.

— and Schankerman, M. (1984). An exploration into the determinants of research intensity. In Z. Griliches (ed.), *R&D, Patents, and Productivity*, vol. I, University of Chicago Press, pp. 209–232.

Pedersen, A. R. (1995). A New Approach to Maximum Likelihood Estimation for Stochastic Differential Equations Based on Discrete Observations. *Scandinavian Journal of Statistics*, **22** (1), 55–71.

Perla, J. and Tonetti, C. (2014). Equilibrium Imitation and Growth. *Journal of Political Economy*, **122** (1), 1–21.

—, — and Waugh, M. E. (2015). Equilibrium Technology Diffusion, Trade, and Growth.

Peters, M. (2011). Heterogeneous Mark-Ups and Endogenous Misallocation.

Prescott, E. C. and Visscher, M. (1980). Organization Capital. *Journal of Political Economy*, **88** (3), 446–461.

Restuccia, D. and Rogerson, R. (2008). Policy distortions and aggregate productivity with heterogeneous establishments. *Review of Economic Dynamics*, **11** (4), 707–720.

Romer, P. M. (1986). Increasing Returns and Long-Run Growth. *Journal of Political Economy*, **94** (5), 1002–1037.

— (1990). Endogenous Technological Change. *Journal of Political Economy*, **98** (S5), S71.

Sampson, T. (2016). Dynamic Selection: An Idea Flows Theory of Entry, Trade and Growth. *Quarterly Journal of Economics*, (January).

Sato, R. (1975a). The most general class of ces functions. *Econometrica*, **43** (5-6), 999–1003.

— (1975b). The Most General Class of CES Functions. *Econometrica*, **43** (5), 999–1003.

Scherer, F. M. (1982). Demand-pull and technological invention: Schmookler revisited. *Journal of Industrial Economics*, **34** (2), 225–237.

Schmookler, J. (1966). *Invention and Economic Growth*. Harvard University Press.

Simon, H. A. and Bonini, C. P. (1958). The size distribution of business firms. *American Economic Review*, **48** (1), 1–17.

Snow, A. and Warren, R. S. (2015). Pigou's law and the proportionality of income and price elasticities of demand. *Economics Letters*, **132** (C), 136–138.

Stockey, N. L., Lucas, R. E. and Prescott, E. C. (1989). *Recursive Methods in Economic Dynamics*. Harvard University Press.

Święcki, T. (2014). Determinants of Structural Change, mimeo.

Syverson, C. (2011). What Determines Productivity? *Journal of Economic Literature*, **49** (2), 326–365.

Vries, K. d., Vries, G. d. and Timmer, M. (2014). *Patterns of Structural Change in Developing Countries*. GGDC Research Memorandum GD-149, Groningen Growth and Development Centre, University of Groningen.

Wilson, D. J. (2009). Beggar Thy Neighbor? The In-State, Out-of-State, and Aggregate Effects of R&D Tax Credits. *Review of Economics and Statistics*, **91** (May), 431–436.

Wooldridge, J. M. (2001). *Econometric Analysis of Cross Section and Panel Data*, *MIT Press Books*, vol. 1. The MIT Press.

World Bank (2010). *Innovation policy: A guide for developing countries*. Tech. rep., World Bank Group.

Xu, D. Y. (2008). A Structural Empirical Model of R&D, Firm Heterogeneity, and Industry Evolution. *Working Paper*, (January), 1–40.

Young, A. (2012a). The african growth miracle. *Journal of Political Economy*, **120**, 696–739.

— (2012b). The African Growth Miracle. *Journal of Political Economy*, **120** (4), 696–739.

— (2013a). Inequality, the urban-rural gap, and migration. *Quarterly Journal of Economics*, **128**, 1727–1785.

— (2013b). Inequality, the urban-rural gap, and migration. *Quarterly Journal of Economics*, **128** (4), 1727–1785.

Zhelobodko, E., Kokovin, S., Parenti, M. and Thisse, J.-F. (2012). Monopolistic Competition: Beyond the Constant Elasticity of Substitution. *Econometrica*, **80** (6), 2765–2784.

Appendix A

Appendix to Chapter 1

A.1 Nonhomothetic CES Preferences

In this section of the appendix, we provide an overview of the properties of the general family of nonhomothetic Constant Elasticity of Substitution (CES) preferences. We first introduce the general family of nonhomothetic CES preferences in Section A.1.1. We then specialize them to the case of isoelastic nonhomothetic CES functions in Section A.1.2 and contrast the latter with Stone-Geary and PIGL preferences in Section A.1.3.

These preferences have a number of distinctive features, extensively studied in an early literature on theoretical foundations of preferences and production functions. Sato (1975a) derived a general family of CES functions as the solution to a partial differential equation that imposes the constancy of elasticity of substitution. This family includes standard homothetic CES functions as well as two classes of separable and non-separable nonhomothetic functions. Hanoch (1975a) showed that additivity of the direct or indirect utility (or production) function results in price and income effects that are non-trivially dependent on each other. He then introduced *implicit additivity* and derived in a family of functions where the income elasticity of demand is not fully dependent on the elasticity of substitution. Our nonhomothetic CES functions correspond to the non-separable class of functions in the sense of Sato (1975a), which also satisfy the condition of implicit additivity

in the sense of Hanoch (1975a).

Finally, Blackorby and Russell (1981) have proved an additional property that is unique to this class of functions. In general, different generalizations of the elasticity of substitution to cases involving more than two variables, e.g., the Allen-Uzawa definition or the Morishima definition, are distinct from each other. However, for the class of nonhomothetic CES functions they become identical and elasticity of substitution can be uniquely defined similar to the case of two-variable functions.

A.1.1 General Nonhomothetic CES Preferences

Consider preferences over a bundle $\boldsymbol{C} = \{C_1, C_2, \cdots, C_I\}$ of goods defined through an implicit utility function:

$$\sum_{i=1}^{I} \Omega_i^{\frac{1}{\sigma}} \left(\frac{C_i}{g_i(U)} \right)^{\frac{\sigma-1}{\sigma}} = 1, \tag{A.1}$$

where functions g_i's are differentiable in U and $\sigma \neq 1$ and $\sigma > 0$.[1] Standard CES preferences are a specific example of Equation (A.1) with $g_i(U) = U$ for all i's. These preferences were first introduced, seemingly independently, by Sato (1975a) and Hanoch (1975a) who each characterize different properties of these functions. Here, we state and briefly prove some of the relevant results to provide a self-contained exposition of our theory in this paper.

Lemma 6. *If $\sigma > 0$ and functions $g_i(\cdot)$ are positive and monotonically increasing for all i, the function $U(\boldsymbol{C})$ defined in Equation (A.1) is monotonically increasing and quasi-concave for all $\boldsymbol{C} \gg 0$.*

Proof. Establishing monotonicity is straightforward. To establish quasi-concavity, assume to the contrary that there exists two bundles of $\boldsymbol{C}'$ and $\boldsymbol{C}''$ and their corresponding utility values U' and U'', such that $U \equiv U(\alpha \boldsymbol{C}' + (1-\alpha)\boldsymbol{C}'')$ is strictly smaller than both U' and

[1] For the case of $\sigma = 1$, the preferences are simply defined according to $\sum_i \Omega_i \log\left(\frac{C_i}{g_i(U)}\right) = 1$.

U''. We then have for the case $\sigma \geq 1$

$$\begin{aligned} 1 &= \sum_i \Omega_i^{1/\sigma} \left(\alpha \frac{C_i'}{g_i(U)} + (1-\alpha) \frac{C_i''}{g_i(U)} \right)^{\frac{\sigma-1}{\sigma}}, \\ &> \sum_i \Omega_i^{1/\sigma} \left(\alpha \frac{C_i'}{g_i(U')} + (1-\alpha) \frac{C_i''}{g_i(U'')} \right)^{\frac{\sigma-1}{\sigma}}, \\ &\geq \alpha \sum_i \Omega_i^{1/\sigma} \left(\frac{C_i'}{g_i(U')} \right)^{\frac{\sigma-1}{\sigma}} + (1-\alpha) \sum_i \Omega_i^{1/\sigma} \left(\frac{C_i''}{g_i(U'')} \right)^{\frac{\sigma-1}{\sigma}}, \end{aligned}$$

where in the second inequality we have used monotonicity of the g_i's and in the third we have used Jensen's inequality and the assumption that $\infty > \sigma > 1$.[2] Since the last line equals 1 from the definition of the nonhomothetic CES functions valued at U' and U'', we arrive at a contradiction. For the case that $0 < \sigma < 1$, we can proceed analogously. In this case, the inequality signs are reversed in both lines and we also reach a contradiction. □

Henceforth, we assume the conditions in Lemma 6 are satisfied. The next lemma characterizes the demand for general nonhomothetic CES preferences and provides the solution to the expenditure minimization problem.

Lemma 7. *Consider any bundle of goods that maximizes the utility function defined in Equation* (A.1) *subject to the budget constraint* $\sum_i p_i C_i \leq E$. *For each good* i, *the real consumption satisfies:*

$$C_i = \Omega_i \left(\frac{p_i}{E} \right)^{-\sigma} g_i(U)^{1-\sigma}, \tag{A.2}$$

and the share in consumption expenditure satisfies:

$$\omega_i \equiv \frac{p_i C_i}{E} = \Omega_i^{\frac{1}{\sigma}} \left(\frac{C_i}{g_i(U)} \right)^{\frac{\sigma-1}{\sigma}} = \Omega_i \left[g_i(U) \left(\frac{p_i}{E} \right) \right]^{1-\sigma}. \tag{A.3}$$

Proof. Let λ and ρ denote the Lagrange multipliers on the budget constraint and constraint (A.1), respectively:

$$\mathcal{L} = U + \rho \left(1 - \sum_i \Omega_i^{\frac{1}{\sigma}} \left(\frac{C_i}{g_i(U)} \right)^{\frac{\sigma-1}{\sigma}} \right) + \lambda \left(E - \sum_i p_i C_i \right).$$

[2] For the case $\sigma = 1$, recall that we have defined a logarithmic function, which is also concave. For the case $\sigma = \infty$ the inequality becomes an equality as we have a linear case.

The FOCs with respect to C_i yields:

$$\rho \frac{1-\sigma}{\sigma} \frac{\omega_i}{C_i} = \lambda p_i, \tag{A.4}$$

where we have defined

$$\omega_i \equiv \Omega_i^{\frac{1}{\sigma}} \left(\frac{C_i}{g_i(U)} \right)^{\frac{\sigma-1}{\sigma}}. \tag{A.5}$$

Equation (A.4) shows that expenditure $p_i C_i$ on good i is proportional to ω_i. Since the latter sums to one from constraint (A.1), it follows that ω_i is the expenditure share of good i, and we have:

$$E = \sum_{i=1}^{I} p_i C_i \quad = \quad \frac{1-\sigma}{\sigma} \frac{\rho}{\lambda}.$$

We can now substitute the definition of ω_i from Equation (A.5) in expression (A.4) and use (A.6) to find (A.2) and (A.3). □

Lemma 7 implies the following relationship, defining the expenditure (and implicitly the indirect utility function) for general Nonhomothetic CES preferences:

$$E = \left[\sum_{i=1}^{I} \Omega_i \left(g_i(U)\, p_i \right)^{1-\sigma} \right]^{\frac{1}{1-\sigma}}. \tag{A.6}$$

The expenditure function is continuous in prices p_i's and U, and homogeneous of degree 1, increasing, and concave in prices. The elasticity of the expenditure function with respect to utility is:

$$\eta_E^U \equiv \frac{U \partial E}{E \partial U} = \sum_i \omega_i \, \eta_{g_i}^U = \overline{\eta_{g_i}^U}, \tag{A.7}$$

which ensures that the expenditure function is increasing in utility if all g_i's are monotonically increasing. It is straightforward to also show that the elasticity of the utility function (A.1) with respect to consumption of good i is also given by:

$$\eta_U^{C_i} \equiv \frac{C_i \partial U}{U \partial C_i} = \frac{\omega_i}{\overline{\eta_{g_i}^U}}, \tag{A.8}$$

where ω_i is the ratio defined in Equation (A.5).

Examining sectoral demand from Equation (A.2) along indifference curves, shows the

main properties of nonhomothetic CES preferences. As expected, the elasticity of substitution is constant:

$$\eta_{C_i/C_j}^{p_i/p_j} \equiv \frac{\partial \log\left(C_i/C_j\right)}{\partial \log\left(p_i/p_j\right)} = \sigma. \tag{A.9}$$

More interestingly, the elasticity of relative demand with respect to utility is in general different from unity:

$$\eta_{C_i/C_j}^{U} \equiv \frac{\partial \log\left(C_i/C_j\right)}{\partial \log U} = \frac{\partial \log\left(g_i/g_j\right)}{\partial \log U}. \tag{A.10}$$

Since utility has a monotonic relationship with real income (and hence expenditure), it then follows that the expenditure elasticity of demand for different goods are different. More specifically, we can use (A.7) to find the expenditure elasticity of demand:

$$\eta_{C_i}^{E} \equiv \frac{\partial \log C_i}{\partial \log E} = \sigma + (1-\sigma)\,\frac{\eta_{g_i}^{U}}{\eta_{g_i}^{U}}. \tag{A.11}$$

The intuition for the normalization in expression (A.11) is that the elasticity $\eta_E^{C_i}$ has to be invariant to all monotonic transformations of utility.

Preferences defined by Equation (A.1) belong to the general class of preferences with *Direct Implicit Additivity*. Hanoch (1975a) shows that the latter family of preferences have the nice property that is illustrated by Equations (A.9) and (A.10): the separability of the income and substitution elasticities of the Hicksian demand. This is in contrast to the stronger requirement of *Explicit Additivity* commonly assumed in nonhomothetic preferences, whereby the utility is explicitly defined as a function $U = F\left(\sum_i f_i(C_i)\right)$. In Section A.1.3 below, we will show examples of how substitution and income elasticities of Hicksian demand are *not* separable for preferences with explicitly additivity in direct utility, e.g., generalized Stone-Geary preferences (Kongsamut *et al.*, 2001), or indirect utility, e.g., PIGL preferences (Boppart, 2014a).

Finally, let us investigate the convexity of the expenditure function in terms of utility. First, we express the second derivative of the expenditure function in terms of elasticities,

$$\frac{\partial^2 E}{\partial U^2} = \frac{E}{U^2}\eta_E^U\left(\eta_E^U + \eta_{\eta_E^U}^U - 1\right), \tag{A.12}$$

where $\eta^U_{\eta^U_E}$ is the second order elasticity of expenditure with respect to utility. We can compute this second order elasticity as follows:

$$\begin{aligned}
\eta^U_{\eta^U_E} &= U\frac{\partial}{\partial U}\log\sum_i \eta_{g_i}\left(U\right)\left(g_i\left(U\right)p_i\right)^{1-\sigma} - \left(1-\sigma\right)\frac{\partial \log E}{\partial \log U}, \\
&= \frac{\sum_i \eta_{\eta_{g_i}}\cdot\eta_{g_i}\left(g_i\left(U\right)p_i\right)^{1-\sigma} + \left(1-\sigma\right)\sum_i \eta^2_{g_i}\left(g_i\left(U\right)p_i\right)^{1-\sigma}}{\sum_i \eta_{g_i}\left(g_i\left(U\right)p_i\right)^{1-\sigma}} - \left(1-\sigma\right)\overline{\eta_{g_i}}, \\
&= \overline{\eta_{g_i}}\left[\frac{\overline{\eta_{\eta_{g_i}}\cdot\eta_{g_i}}}{\left(\overline{\eta_{g_i}}\right)^2} + \left(1-\sigma\right)Var\left(\frac{\eta_{g_i}}{\overline{\eta_{g_i}}}\right)\right], \qquad \text{(A.13)}
\end{aligned}$$

where $\overline{X_i}$ and $Var\left(X_i\right)$ denote the expected value and variance of variable X_i across sectors with weights given by expenditure shares ω_i for prices $\boldsymbol{p}$ and utility U.

To make sense of (A.13), consider the choice of $g_i(U) \equiv g(U)^{\epsilon_i}$ for some monotonically increasing function $g(\cdot)$ (which corresponds to the aggregator introduced in Section **??** of the online appendix). We have that $\eta_{g_i} = \eta_g\,\varepsilon_i$ and $\eta_{\eta_{g_i}} = \eta_{\eta_g}$, implying:

$$\eta^U_{\eta^U_E} = \eta_g\overline{\varepsilon_i}\left[\eta_{\eta_g} + \left(1-\sigma\right)Var\left(\frac{\varepsilon_i}{\overline{\varepsilon_i}}\right)\right]. \qquad \text{(A.14)}$$

A.1.2 Isoelastic Nonhomothetic CES Preferences

Now, consider the specific case used in our basic model in Section 1.2, where the isoelastic functions g_i are defined as:

$$g_i(U) = U^{\frac{\epsilon_i-\sigma}{1-\sigma}}, \qquad \text{(A.15)}$$

where $\eta^U_{g_i} = (\epsilon_i - \sigma)/(1-\sigma)$, and we retrieve standard CES preferences when $\epsilon_i = 1$ for all i's. To tie our exposition more closely to the discussions in Section 1.2, let us for now identify utility with C, aggregate real income and define a corresponding aggregate price index $P \equiv E/C$. From Equations (A.2) and (A.3), we find demand to be:

$$\begin{aligned}
C_i &= \left(\frac{p_i}{P}\right)^{-\sigma} C^{\epsilon_i}, \qquad \text{(A.16)} \\
\omega_i &= \frac{p_i C_i}{PC} = \left(\frac{p_i}{P}\right)^{1-\sigma} C^{\epsilon_i - 1}. \qquad \text{(A.17)}
\end{aligned}$$

The aggregate price index is

$$P \equiv \frac{E}{C} = \left(\sum_{i=1}^{I} C^{\epsilon_i - 1} p_i^{1-\sigma} \right)^{\frac{1}{1-\sigma}}. \tag{A.18}$$

From (A.7), the real income elasticity of the expenditure function is:

$$\eta_E^C \equiv \frac{C}{E} \frac{\partial E}{\partial C} = \frac{\bar{\epsilon} - \sigma}{1 - \sigma}, \tag{A.19}$$

where $\bar{\epsilon} = \sum_i \omega_i \epsilon_i$. Therefore, a sufficient condition for the function $E\left(C; \{p_i\}_{i=1}^I\right)$ to be a one-to-one mapping for all positive prices is that all sectors have an income elasticity larger than the elasticity of substitution $\epsilon_i > \sigma$ if $\sigma < 1$ (and $\epsilon_i < \sigma$ if $\sigma > 1$). This directly follows from Lemma 6.

Combining Equations(A.12) and (A.13), we find

$$\frac{\partial^2 E}{\partial C^2} = \frac{E}{C^2} \frac{\bar{\epsilon} - \sigma}{1 - \sigma} \left(\frac{\bar{\epsilon} - \sigma}{1 - \sigma} - 1 \right) + \frac{Var(\epsilon)}{\bar{\epsilon} - \sigma}. \tag{A.20}$$

Therefore, a sufficient condition for the expenditure function to be convex in C for all prices is that $\epsilon_{min} \geq 1$.

The income elasticities of demand are given by Equations (A.10) and (A.11):

$$\eta_{C_i/C_j}^C = \epsilon_i - \epsilon_j, \tag{A.21}$$

$$\eta_{C_i}^E = \sigma + (1 - \sigma) \frac{\epsilon_i - \sigma}{\bar{\epsilon} - \sigma}. \tag{A.22}$$

Each good i is characterized by a parameter $\epsilon_i \in \mathbb{R}$ that is a measure of its real income elasticity.

More generally, the relationship between utility U and real aggregate consumption C in Expression (A.15) can be defined by any monotonic function G such that $U = G(C)$. In particular, let us define $G(\cdot)$ such that C corresponds to consumption expenditure at constant prices $\{q_i\}_i$ such that

$$C^{1-\sigma} = \sum_{i=1}^{I} \Omega_i \, G(C)^{\epsilon_i - 1} q_i^{1-\sigma}. \tag{A.23}$$

Assuming $\sigma \in (0, 1)$, if $\epsilon_i > \sigma$ for all i, function $G(\cdot)$ defined through Equation (A.23) is

monotonically increasing for all positive C. Therefore, we can approximate the relationship as:

$$\begin{aligned}\log G(C) &\approx \log G(\tilde{C}) + \left.\frac{\partial \log G}{\partial \log C}\right|_{C=\tilde{C}} \cdot \left(\log C - \log \tilde{C}\right), \\ &= \frac{1-\sigma}{\bar{\epsilon}-\sigma}\log C + \text{const.}, \end{aligned} \tag{A.24}$$

where $\bar{\epsilon}$ is the average elasticity parameter at constant price q and real income $\tilde{C}$.

A.1.3 Comparison to Generalized Stone-Geary and PIGL Preferences

For comparison, now consider generalized Stone-Geary preferences that have been widely used in previous work on structural change (see, e.g., Kongsamut *et al.*, 2001):

$$C = \left(\sum_{i=1}^{I} \left(C_i - \underline{C}_i\right)^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}}, \tag{A.25}$$

where $\underline{C}_i$ are the usual coordinate shifters.[3] The expenditure elasticity of demand for good i is given by:

$$\eta^E_{C_i} = 1 - \frac{\underline{C}_i}{C_i}, \tag{A.26}$$

which is different from 1 as long as the shifter $\underline{C}_i \neq 0$. However, note that due to constancy of $\underline{C}_i$, this elasticity converges to unity at the same rate as the rate of growth of C_i. Therefore, nonhomotheticity is a short-run feature of Stone-Geary preferences: as the income grows Stone-Geary preferences asymptote to homothetic CES preferences.

Another important feature of the nonhomothetic CES preferences is the fact that elasticity of substitution σ_{ij} between all goods i and j remains constant σ and remains independent of expenditure (income) elasticities. In contrast, for Stone-Geary preferences we find:

$$\sigma_{ij} = \sigma \cdot \frac{E}{E - \sum_k p_k \underline{C}_k} \cdot \left(1 - \frac{\underline{C}_i}{C_i}\right) \cdot \left(1 - \frac{\underline{C}_j}{C_j}\right), \tag{A.27}$$

$$\eta^E_{C_i} = \frac{E}{E - \sum_k p_k \underline{C}_k} \cdot \left(1 - \frac{\underline{C}_i}{C_i}\right), \tag{A.28}$$

[3] In particular, standard 3-sector models of structural transformation generally assume $\underline{C}_a > 0$, $\underline{C}_s < 0$ and $\underline{C}_m = 0$.

where η_i is the nominal income elasticity of demand in sector i (Hanoch, 1975a). It then follows that the elasticities of substitution between goods i and j always satisfies the following equality:

$$\sigma_{ij} = \sigma \eta^E_{C_i} \eta^E_{C_j}, \tag{A.29}$$

creating a direct linkage between elasticities of substitution and expenditure for different sectors. As expected, when E goes to infinity we find that $\sigma_{ij} \to \sigma$ and $\eta^E_{C_i} \to 1$ for all sectors.

An alternative specification for nonhomothetic preferences in the structural change literature, recently used by Boppart (2014a), is the Price Independent Generalized Linear (PIGL) preferences. The canonical definition for these preferences involves a two-good system. In general, no closed-form representation for the utility function exists, but the indirect utility/expenditure function relationship can be specified as:

$$C + \frac{\vartheta}{\varrho}\left[\left(\frac{p_1}{p_2}\right)^{\varrho} - 1\right] = \frac{1}{\xi}\left[\left(\frac{E}{p_2}\right)^{\xi} - 1\right], \tag{A.30}$$

where $\boldsymbol{p} = (p_1, p_2)$ is the pair of good prices, C is the aggregator (utility) and E is expenditure, $0 \leq \xi \leq \varrho < 1$, and $\vartheta > 0$.[4] For these preferences, the expenditure elasticity of demand for good i is constant and less than unity: $\eta^E_{C_1} = 1 - \xi < 1$. Therefore, like nonhomothetic CES preferences and unlike Stone-Geary, PIGL preferences also feature nonhomotheticity at all levels of income. In contrast to nonhomothetic CES, however, there is no generalization of PIGL preferences to more than two good demand systems that preserves the independence of income elasticities across different goods.[5]

Needless to say, since PIGL preferences are outside the CES family, the elasticity of substitution varies with income and prices. As Boppart (2014a) shows, the elasticity of

[4]PIGL preferences are *not* additive in the sense of Hanoch (1975a).

[5]As a reminder, from Engel aggregation we know that we can have up to $I - 1$ independent income elasticities in a demand system involving I goods. This is why we have one degree of freedom in specifying the I income elasticity parameters $\bar{\epsilon}$ in nonhomothetic CES preferences defined in Section A.1.2.

substitution between goods 1 and 2 are given by

$$\sigma = 1 - \varrho - (\varrho - \xi) \frac{\vartheta \left(\frac{p_1}{p_2}\right)^{\varrho}}{\left(\frac{E}{p_2}\right)^{\xi} - \vartheta \left(\frac{p_1}{p_2}\right)^{\varrho}}. \tag{A.31}$$

As a result, when PIGL preferences are embedded in a growth model, along an equilibrium path that involves growing income the elasticity of substitution will be monotonically increasing and converges toward $1 - \varrho$. Therefore, the choice of PIGL preferences involves specific assumptions about the dynamics of substitution elasticities in a two-good model.

A.2 Proofs of Propositions and Lemmas

Proof of Lemma 1. First, we show that the household problem has a unique solution that is characterized by an Euler equation along with a standard transversality condition. Let $E_t = w_t + (1 + r_t)\mathcal{A}_t - \mathcal{A}_{t+1}$ be the consumption expenditure when the representative household has current stock of assets $\mathcal{A}_t$ and chooses an allocation $\mathcal{A}_{t+1}$ of assets for the next period. We can decompose the problem into two independent parts. The intratemporal problem involves allocating the expenditure E_t across I goods so as to maximize the aggregator C_t defined by Equation (1.2). The solution is given by Equations (A.16) and (A.18).

Let $\tilde{C}_t(E) \equiv \max C_t$ subject to the constraint $E = \sum_{i=1}^{I} p_{it} C_{it}$. The intertemporal problem then involves finding the sequence of assets $\{\mathcal{A}_{t+1}\}_{t=0}^{\infty}$ such that

$$\max \sum_{t=0}^{I} \beta^t \frac{\tilde{C}_t \left(w_t + (1 + r_t)\mathcal{A}_t - \mathcal{A}_{t+1}\right)^{1-\theta} - 1}{1 - \theta}. \tag{A.32}$$

From Section A.1.2, we know that when $\epsilon_i \geq 1$ for all i, the expenditure function is monotonically increasing and strictly convex for all prices. Therefore, its inverse, the indirect aggregate consumption function $\tilde{C}(E; \boldsymbol{p}_t)$ exists and is monotonically increasing and strictly concave for all prices. Standard results from discrete dynamic programming (e.g., see

Acemoglu, 2008, Chapter 6) then imply that the Euler equation

$$C_t^{-\theta}\frac{\partial \tilde{C}_t}{\partial E_t} = \beta\left(1+r_t\right)C_{t+1}^{-\theta}\frac{\partial \tilde{C}_{t+1}}{\partial E_{t+1}},$$

and the transversality condition

$$\lim_{t\to\infty}\beta^t\left(1+r_t\right)\mathcal{A}_t C_t^{-\theta}\frac{\partial \tilde{C}_t}{\partial E_t}=0, \tag{A.33}$$

provide necessary and sufficient condition for a sequence $\{\mathcal{A}_{t+1}\}_{t=0}^{\infty}$ to characterize the solution.

Using Equation (A.19) we can simplify the Euler equation above to

$$C_t^{-\theta}\frac{C_t}{E_t}\frac{1-\sigma}{\bar{\epsilon}_t-\sigma}=\beta\left(1+r_t\right)C_{t+1}^{-\theta}\frac{C_{t+1}}{E_{t+1}}\frac{1-\sigma}{\bar{\epsilon}_{t+1}-\sigma},$$

and the transversality condition to

$$\lim_{t\to\infty}\beta^t\left(1+r_t\right)\frac{\mathcal{A}_t}{E_t}C_t^{1-\theta}\frac{1-\sigma}{\bar{\epsilon}_t-\sigma}=0.$$

□

Proof of Proposition 1. Our proof for the proposition involves two steps. First, we use the second Welfare Theorem and consider the equivalent centralized allocation by a social planner. Due to the concavity of the aggregator C_t as a function of the bundle of goods $(C_1,\cdots,C_I)$, which is ensured by the condition $\epsilon_i \geq 1$ for all i, we can use standard arguments to establish the uniqueness of the equilibrium allocations (see Stockey *et al.*, 1989, p. 291). Next, we construct a unique constant growth path (steady state) that satisfies the equilibrium conditions. It then follows that the equilibrium converges to the constructed Constant Growth Path (CGP).

Consider an equilibrium path along which consumption expenditure E_t, aggregate stock of capital K_t, and the capital allocated to the investment sector K_{0t} all asymptotically grow at rate $(1+\gamma_0)^{\frac{1}{1-\alpha_0}}$, and the labor employed in the investment sector asymptotically converges to $L_0^* \in (0,1)$. Henceforth, we use the tilde variables to denote normalization $A_{0t}^{-\frac{1}{1-\alpha_0}}$, for instance, $\tilde{K}_t \equiv A_{0t}^{-\frac{1}{1-\alpha_0}}K_t$. Accordingly, we can write the law of evolution of aggregate stock

of capital as

$$\tilde{K}_{t+1} = \frac{1-\delta}{(1+\gamma_0)^{1/(1-\alpha_0)}}\tilde{K}_t + \frac{1}{(1+\gamma_0)^{1/(1-\alpha_0)}}\tilde{K}_{0t}^{\alpha_0} L_{0t}^{1-\alpha_0}, \tag{A.34}$$

and the interest rate and wages as

$$r_t = R_t - \delta = \alpha_0 \left(\frac{\tilde{K}_{0t}}{L_{0t}}\right)^{\alpha_0 - 1} - \delta, \tag{A.35}$$

$$\tilde{w}_t = (1-\alpha_0)\,\tilde{K}_{0t}^{\alpha_0} L_{0t}^{-\alpha_0}. \tag{A.36}$$

From the assumptions above, it follows that $\tilde{K}_{0t}/L_{0t}$ asymptotically converges to a constant, which from Equation (A.35) implies that the rate of interest also converges to a constant r^*.

We first derive an expression for the asymptotic growth of nominal consumption expenditure shares (and sectoral employment shares) of different sectors, using in equation (1.6),

$$\begin{aligned} 1 + \xi_i \equiv \lim_{t\to\infty} \frac{\omega_{it+1}}{\omega_{it}} &= \lim_{t\to\infty} \left(\frac{E_t}{E_{t+1}}\right)^{1-\sigma} \left(\frac{p_{it+1}}{p_{it}}\right)^{1-\sigma} \left(\frac{C_{t+1}}{C_t}\right)^{\epsilon_i - \sigma}, \\ &= \left(\frac{1}{1+\gamma_0}\right)^{\frac{1-\sigma}{1-\alpha_0}} \left(\frac{(1+\gamma_0)^{\frac{1-\alpha_i}{1-\alpha_0}}}{1+\gamma_i}\right)^{(1-\sigma)} (1+\gamma^*)^{\epsilon_i - \sigma}, \\ &= \frac{(1+\gamma^*)^{\epsilon_i - \sigma}}{\left[(1+\gamma_0)^{\frac{\alpha_i}{1-\alpha_0}} (1+\gamma_i)\right]^{1-\sigma}}, \end{aligned} \tag{A.37}$$

where in the second line we have used the definition of the constant growth path as well as the fact that from Equations (A.35) and (A.36), the relative labor-capital price grows as rate $(1+\gamma_0)^{\frac{1}{1-\alpha_0}}$ and therefore from Equation (1.16) we have

$$\lim_{t\to\infty} \frac{p_{it+1}}{p_{it}} = \frac{1+\gamma_0}{1+\gamma_i} (1+\gamma_0)^{\frac{\alpha_0 - \alpha_i}{1-\alpha_0}}. \tag{A.38}$$

Equation (A.37) shows that the expenditure shares asymptotically grow (or diminish) monotonically. Since the shares belong to the compact $I-1$ dimensional simplex, they asymptotically converge to a time-constant set of shares.

Since shares have to add up to 1, we need to have that $\xi_i \leq 0$ for all i. Moreover, this inequality has to be satisfied with equality at least for one non-vanishing sector. Now, consider the expression defined in (1.21) for the growth rate of real consumption. For sectors

$i \in \mathcal{I}^*$ that achieve the minimum, the growth of nominal expenditure share becomes zero, and their shares converge to constant values ω_i^*. For sectors $i \notin \mathcal{I}^*$, we find the following expression for the growth rate of nominal shares ξ_i in Equation (A.37) becomes negative. Assuming $\sigma < 1$ and $\epsilon_i > \sigma$, the expression on the right hand side becomes strictly less than 1, since we know sector i does not achieve the minimum in (1.21). Therefore, $\xi_i < 0$ and the nominal shares asymptotically vanish for $i \notin \mathcal{I}^*$.

Asymptotically, the expenditure-weighted average income elasticity and expenditure-weighted capital intensity in the consumption sector both converge to constants $\bar{\epsilon}^* \equiv \lim_{t\to\infty} \sum_{i=1}^{I} \epsilon_i \omega_{it} = \sum_{i\in\mathcal{I}^*} \epsilon_i \omega_i^*$ and $\bar{\alpha}^* \equiv \lim_{t\to\infty} \sum_{i=1}^{I} \alpha_i \omega_{it} = \sum_{i\in\mathcal{I}^*} \alpha_i \omega_i^*$. Henceforth, we extend our notation to use tilde to indicate variables normalized by their corresponding asymptotic rate of growth (or decline) along our proposed constant growth path. For instance, we let $\tilde{p}_{it} \equiv p_{it}(1+\gamma_0)^{-\frac{1-\alpha_i}{1-\alpha_0}t}(1+\gamma_i)^{-t}$ and $\tilde{C}_t \equiv C_t(1+\gamma^*)^{-t}$. Furthermore, we define starred notation to indicate the asymptotic value of each variable along the constant growth path, for example, we let $p_i^* \equiv \lim_{t\to\infty} \tilde{p}_{it}$ and $\tilde{C}^* \equiv \lim_{t\to\infty} \tilde{C}_t$.

We now show that a constant growth path exists and is characterized by γ^* as defined by equation (1.21). We also show the existence of the asymptotic values $\{\tilde{K}^*, \tilde{C}^*, \tilde{K}_0^*, L_0^*\}$. From the Euler equation (1.8), we have that asymptotically

$$(1+\gamma^*)^{1-\theta} = \frac{(1+\gamma_0)^{\frac{1}{1-\alpha_0}}}{\beta\,(1+r^*)}, \tag{A.39}$$

which pins down r^*, the asymptotic real interest rate in terms of γ^* given by Equation (1.21). Then from Equation (A.35), we find the asymptotic capital-labor ratio in the investment sector in terms of the asymptotic real interest rate

$$\kappa \equiv \frac{\tilde{K_0}^*}{L_0^*} = \left(\frac{\alpha_0}{r^*+\delta}\right)^{\frac{1}{1-\alpha_0}}. \tag{A.40}$$

This gives us the asymptotic relative labor-capital price from Equations (A.35) and (A.36) as

$$\frac{\tilde{w}^*}{R^*} = \frac{1-\alpha_0}{\alpha_0}\frac{\tilde{K}^*}{L_0^*} = \frac{1-\alpha_0}{\alpha_0}\left(\frac{\alpha_0}{r^*+\delta}\right)^{\frac{1}{1-\alpha_0}}. \tag{A.41}$$

From Equation (1.16), we find

$$\tilde{p}_i^* = \frac{\alpha_0^{\alpha_0}\left(1-\alpha_0\right)^{1-\alpha_0}}{\alpha_i^{\alpha_i}\left(1-\alpha_i\right)^{1-\alpha_i}}\left(\frac{\tilde{w}^*}{R^*}\right)^{\alpha_0-\alpha_i}\frac{A_{0,0}}{A_{i,0}}, \tag{A.42}$$

where $\tilde{w}^*/R^*$ is given by Equations (A.41) and (A.39) and $A_{i,0}$ denotes the initial state of technology in sector i and $A_{0,0} \equiv 1$. Given asymptotic prices

$$\tilde{E}^* = \left[\sum_{i\in\mathcal{I}^*}\left(\tilde{C}^*\right)^{\epsilon_i-\sigma}\left(\tilde{p}_i^*\right)^{1-\sigma}\right]^{\frac{1}{1-\sigma}}, \tag{A.43}$$

and

$$\omega_i^* = \left(\frac{\tilde{p}_i^*}{\tilde{E}^*}\right)^{1-\sigma}\left(\tilde{C}^*\right)^{\epsilon_i-\sigma}. \tag{A.44}$$

Next, we combine the equation for accumulation of capital (A.34), the household budget constraint (1.10) the market clearing condition of consumption goods to establish that these exists a unique $\{\tilde{K}^*, \tilde{C}^*, \tilde{K}_0^*, L_0^*\}$ satisfying the asymptotic equilibrium conditions and $\kappa = \tilde{K}_0^*/L_0^*$ where κ is given by Equation (A.40). From market clearing, the sum of payments to labor in the consumption sector is $\sum_{i=1}^{I}(1-\alpha_i)\omega_{it}E_{it}$, which implies $\left(1-\bar{\alpha}_t\right)\tilde{E}_t = \tilde{w}_t\left(1-L_{0t}\right)$. Asymptotically, we find that

$$\left(1-\bar{\alpha}^*\right)\tilde{E}^* = \left(1-\alpha_0\right)\kappa^{\alpha_0}\left(1-L_0^*\right). \tag{A.45}$$

Similarly, from Equation (A.34) it follows that $\left[\left(1+\gamma_0\right)^{\frac{1}{1-\alpha_0}}-\left(1-\delta\right)\right]\tilde{K}^* = \kappa^{\alpha_0}L_0^*$. Defining the expression within the square brackets at a positive constant ϑ, we use write the asymptotic employment in the investment sector in terms of the aggregate stock of capital as

$$L_0^* = \vartheta\kappa^{-\alpha_0}\tilde{K}^*. \tag{A.46}$$

Finally, using the market clearing condition in the assets market $\mathcal{A}_t = K_t$ and Equation (1.10), we find that $\tilde{E}_t = \tilde{w}_t + R_t\tilde{K}_t - \left(\frac{\tilde{K}_{0t}}{L_{0t}}\right)^{\alpha_0}L_{0t}$ for all t. Taking the limit, it follows that

$$\tilde{E}^* = \left(1-\alpha_0\right)\kappa^{\alpha_0} + \alpha_0\kappa^{\alpha_0-1}\tilde{K}^* - \kappa^{\alpha_0}L_0^*. \tag{A.47}$$

Substituting from Equation (A.46) into Equations (A.45) and (A.47) yields,

$$\bar{\alpha}^* \tilde{E}^* = \alpha_0 \left(\kappa^{\alpha_0 - 1} - \vartheta \right) \tilde{K}^*. \tag{A.48}$$

We can show that the left hand side of this equation is a monotonically increasing function of $\tilde{C}^*$ with a given κ.[6] From condition (1.22), we have that $\kappa^{\alpha_0 - 1} - \vartheta > 0$ and therefore the right hand side is a linear increasing function of $\tilde{K}^*$. Therefore, Equation (A.48) defines $\tilde{C}^*$, and correspondingly $\tilde{E}^*$, as an increasing function of $\tilde{K}^*$. Finally, substituting this function and Equation (A.46) in Equation (A.47), we find

$$\tilde{E} + \left(\vartheta - \alpha \kappa^{\alpha_0 - 1} \right) \tilde{K} = (1 - \alpha_0) \kappa^{\alpha_0}. \tag{A.49}$$

From condition (1.22), we know that the left hand side is a monotonically increasing function of $\tilde{K}^*$ for constant κ. This function is 0 when $\tilde{K}^*$ and limits to infinity as the latter goes to infinity. Therefore, Equation (A.49) uniquely pins down $\tilde{K}^*$ as a function of κ, which in turn is given by Equation (A.40). Condition (1.22) also ensures that the transversality condition (1.9) is satisfied. Finally, we verify that $L_0^* \in (0, 1)$. Combining equations (A.46), (A.45) and (A.47) we obtain that

$$L_0^* = \frac{\bar{\alpha}}{\left[\frac{1 - \bar{\alpha}}{1 - \alpha_0} \left(\alpha_0 \kappa^{\alpha_0 - 1} \vartheta^{-1} - 1 \right) + 1 \right]} \tag{A.50}$$

Assuming that the term in square brackets is positive, we have that $L_0^* \in (0, 1)$ if and only if $\vartheta < \kappa^{\alpha_0 - 1}$, which in terms of fundamental parameters requires that $\beta (1 + \gamma^*)^{1 - \theta} < \frac{(1+\gamma_0)^{-\frac{\alpha_0}{1-\alpha_0}}}{\alpha_0 + (1-\alpha_0)(1+\gamma_0)^{-\frac{1}{1-\alpha_0}}(1-\delta)}$ which is the condition stated in (1.22). Also, it is readily verified that as long as $\vartheta < \kappa^{\alpha_0 - 1}$, L_0^* cannot be negative.

Therefore, we constructed a unique costant growth path that asymptotically satisfies the equilibrium conditions whenever the parameters of the economy satisfy Equation (1.22). Together with the uniqueness of the competitive equilibrium, this completes the proof.

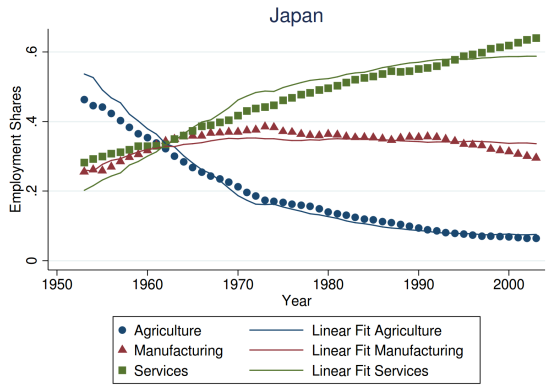
□

[6] We have that $\frac{\partial (\bar{\alpha}^* \tilde{E}^*)}{\partial \tilde{C}^*} = \frac{\bar{\alpha}^* \tilde{E}^*}{\tilde{C}^*} \frac{\bar{\epsilon} - \sigma}{1 - \sigma} \left[1 + (1 - \sigma) \rho_{\epsilon_i - \sigma, \alpha_i} \right]$ where $\rho_{\epsilon_i - \sigma, \alpha_i}$ is the correlation coefficient between $\varepsilon_i - \sigma$ and α_i under a distribution implied by expenditure shares (see online Appendix for details of the derivation). Therefore, the derivative is always positive and the function is a monotonic of $\tilde{C}^*$.

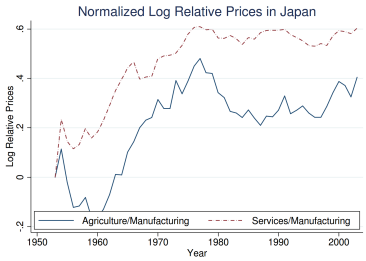
A.3 Additional Figures

Figure A.1: *Regression Fit for Japan using common world parameters* $\{\sigma, \epsilon_a - \epsilon_m, \epsilon_s - \epsilon_m\}$

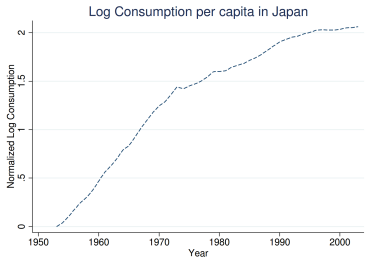
(a) *Regression Fit using all regressors*



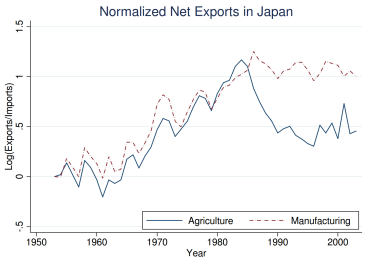
(b) *Relative Prices*



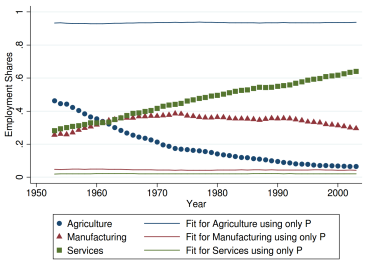
(c) *Consumption*



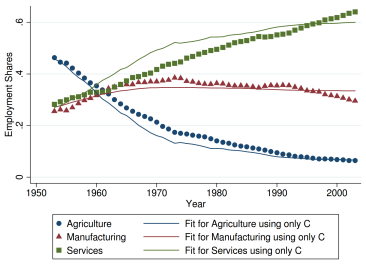
(d) *Net Exports*



(e) *Partial fit: Prices only*



(f) *Partial fit: Consumption only*



(g) *Partial fit: Net Exports only*

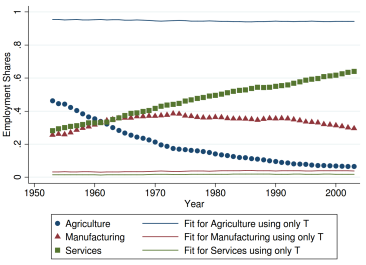


Figure A.2: *Baseline Country Fit*

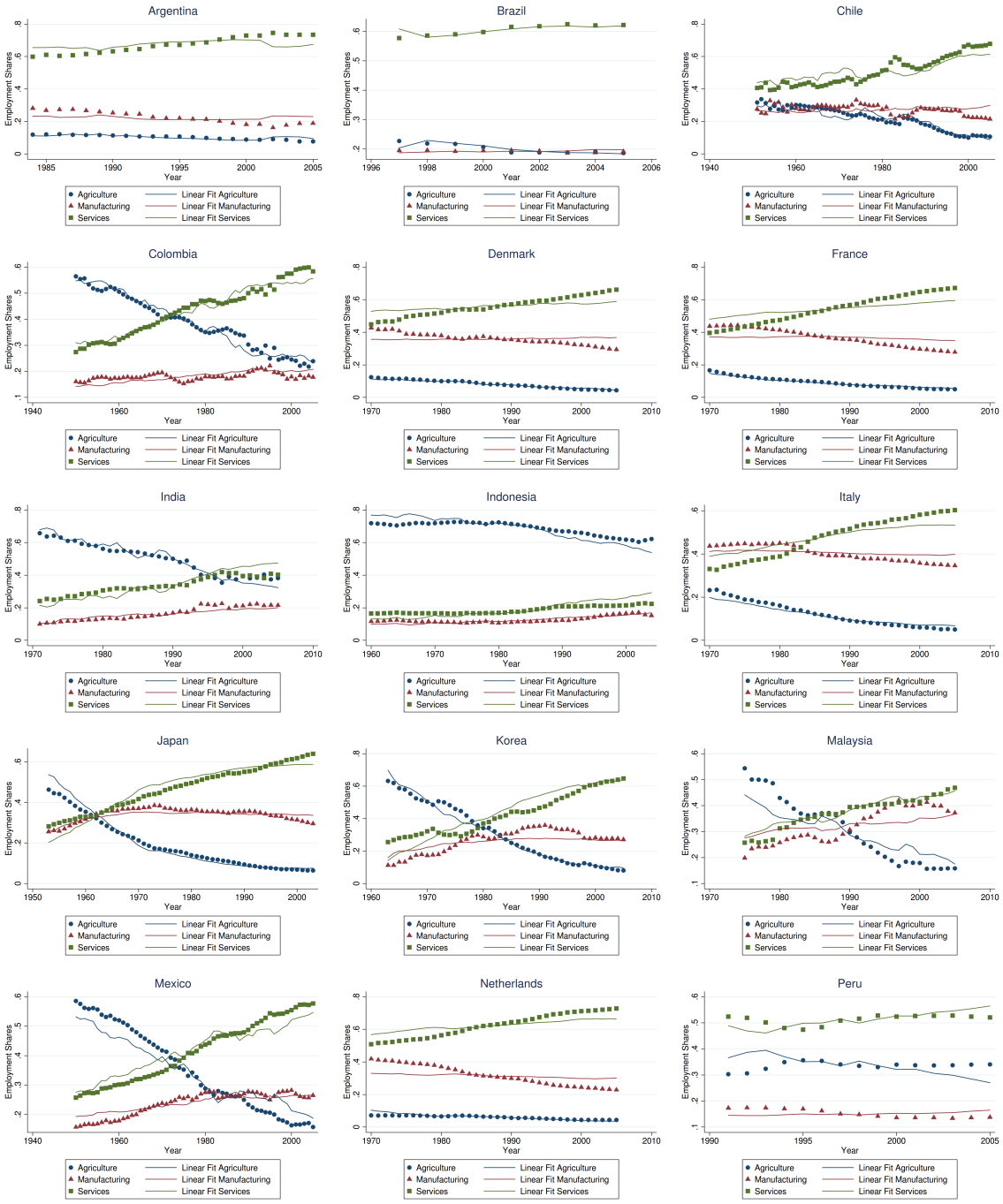


Figure A.3: *Baseline Country Fit (continued)*

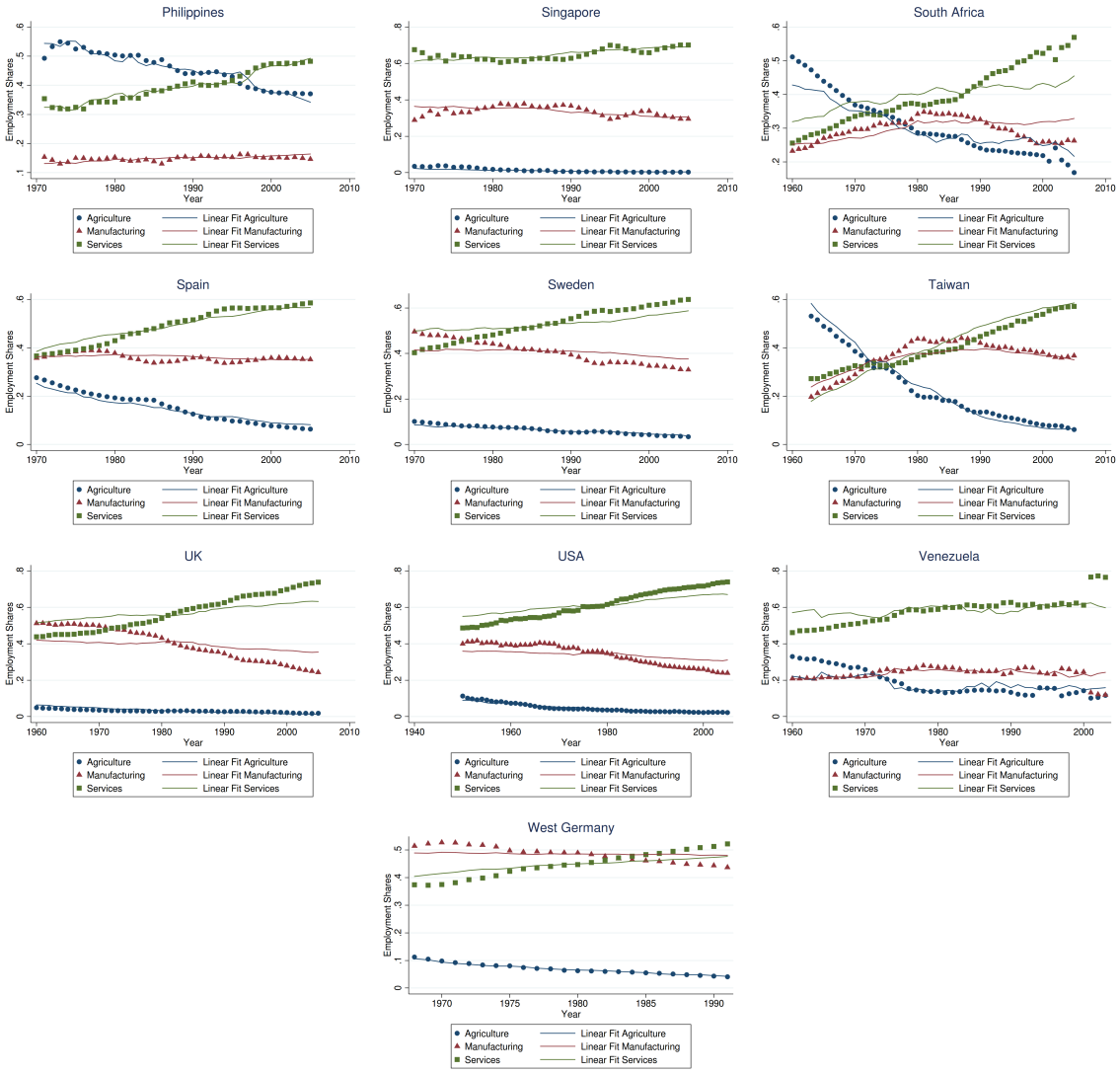


Table A.1: *Residual Relative Log-Expenditure Shares on Residual Aggregate Consumption*

Dep. Var.:	$\log\left(\frac{\text{Agriculture}}{\text{Manufact.}}\right)$		$\log\left(\frac{\text{Services}}{\text{Manufact.}}\right)$	
Residual Log-Expenditure	(1)	(2)	(3)	(4)
Residual Log Aggregate Income	-0.34	-0.42	0.27	0.23
	(0.11)	(0.12)	(0.08)	(0.07)
Residual Log Agg. Income below Median Income		0.19		0.09
		(0.12)		(0.13)
R^2	0.27	0.29	0.20	0.21
Observations	513	513	513	513

Note: Standard errors clustered by country. Residual Aggregate Income is constructed by taking the residuals of the following OLS regresssion: $\log Y_t^c = \alpha \log p_{at}^c + \beta \log p_{mt}^c + \gamma \log p_{st}^c + \xi^c + \nu_t^c$ where superscript c denotes country, and subscript t, time. p_{at}^c denotes price of agriculture in country c at time t. Likewise p_{mt}^c and p_{st}^c denote the prices of manufacturing and services, respectively. ξ^c denotes a country fixed effect and ν_t^c the error term. Residual log-expenditures are constructed in an analogous manner.

Table A.2: *Contribution of Relative Prices and Consumption*

Specification	Log-Likelihood	LR Test		AIC	BIC
		χ^2	p-value		
FE Only	-324.28	–	–	754.56	1010.02
FE + Prices	-270.30	107.96	0.00	648.61	908.89
FE + Consumption	363.34	1375.25	0.00	-616.68	-351.58
Full Specification	412.08	1472.71 97.47	0.00 0.00	-712.15	-442.23

Note: AIC refers to the Akaike Information Criterion, BIC refers to the Bayesian Information Criterion. The first two Likelihood Ratio Tests are done against the model that has only country-(relative)sector fixed effects. The last Likelihood Ratio Test compares the full model against one with fixed effects and consumption.

Table A.3: *Partial Correlations*

Regression Equation	Consumption		Relative Prices	
	Partial Corr.	Partial Corr.^{2}	Partial Corr.	Partial Corr.^{2}
L_a/L_m	-0.85	0.72	0.11	0.01
L_s/L_m	0.46	0.21	0.17	0.03

Note: Country*(relative)sector fixed effects included. Suppose that y is determined by $x_1, x_2, \ldots, x_k$. The partial correlation between y and x_1 is estimates the correlation that would be observed between y and x_1 if the other x's did not vary. The squared correlations estimate the proportion of the variance of y that is explained by each.

Table A.4: *Heterogeneous Price Elasticity of Substitution*

Dep. Var: Emp. Shares	Agri.-Manu. (1)	Serv.-Manu (2)
σ_{am}	0.67	
	(0.12)	
σ_{sm}		0.78
		(0.18)
$\varepsilon_a - \varepsilon_m$	-1.01	
	(0.13)	
$\varepsilon_s - \varepsilon_m$		0.33
		(0.13)
$c \cdot sm$ Fixed Effects	Y	Y
Trade Controls	Y	Y

Note: Standard errors clustered by country.

Table A.5: *Value Added Regressions*

Dep. Var: Value Added	(1)	(2)
σ	0.91	0.51
	(0.25)	(0.14)
$\varepsilon_a - \varepsilon_m$	-1.00	-1.17
	(0.21)	(0.15)
$\varepsilon_s - \varepsilon_m$	0.22	0.10
	(0.08)	(0.15)
Observations	1043	1043
$c \cdot sm$ Fixed Effects	N	Y

Note: Standard errors clustered by country.

Table A.6: *Stone-Geary Estimation Results*

Dep. Var.: Emp. Shares	(1)	(2)
σ	0.21	0.21
	(0.05)	(0.05)
$\bar{c}_a$	-208	-201
	(205)	(194)
$\bar{c}_s$	415	395
	(475)	(452)
Log-Likelihood	173677	173645
Akaike IC	-347348	-347274
Bayesian IC	-347333	-347236
Sectoral Trade Control	N	Y
Measures of Non-homotheticity		
Entire Sample		
max $\frac{p_s\bar{c}_s}{\sum_{i\in\{a,m,s\}} p_i C_i}$	-0.00	-0.00
min $\frac{p_s\bar{c}_s}{\sum_{i\in\{a,m,s\}} p_i C_i}$	-0.11	-0.11
max $\frac{p_a\bar{c}_a}{\sum_{i\in\{a,m,s\}} p_i C_i}$	-0.00	-0.00
min $\frac{p_a\bar{c}_a}{\sum_{i\in\{a,m,s\}} p_i C_i}$	-0.11	-0.11
United States		
max $\frac{p_s^{U.S.}\bar{c}_s}{\sum_{i\in\{a,m,s\}} p_i^{U.S.} C_i^{U.S.}}$	4.9e-04	4.9e-04
min $\frac{p_a^{U.S.}\bar{c}_a}{\sum_{i\in\{a,m,s\}} p_i^{U.S.} C_i^{U.S.}}$	-5.4e-04	-5.4e-04

Note: Standard errors clustered by country.

Table A.7: *Contribution of Prices and Consumption for OECD Countries*

Specification	Log-Likelihood	LR Test χ^2	LR Test p-value	AIC	BIC
FE Only	43.84	–	–	-45.68	33.91
FE + Prices	119.21	150.74	0.00	-194.41	-111.03
FE + Consumption	428.92	770.16	0.00	-811.84	-724.67
Full Specification	441.01	794.35 24.18	0.00 0.00	-834.03	-743.07

Note: AIC refers to the Akaike Information Criterion, BIC refers to the Bayesian Information Criterion. The first two Likelihood Ratio Tests are done against the model that has only country-(relative)sector fixed effects. The last Likelihood Ratio Test compares the full model against one with fixed effects and consumption.

Regression Equation	Consumption Partial Corr.	Consumption Partial Corr.^{2}	Relative Prices Partial Corr.	Relative Prices Partial Corr.^{2}
L_a/L_m	-0.52	0.27	-0.24	0.06
L_s/L_m	0.88	0.77	0.02	0.00

Note: Country*(relative)sector fixed effects included. Suppose that y is determined by $x_1, x_2, \ldots, x_k$. The partial correlation between y and x_1 is estimates the correlation that would be observed between y and x_1 if the other x's did not vary. The squared correlations estimate the proportion of the variance of y that is explained by each.

Figure A.4: *Predicted Evolution of Expenditure Shares for the U.S.*

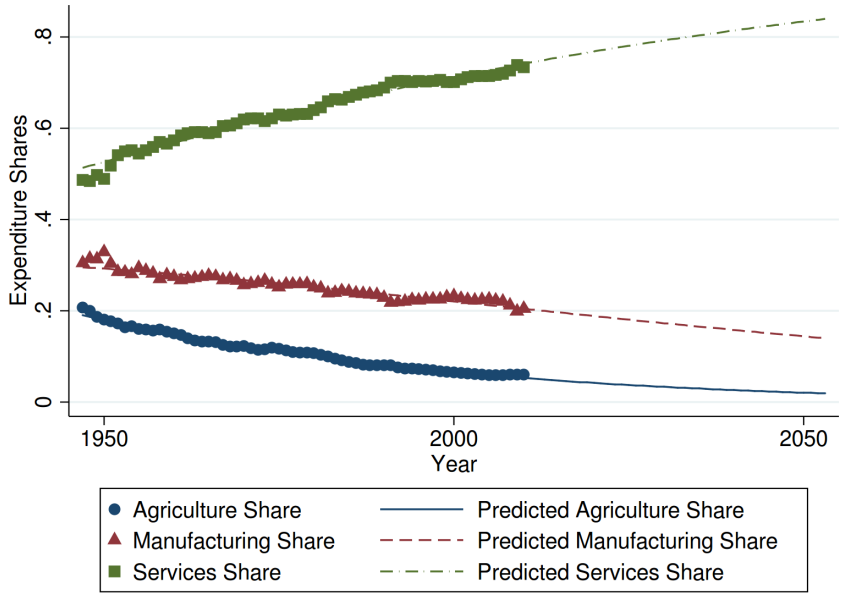


Table A.8: *Contribution of Prices and Consumption for Non-OECD Countries*

Specification	Log-Likelihood	LR Test		AIC	BIC
		χ^2	p-value		
FE Only	-186.07	–	–	442.13	595.38
FE + Prices	-182.83	6.47	0.01	437.66	595.28
FE + Consumption	261.90	895.94	0.00	-449.81	-287.80
Full Specification	280.09	932.32 36.38	0.00 0.00	-484.19	-317.81

Note: AIC refers to the Akaike Information Criterion, BIC refers to the Bayesian Information Criterion. The first two Likelihood Ratio Tests are done against the model that has only country-(relative)sector fixed effects. The last Likelihood Ratio Test compares the full model against one with fixed effects and consumption.

Regression Equation	Consumption		Relative Prices	
	Partial Corr.	Partial Corr.^{2}	Partial Corr.	Partial Corr.^{2}
L_a / L_m	-0.88	0.78	0.17	0.03
L_s / L_m	0.22	0.05	0.00	0.00

Note: Country*(relative)sector fixed effects included. Suppose that y is determined by $x_1, x_2, \ldots, x_k$. The partial correlation between y and x_1 is estimates the correlation that would be observed between y and x_1 if the other x's did not vary. The squared correlations estimate the proportion of the variance of y that is explained by each.

Table A.9: *Growth Rates of Relative Prices in the Country Panel*

$$\log\left(\frac{p^c_{i,t}}{p^c_{m,t}}\right) = \alpha^c_{im} + \beta_i \cdot \text{Year} + \varepsilon^c_{im,t}, \qquad i = \{s, a\}$$

	$\log\left(\frac{p^c_a}{p^c_m}\right)$	$\log\left(\frac{p^c_s}{p^c_m}\right)$
Year	-0.59	0.13
	(0.05)	(0.04)
Country-Sector FE	Yes	Yes
R^2	0.49	0.41
Observations	1680	1680

Note: Year has been re-scaled to Year/100.

Figure A.5: *Country Fit Using Stone-Geary Preferences*

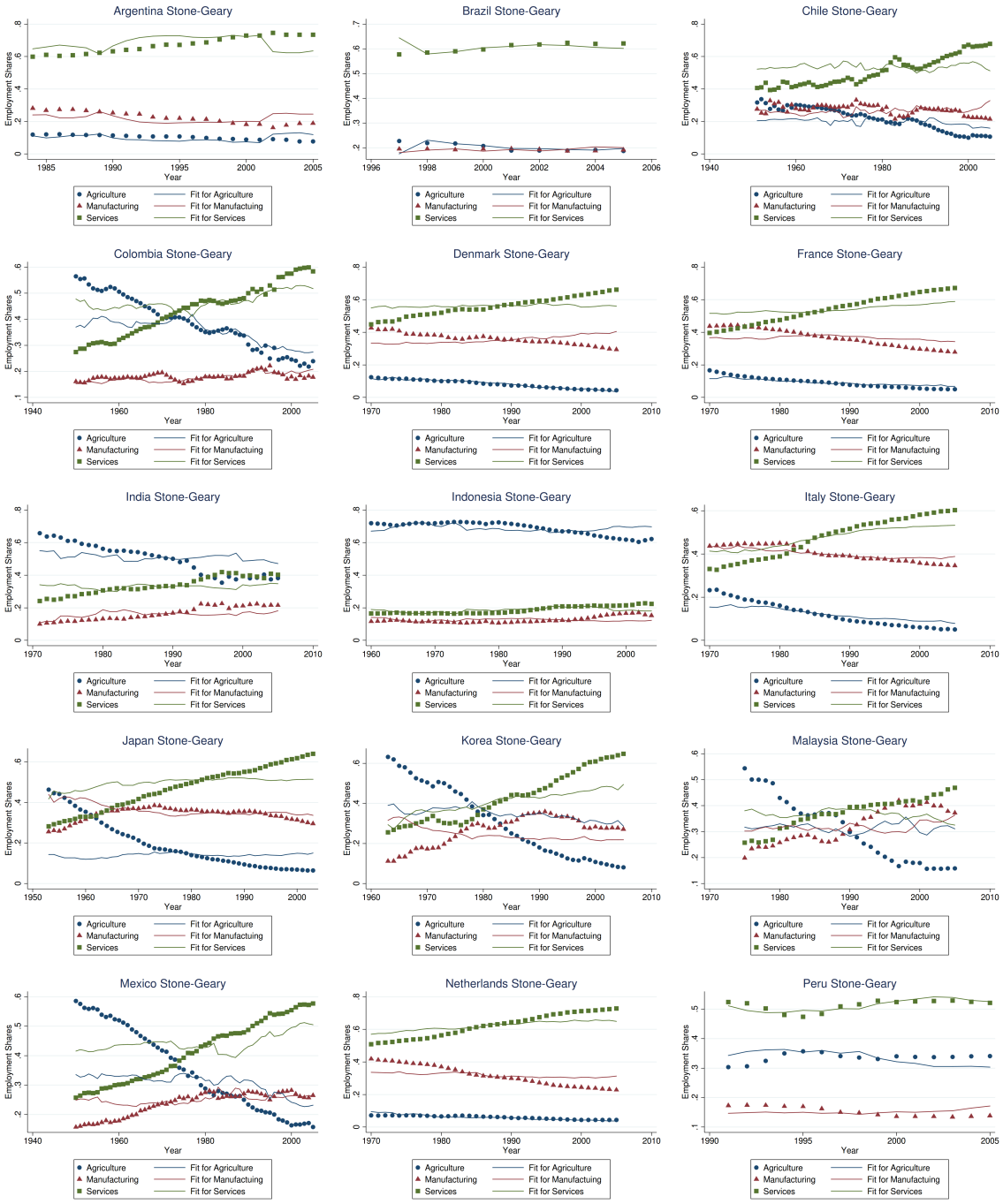
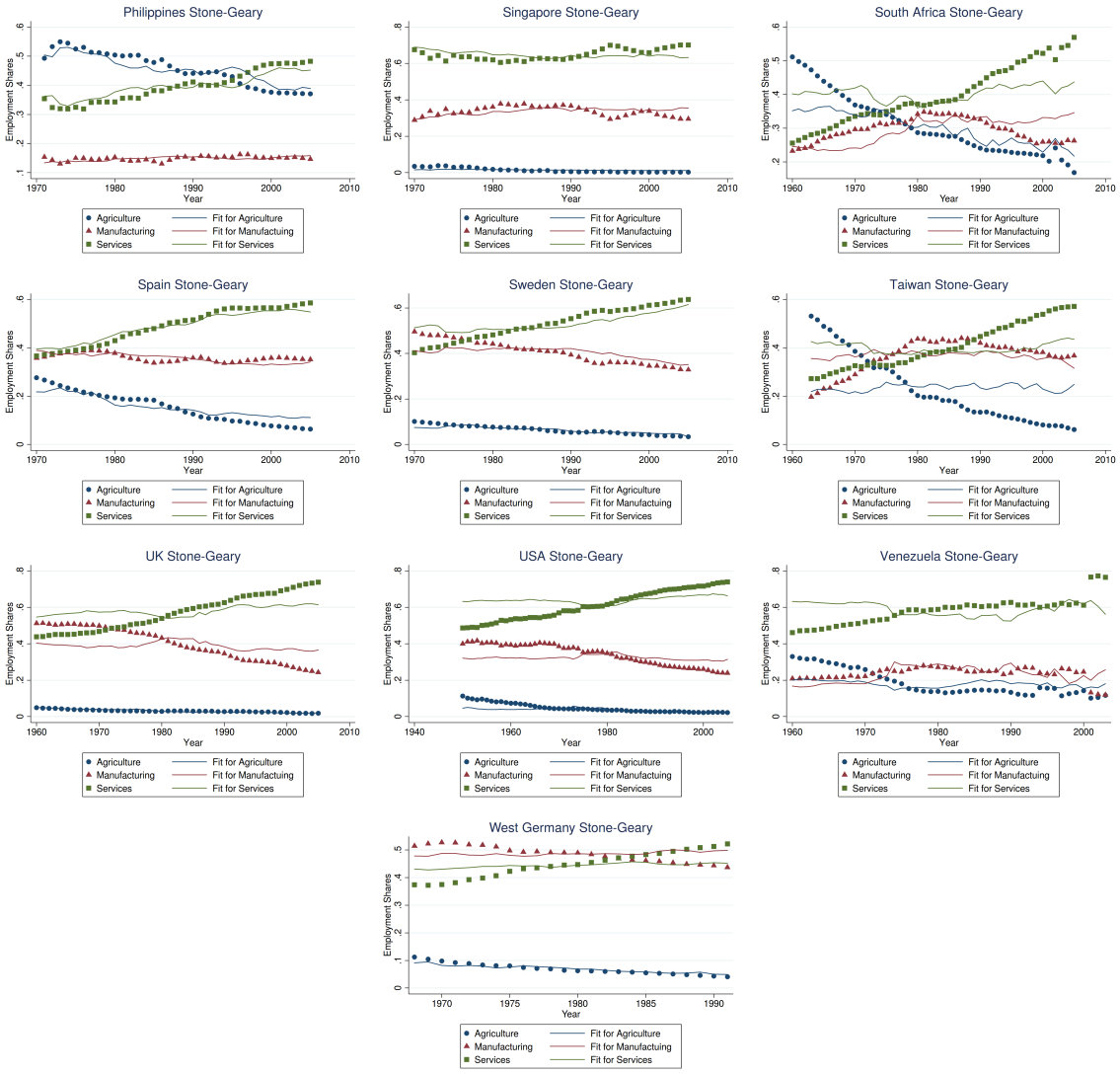


Figure A.6: *Country Fit Using Stone-Geary Preferences*



Appendix B

Appendix to Chapter 2

B.1 Proofs

Proposition. *Proposition 2.*

Proof. The derivations for the intratemporal allocation as given by Equation (2.3) with the price index as defined by Equation (2.4) are straightforward (see Comin *et al.*, 2015). We only present the proof on the intertemporal component of the solution.

For a given path of real interest rate $[r(t)]_{t=0}^{\infty}$ and sectoral good prices $[\boldsymbol{P}(t)]_{t=0}^{\infty}$, the current-value Hamiltonian for the consumer problem (2.1) may be written as

$$\hat{\mathcal{H}} \equiv \frac{C(t)^{1-\theta}-1}{1-\theta} + \lambda(t)\left[1 + r(t)A(t) - \mathcal{E}(C(t);\boldsymbol{P}(t))\right],$$

where we have defined the expenditure function $\mathcal{E} \equiv \mathcal{P}(C(t);\boldsymbol{P}(t)) \times C(t)$, where the price-index function is defined by Equation (2.4). Let us start with the necessary conditions. The FOCs for the Hamiltonian are as follows:

$$\frac{\partial \hat{\mathcal{H}}}{\partial C} = 0 \quad \Rightarrow \; C^{-\theta} - \lambda \frac{\partial \mathcal{E}}{\partial C} = 0, \tag{B.1}$$

$$\frac{\partial \hat{\mathcal{H}}}{\partial A} = \rho\lambda - \dot{\lambda} \quad \Rightarrow \; -\frac{\dot{\lambda}}{\lambda} = r - \rho. \tag{B.2}$$

In addition, we impose that the solution satisfy the transversality condition:

$$\lim_{t\to\infty} e^{-\rho t}\lambda\left(t\right)A\left(t\right)=0. \tag{B.3}$$

Equations (B.1) and (B.2) together with the law of evolution of assets (2.5) and the transversality equation (2.6) characterize paths of per capita real aggregate consumption and asset holdings $[C\left(\cdot\right),A\left(\cdot\right)]$, and costate $\lambda\left(\cdot\right)$ that satisfy necessary conditions for optimality.

Next, we show the conditions that ensure the solution above indeed corresponds to the unique solution to the household utility maximization problem. A standard argument (using (B.2) and the No-Ponzi constraint) shows that *for all feasible pairs* $[C\left(\cdot\right),A\left(\cdot\right)]$, we have that $\lim_{t\to\infty}\exp\left(-\rho t\right)\lambda\left(t\right)A\left(t\right)\geq 0$.[1] Therefore, we can establish that the pair characterized by Equations (B.1), (B.2), and (B.3) indeed correspond to the optimum if the Hamiltonian is concave in C. Furthermore, since the Hamiltonian is separable in (C,A) and linear A, strict concavity in C implies the uniqueness of the optimum for the household problem.

Combining Equations (B.1), (B.2), and (B.3) with the definition of the expenditure function $\mathcal{E}=\mathcal{P}C$ and the price index (2.4) gives the Euler equation (2.7) and the transversality condition (2.8). It remains for us to find conditions that ensure the strict concavity of $\mathcal{E}$ in C to ensure the sufficiency of the conditions above and uniqueness of the solution.

The second order condition for C is

$$\begin{aligned} o &> -\theta C^{-(\theta+1)}-\lambda\frac{\partial^2\mathcal{E}}{\partial C^2}, \\ &= -C^{-(\theta+1)}\left(\theta+\eta_{\mathcal{E}_C}\right), \end{aligned}$$

where $\eta_{\mathcal{E}_C}\equiv\frac{C\frac{\partial^2\mathcal{E}}{\partial C^2}}{\frac{\partial\mathcal{E}}{\partial C}}$ denotes the elasticity of marginal expenditure with respect to real consumption C. In the equality above, we have substituted for $\lambda=\left(C^{\theta}\partial\mathcal{E}/\partial C\right)^{-1}$ from Equation (B.1). The second order condition therefore implies that a sufficient condition for the strict concavity of the Hamiltonian to be $\eta_{\mathcal{E}_C}>-\theta$.

[1] For a discussion of necessary and sufficient conditions for the optimality of a solution to similar continuous-time dynamic programming problems, see Acemoglu (2009).

Using Equation (2.4), we can compute the elasticity to find

$$\eta_{\mathcal{E}_C} = \overline{\epsilon}(t)\left[1 + (1-\sigma)\, Var\left(\frac{\epsilon_i}{\overline{\epsilon}(t)}, t\right)\right] - 1.$$

It then follows that conditions $\sigma \in (0,1]$ and $\epsilon_i > 1-\theta$ for all i are sufficient to ensure the strict concavity of the Hamiltonian in C. □

Lemma. *See Lemma 2.*

Proof. We can substitute for the rate of growth of R&D productivity in sector i as

$$\dot{s}_i(t) = \delta \cdot \dot{n}_i(t) + (1-\delta) \cdot \dot{s}(t).$$

Substituting this expression in Equation (2.28), the rate of growth of productivity in sector i should satisfy

$$\dot{n}_i(t) = \frac{1}{\delta}\left(-\left[r(t) + (1-\delta)\,\dot{s}(t)\right] + \frac{1}{\zeta A(t)}\frac{L(t)}{H}\frac{\Omega_i(t)}{\Lambda_i(t)}\right). \tag{B.4}$$

Once again, using the free entry condition, we can write the rate of sectoral productivity growth as

$$\dot{n}_i(t) = \frac{\dot{N}_i(t)}{N_i(t)} = \frac{Z_i(t)}{N_i(t)\, V_i(t)} = \frac{1}{A(t)}\frac{Z_i(t)}{H}\frac{1}{\Lambda_i(t)}, \tag{B.5}$$

where we have combined Equations (2.12) and (2.22). Using this result, multiplying both sides of Equation (B.4) by $A(t)\,\Lambda_i(t)$ and summing over i we can find the expression inside the brackets in Equation (B.4) to be given by □

$$A(t) \cdot \left[r(t) + (1-\delta) \cdot \dot{s}(t)\right] = -\delta\frac{Z(t)}{H} + \frac{1}{\zeta}\frac{L(t)}{H}. \tag{B.6}$$

Substituting this expression back in Equation (B.4) and using Equation (B.5) we find that, if $Z_i(t) > 0$ for all i, the allocations of R&D employment and the rates of sectoral growth are given by Equations (2.30) and (2.31).

To find the expression for the interest rate $r(t)$, we first compute the rate of growth of

the general-purpose stock of knowledge,

$$\begin{aligned}\dot{s} &= \frac{\sum_i \vartheta_i^{1-\varsigma} N_i^{\varsigma} \dot{n}_i}{\sum_i \vartheta_i^{1-\varsigma} N_i^{\varsigma}}, \\ &= \frac{1}{A}\frac{Z}{H} + \frac{1}{\delta\zeta}\frac{L}{H}\left[\sum_i \vartheta_i^{1-\varsigma}\eta_i \left(\frac{N_i}{S}\right)^{\varsigma+\delta-1} \Omega_i - \frac{1}{A}\right],\end{aligned}$$

where in the second expression, we have substituted from Equation (2.31) and have used the fact that $\Lambda_i \equiv \frac{\eta_i}{A}\left(N_i/S\right)^{1-\delta}$ and $S^{\varsigma} \equiv \sum_i \vartheta_i N_i^{\varsigma}$. Substituting this expression in Equation (B.6), we find the expression for the interest rate to be given by Equation (2.32).

Finally, we can rewrite the rates

$$\dot{n}_i\left(t\right) = \frac{1}{\delta}\left[-r\left(t\right) - \left(1-\delta\right)\dot{s}\left(t\right) + \frac{1}{1+\zeta}\frac{P_i\left(t\right)Y_i\left(t\right)}{Z_i\left(t\right)}\dot{n}_i\left(t\right)\right]$$

of technical growth as

$$\dot{n}_i\left(t\right) = \frac{r\left(t\right) + \left(1-\delta\right)\dot{s}\left(t\right)}{\frac{1}{1+\zeta}\frac{P_i(t)Y_i(t)}{Z_i(t)} - \delta},$$

suggesting a one-to-one relationship between sectoral R&D intensity ($\frac{Z_i}{P_iY_i}$) and the sectoral rate of technological growth. Note that this expression suggests the following constraint on the R&D intensity of sectors, satisfied everywhere along an equilibrium path

$$\frac{Z_i\left(t\right)}{P_i\left(t\right)Y_i\left(t\right)} < \frac{1}{\delta\left(1+\zeta\right)}.$$

Lemma. *See Lemma 3.*

Proof. Along a CGP as characterized by Definition (2.3.6), the growth in the relative shares of sectors i and j can be found from Equation (2.3) to be

$$\lim_{t\to\infty} \dot{\omega}_i\left(t\right) - \dot{\omega}_j\left(t\right) \quad \equiv \quad \left(1-\sigma\right)g^*\left[\left(\epsilon_i - \epsilon_j\right) - \frac{\gamma_i - \gamma_j}{\zeta}\right],$$

which is a constant. Therefore, relative shares asymptotically evolve in a monotonic fashion; their time derivatives maintaining their signs. Since shares are nonnegative numbers that sum to one, they belong to a compact set and therefore have to converge to constant shares $\{\Omega_i^*\}_{i=1}^I$.

Under the asymptotic distribution $\{\Omega_i^*\}_{i=1}^I$, we can define $\overline{\gamma}^*$ and $\overline{\epsilon}^*$ to be average

expenditure-weighted technical growth rates and income elasticity parameters across sectors. From Equation (2.9), we find that the growth rate of consumption expenditure is given by

$$\lim_{t\to\infty} \dot{e}(t) = \left(\overline{\epsilon}^* - \frac{1}{\zeta}\overline{\gamma}^*\right) g^*,$$

which, again, suggests that $E(t)$ either asymptotically grows or falls at a constant rate. From the market clearing we know that $E(t) = \frac{1+\zeta}{\zeta}\frac{L(t)}{H} \leq \frac{1+\zeta}{\zeta}$ and therefore $E(t)$ also belongs to a closed and bounded set. Therefore, $\lim_{t\to\infty} E(t) = E^* > 0$, where the strict positivity follows from the transversality condition (2.8). This implies that the production employment also converges to a constant $L^* > 0$. Furthermore, the growth of $E(t)$ has to asymptotically be zero, that is,

$$\zeta = \frac{\overline{\gamma}^*}{\overline{\epsilon}^*}. \tag{B.7}$$

Since all shares are converge to constants, we have that

$$\begin{aligned}
\lim_{t\to\infty} \dot{\omega}_i(t) &= (1-\sigma) g^* \left((\epsilon_i - \overline{\epsilon}^*) - \frac{\gamma_i - \overline{\gamma}^*}{\zeta}\right), \\
&= (1-\sigma) g^* \left(\epsilon_i - \frac{\gamma_i}{\zeta}\right) \leq 0,
\end{aligned}$$

where we have used Equation (B.7) in the second equality . This suggests that $\zeta\epsilon_i \leq \gamma_i$ and therefore $\zeta \leq \gamma_i/\epsilon_i$ for all i, implying the result

$$\zeta = \min_i \left\{\frac{\gamma_i}{\epsilon_i}\right\}.$$

Finally, Equations (2.37) and (2.38) follow from the relations above and Equations (2.3) and (2.4).

□

Lemma. *See Lemma 4.*

Proof. First, from $\dot{n}_i(t) = S_i(t) Z_i(t) / N_i(t) \longrightarrow \gamma_i g^*$ and from the previous lemma, we know that $\gamma_i \geq \zeta\epsilon_i > 0$. We then asymptotically find

$$\gamma_i^S - \gamma_i + \lim_{t\to\infty} \dot{z}_i(t) = 0.$$

Since the R&D employment Z_i is bounded by Z^*, its rate of growth above cannot be positive. Therefore, we have that $Z_i(t)$'s converge to constants and we have

$$\gamma_i^S \geq \gamma_i, \tag{B.8}$$

for all i. For a sector $i \in \mathcal{I}^\dagger$ for which the asymptotic share of R&D employment is not zero, we have $\gamma_i^S = \gamma_i$ and its size of R&D employment is given by

$$Z_i^* = \gamma_i g^* \times \lim_{t\to\infty}\left(\frac{N_i(t)}{S_i(t)}\right)$$

□

Proof. We can directly employ the free entry condition to study the properties of the CGP. First, notice that

$$\begin{aligned}
\lim_{t'\to\infty}\frac{\Pi_i(t'+t)}{\Pi_i(t')} &= \lim_{t'\to\infty}\left(\frac{N_i(t')}{N_i(t'+t)}\right)\left(\frac{\Omega_i(t'+t)}{\Omega_i(t')}\right)\frac{Y(t'+t)}{Y(t')}, \\
&= \exp\left[\left(\underbrace{-\gamma_i}_{\text{technology}} \underbrace{+(1-\sigma)\left(\epsilon_i-\frac{\gamma_i}{\zeta}\right)}_{\text{market size}}\right)g^* t\right], \\
&= e^{-\left[\left(1+\frac{1-\sigma}{\zeta}\right)\gamma_i-(1-\sigma)\epsilon_i\right]g^* t},
\end{aligned} \tag{B.9}$$

where in the second line, we have used the fact that $\lim_{t\to\infty} Y(t) = HE^*$ and have substituted for the asymptotic rate of growth of sectoral output shares from Equation (2.36). Note that since $\Omega_i(t)$ is a share variable, we have that $\lim \dot{\omega}_i \leq 0$ with equality being satisfied at least by one sector. Now, we find the asymptotic value of owning a firm in sector i to be

$$\begin{aligned}
\lim_{t\to\infty}\frac{V_i(t)}{\Pi_i(t)} &= \left(\lim_{t\to\infty}\int_0^\infty e^{-\left(\frac{1}{t'}\int_0^{t'} r(t+t')\right)t'}\left(\frac{\Pi_i(t+t')}{\Pi_i(t)}\right)dt',\right) \\
&= \int_0^\infty e^{-\left[r^*+g^*\left(\gamma_i-(1-\sigma)\left(\epsilon_i-\frac{\gamma_i}{\zeta}\right)\right)\right]t'}dt', \\
&= \frac{1}{r^*+g^*\left(\gamma_i-(1-\sigma)\left(\epsilon_i-\frac{\gamma_i}{\zeta}\right)\right)},
\end{aligned}$$

where in the first equality we have used the definition of $V_i(t)$ from Equation (2.20), and in the second equality, we have used Equation (B.9) and the fact that the Euler equation and

Lemma 3 imply asymptotically constant the real interest rate r^*.

Now, we can use the free entry condition, either in its direct form (2.22), or through Equation (2.21) to find

$$\begin{aligned} r^* &= \lim_{t\to\infty} -\dot{s}_i(t) + \frac{\Pi_i(t)}{V_i(t)}, \\ &= -\gamma_i^S + r^* + g^*\left(\gamma_i - (1-\sigma)\left(\epsilon_i - \frac{\gamma_i}{\zeta}\right)\right), \end{aligned} \tag{B.10}$$

which implies Equation (2.40). □

Proposition. *See Proposition 11.*

Proof. First, it is easy to see that $\gamma_{max} = \zeta\epsilon_{max}$. Note that from the definition (3) and for $S(t)$ defined for $\varsigma > 0$, we have $\lim \dot{s} = \gamma_{max}$ and we find

$$\gamma_i^S - \gamma_i = \delta\gamma_i + (1-\delta)\gamma_{max} - \gamma_i = (1-\delta)(\gamma_{max} - \gamma_i).$$

This means that for $i \in \mathcal{I}^\dagger$, we need to have $\gamma_i = \gamma_{max} = \zeta\epsilon_i$. Now, consider the sector j with $\epsilon_j = \epsilon_{max}$ and, to find a contradiction, assume that $j \notin \mathcal{I}^\dagger$. This implies that $j \notin \mathcal{I}^*$ and therefore $\gamma_j > \zeta\epsilon_{max}$. This is a contradiction, because it implies that $\gamma_j > \gamma_{max}$. Therefore, we find

$$\mathcal{I}^* = \mathcal{I}^\dagger = \{i \,|\, \epsilon_i = \epsilon_{max}\}.$$

From Equation (2.40), we then have

$$(1-\delta)(\zeta\epsilon_{max} - \gamma_i) = (1-\sigma)\left(\frac{\gamma_i}{\zeta} - \epsilon_i\right),$$

implying

$$\begin{aligned} \gamma_i &= \left(1 - \delta + \frac{1-\sigma}{\zeta}\right)^{-1}\left[(1-\delta)\zeta\epsilon_{max} + (1-\sigma)\epsilon_i\right], \\ &= \zeta\left[\frac{1-\sigma}{\zeta(1-\delta)+1-\sigma}\epsilon_i + \frac{\zeta(1-\delta)}{\zeta(1-\delta)+1-\sigma}\epsilon_{max}\right]. \end{aligned}$$

The remainder of the proof follows as a special case of the proof of the Proposition 11 in Section B.2. □

B.2 Generalized Spillover Function

In this section, we examine the properties of our model for a more general class of spillover functions.

Definition 10. Consider spillover functions $\boldsymbol{S}$ defined according to

$$S_i(\boldsymbol{N}) \equiv \frac{1}{\eta_i}\left[\delta_i^{1-\psi_i} N_i^{\psi_i} + (1-\delta_i)^{1-\psi_i}\,\tilde{S}_i(\boldsymbol{N})^{\psi_i}\right]^{\frac{1}{\psi_i}}, \tag{B.11}$$

$$\tilde{S}_i(\boldsymbol{N}) \equiv \left(\sum_j \vartheta_{ij}^{1-\varsigma_i} N_j^{\varsigma_i}\right)^{\frac{1}{\varsigma_i}}, \tag{B.12}$$

for all i, where $\delta \in (0,1)$, $\vartheta_{ij} > 0$ for all $i, j \in \{1, \cdots, I\}$, and $\sum_j \vartheta_{ij} = 1$ for all i. We *rule out* the set of model parameters that for any sector i satisfy $\psi_i \leq 0$ and $\varsigma_i \leq 0$.

According to Equations (B.11) and (B.12), innovation spillovers to each sector have two components: the spillovers from the same sector, the $\delta_i^{1-\psi_i} N_i^{\psi_i}$ term, and the spillovers from all other sectors, the term involving $\tilde{S}_i$. The degree of substitutability between the two types of spillovers is determined by ψ_i. If $\psi_i > 0$, N_i and $\tilde{S}_i$ are gross substitutes and function S_i is log-submodular in the two arguments. If $\psi_i < 0$, N_i and $\tilde{S}_i$ are gross complements and function S_i is log-supermodular in the two arguments. Similarly, parameter ς_i for sector i determines the degree of substitutability among technologies of different sectors in their innovation spillovers to sector i. We will see below that the two parameters (ψ_i, ς_i) play an important role in the characterization of the asymptotic rate of technical growth for sector i. The case of $\psi_i \to 0$, when $\delta_i \equiv \delta$, $\varsigma_i \equiv \varsigma$, and $\vartheta_{ij} \equiv \vartheta_j$ for all $i$and j corresponds to the case of Definition 3, with the stock of general-purpose technologies S defined correspondingly. We rule out the case where both N_t and $\tilde{S}_i$ are gross substitutes and $\tilde{S}_i$ is log-submodular in its arguments, since we show below that this case does not allow for the existence of a CGP.

The following proposition characterizes the long-run behavior of the economy for this class of spillover functions.

Proposition 11. *Consider an economy with the intersectoral innovation spillover function introduced in Example 10 and a set of income elasticity parameters that consists of a set of distinct values, that*

is, $\varepsilon_i \neq \varepsilon_{i'}$ for any pair of $i \neq i'$.

Consider then a potential CGP as one characterized by a set of parameters

$$\left(\mathcal{I}^* \subset \mathcal{I}, \{\Omega_i^*, \Lambda_i^*\}_{i\in\mathcal{I}^*}, g^*, L^*, \boldsymbol{\gamma}, \mathbf{N}^*\right).$$

First, the set of sectors that asymptotically constitute a nonvanishing share of economic activity $\mathcal{I}^*$ consists of

(1) Any sector i with $\varsigma_i > 0$ and $\psi_i < 0$, or $\varsigma_i < 0$ and $\psi_i > 0$,

(2) Any sector i with $\varsigma_i > 0$ and $\psi_i \geq 0$ if $\epsilon_i \geq \epsilon_{i'}$ for all i'.

For all of these sectors, the rate of technological growth is asymptotically given by $\gamma_i = \zeta\epsilon_i$. Any other sector has a vanishing share of employment in production and output, and a vanishing share in total assets. The asymptotic rate of technological growth for such a sector i is given by

$$\gamma_i = \zeta\left(\varrho_i\epsilon_i + (1-\varrho_i)\,\epsilon_{max}\right), \qquad \varrho_i \equiv \begin{cases} \frac{1-\sigma}{\zeta+1-\sigma}, & \psi_i > 0, \\ \frac{1-\sigma}{\zeta(1-\delta_i)+1-\sigma}, & \psi_i \to 0. \end{cases}$$

Denote by $\langle\epsilon_i\rangle^*$ and $Var\,\langle\epsilon_i\rangle^*$ the average and the variance of income elasticity parameters under the sectoral distribution implied by the asymptotic sectoral shares of corporate assets $\{\Lambda_i^*\}_{i\in\mathcal{I}^*}$.

If the discount rate ρ satisfies

$$\frac{1}{\zeta A^*}\left(1 - \frac{\theta + \langle\epsilon_i\rangle^* - 1 + (\zeta+1)\langle\epsilon_i\rangle^*}{1 + (\zeta+1)\langle\epsilon_i\rangle^*}\right) < \rho < \frac{1}{\zeta A^*}, \tag{B.13}$$

the CGP exists and is unique. In this CGP, the asymptotic real interest rate r^* and the consumption expenditure weighted income elasticity parameters $\overline{\epsilon}^*$ will be found by the unique solution to the system of equations

$$\begin{aligned} \overline{\epsilon}^* &= \langle\epsilon_i\rangle^* + \frac{Var\,\langle\epsilon_i\rangle^*}{r^* + \zeta\langle\epsilon_i\rangle^*}, \\ r^* &= \frac{1}{\zeta A^*} - \frac{(1+\zeta)\langle\epsilon_i\rangle^*}{\overline{\epsilon}^* + \theta - 1 + (1+\zeta)\langle\epsilon_i\rangle^*}\left(\frac{1}{\zeta A^*} - \rho\right). \end{aligned}$$

The asymptotic rate of growth in aggregate consumption is given by

$$g^* = \frac{1/\zeta A^* - \rho}{\theta + \overline{\epsilon}^* - 1 + (\zeta+1)\langle\epsilon_i\rangle^*},$$

and the asymptotic share of employment in production by

$$\frac{L^*}{H} = \frac{\rho\zeta A^* \langle\epsilon_i\rangle^* + \theta + \overline{\epsilon}^* - 1 + \zeta \langle\epsilon_i\rangle^*}{\theta + \overline{\epsilon}^* - 1 + (\zeta+1) \langle\epsilon_i\rangle^*}.$$

Finally, $(\Omega_i, N_i^*)_{i\in\mathcal{I}^*}$ *are given by Equations (2.25) and (2.26) and*

$$\Omega_i^* = \frac{\zeta}{L^*}\left(r^* + \zeta g^* \epsilon_i\right)\Lambda_i^*, \qquad \text{for all } i \in \mathcal{I}^*.$$

Proof. Let γ_{min} and γ_{max} denote the minimum and the maximum in the set of all relative rates of technical growth $\{\gamma_i\}_{i=1}^I$, respectively, and sets $\mathcal{I}_{min}$ and $\mathcal{I}_{max}$ to include the sectors that achieve the correspondingly values. Let us further define $\gamma_i^S \equiv \lim_{t\to\infty} \frac{\dot{s}_i(t)}{g^*} = \sum_j \Sigma_{ij}^* \gamma_i$. We can now rewrite Equation (B.10) as

$$\left(1 + \frac{1-\sigma}{\zeta}\right)\gamma_i - \gamma_i^S = (1-\sigma)\,\epsilon_i, \qquad \text{for all } i.$$

From Equation (2.39) we know that $\gamma_i \leq \gamma_i^S$ and therefore $\gamma_i \geq \zeta\epsilon_i$.

Now, we will group sectors based on the degree of substitutability parameters (ψ_i, ς_i) in their respective intersectoral innovation spillover functions.

1. *Sectors with* $\psi_i, \varsigma_i < 0$*:* In this case, S_i asymptotically grows at the rate dictated by the slowest (technologically) growing sectors $\mathcal{I}_{min}$. Therefore, for all such sectors we have $\gamma_i^S = \gamma_{min}$. However, in order to satisfy condition (2.39), we need to ensure that $\gamma_i = \zeta\epsilon_i = \gamma_{min}$. But since $\gamma_i \geq \zeta\epsilon_i$ for all i, this could be the case only if $\epsilon_i = \epsilon_{min}$. We can further compute the asymptotic corporate assets held in sector i as

$$\begin{aligned}
\Lambda_i^* &= \lim_{t\to\infty} \frac{N_i(t)}{S_i(t)}, \\
&= \lim_{t\to\infty} \eta_i \left(\delta_i^{1-\psi_i} + (1-\delta_i)^{1-\psi_i}\left(\sum_{j\neq i} \vartheta_{ij}^{1-\varsigma_i}\left(\frac{N_j(t)}{N_i(t)}\right)^{\varsigma_i}\right)^{\frac{\psi_i}{\varsigma_i}}\right)^{-\frac{1}{\psi_i}}, \\
&= \eta_i \left(\delta_i^{1-\psi_i} + (1-\delta_i)^{1-\psi_i}\left(\sum_{j\neq i:\, \epsilon_j=\epsilon_{min}} \vartheta_{ij}^{1-\varsigma_i}\left(\frac{N_j^*}{N_i^*}\right)^{\varsigma_i}\right)^{\frac{\psi_i}{\varsigma_i}}\right)^{-\frac{1}{\psi_i}}.
\end{aligned}$$

2. *Sectors with* $\psi_i < 0$ *and* $\varsigma_i > 0$*:* In this case S_i asymptotically grows at a rate dictated

by the minimum of the rate of growths of N_i and $\tilde{S}_i$. The spillover term $\tilde{S}_i$ in turn asymptotically grows at the rate dictated by the fastest (technologically) growing sectors γ_{max} (note that condition stated in the proposition rules out the case where all sectors other than i technologically grow at strictly lower rates). Therefore, we always have that $\gamma_i^S = \gamma_i = \zeta\epsilon_i$.

$$\begin{aligned}\Lambda_i^* &= \lim_{t\to\infty} \frac{N_i(t)}{S_i(t)}, \\ &= \eta_i \left(\delta_i^{1-\psi_i} + (1-\delta_i)^{1-\psi_i} \mathbb{I}\{\epsilon_i = \epsilon_{max}\} \left(\sum_{j\neq i:\, \epsilon_j=\epsilon_i} \vartheta_{ij}^{1-\varsigma_i} \left(\frac{N_j^*}{N_i^*} \right)^{\varsigma_i} \right)^{\frac{\psi_i}{\varsigma_i}} \right)^{-\frac{1}{\psi_i}},\end{aligned}$$

where $\mathbb{I}\{\epsilon_i = \epsilon_{max}\} = 1$ only if $\epsilon_i = \epsilon_{max}$.

3. *Sectors with $\psi_i > 0$ and $\varsigma_i < 0$:* In this case S_i asymptotically grows at a rate dictated by the maximum of the rates of growth of N_i and $\tilde{S}_i$. However, the rate of growth of $\tilde{S}_i$ is now determined by the slowest (technologically) growing sector other than i. Therefore, once again we will have $\gamma_i^S = \gamma_i = \zeta\epsilon_i$, unless if $\epsilon_{i'} > \epsilon_i$ for all $i' \neq i$. We will have

$$\begin{aligned}\Lambda_i^* &= \lim_{t\to\infty} \frac{N_i(t)}{S_i(t)}, \\ &= \eta_i \left(\delta_i^{1-\psi_i} + (1-\delta_i)^{1-\psi_i} \mathbb{I}\{\epsilon_i = \epsilon_{min}\} \left(\sum_{j\neq i:\, \epsilon_j=\epsilon_i} \vartheta_{ij}^{1-\varsigma_i} \left(\frac{N_j^*}{N_i^*} \right)^{\varsigma_i} \right)^{\frac{\psi_i}{\varsigma_i}} \right)^{-\frac{1}{\psi_i}}.\end{aligned}$$

If if $\epsilon_{i'} > \epsilon_i$ for all $i' \neq i$, then we have that $\gamma_i^S = \gamma_{min}^{-i} \equiv \min\{\gamma_{i'}\}_{i'\in\mathcal{I}\setminus\{i\}}$.

$$\gamma_i = \zeta\epsilon_i \left(1 + \frac{\zeta}{\zeta + 1 - \sigma} \left(\frac{\gamma_{min}^{-i}}{\zeta\epsilon_i} - 1 \right) \right) > \zeta\epsilon_i,$$

since $\zeta\epsilon_i = \gamma_i = \gamma_{min} < \gamma_{min}^{-i}$.

4. *Sectors with $\psi_i > 0$ and $\varsigma_i > 0$:* In this case S_i asymptotically grows at the rate determined by the fastest (technologically) growing sectors $\mathcal{I}_{max}$, and we have $\gamma_i^S = \gamma_{max}$ for all i. This first suggests that $\gamma_{max} = \zeta\epsilon_{max}$ where ϵ_{max} is the sector with the

largest income elasticity. For this sector

$$\begin{aligned}\Lambda_i^* &= \lim_{t\to\infty}\frac{N_i(t)}{S_i(t)},\\ &= \eta_i\left(\delta_i^{1-\psi_i}+(1-\delta_i)^{1-\psi_i}\mathbb{I}\{\epsilon_i=\epsilon_{max}\}\left(\sum_{j\neq i:\,\epsilon_j=\epsilon_i}\vartheta_{ij}^{1-\varsigma_i}\left(\frac{N_j^*}{N_i^*}\right)^{\varsigma_i}\right)^{\frac{\psi_i}{\varsigma_i}}\right)^{-\frac{1}{\psi_i}}.\end{aligned}$$

For all other sectors follows from Equation (B.10) that

$$\gamma_i=\zeta\epsilon_i\left(1+\frac{\zeta}{\zeta+1-\sigma}\left(\frac{\epsilon_{max}}{\epsilon_i}-1\right)\right).$$

5. *Sectors with $\psi_i\to 0$:* In this case

$$S_i=\frac{1}{\eta_i}N_i^{\delta_i}\tilde{S}_i^{1-\delta_i},$$

and $\dot{s}_i=\delta_i\dot{n}_i+(1-\delta_i)\dot{\tilde{s}}_i$. If $\varsigma_i<0$, again, we can only have $\gamma_i=\zeta\epsilon_i$ with $\epsilon_i=\epsilon_{min}$. If $\varsigma_i>0$, we have

$$\gamma_i=\zeta\epsilon_i\left(1+\frac{\zeta(1-\delta_i)}{\zeta(1-\delta_i)+1-\sigma}\left(\frac{\epsilon_{max}}{\epsilon_i}-1\right)\right).$$

In this case, if ϵ_i is the unique maximum in the set $\{\epsilon_i\}_{i\in\mathcal{I}}$, we do not have a CGP.

Note that in cases 1-3 and the case 4 when $\epsilon_i=\epsilon_{max}$, we have that $\gamma_i^S=\gamma_i$ and therefore $i\in\mathcal{I}^\dagger$ as defined by Lemma 4. Only sectors with $\gamma_i>\zeta\epsilon_i$, that is, sectors with $\psi_i\geq 0$ and $\varsigma_i\geq 0$ are not in $\mathcal{I}^\dagger$ and therefore asymptotically vanish.

From the Euler Equation (2.7), we find that asymptotically

$$g^*=\frac{r^*-\rho}{\theta-1+\overline{\epsilon}^*\left[1+(1-\sigma)\,Var^*\left(\frac{\epsilon_i}{\overline{\epsilon}^*}\right)\right]-\frac{1}{\zeta}\overline{\gamma}^*\left[1+(1-\sigma)\,Cov^*\left(\frac{\epsilon_i}{\overline{\epsilon}^*},\frac{\gamma_i}{\overline{\gamma}^*}\right)\right]}.$$

Let us first compute θ^* defined such that $r^*=g^*\theta^*+\rho$, that is

$$\theta^*\equiv\theta-1+\overline{\epsilon}^*\left[1+(1-\sigma)\,Var^*\left(\frac{\epsilon_i}{\overline{\epsilon}^*}\right)\right]-\frac{1}{\zeta}\overline{\gamma}^*\left[1+(1-\sigma)\,Cov^*\left(\frac{\epsilon_i}{\overline{\epsilon}^*},\frac{\gamma_i}{\overline{\gamma}^*}\right)\right].$$

For all sectors i that are included in the moments that appear in θ^*, we have $\gamma_i=\zeta\epsilon_i$.

Therefore, the covariance term is $Cov^*\left(\frac{\epsilon_i}{\overline{\epsilon}^*}, \frac{\epsilon_i}{\overline{\epsilon}^*}\right) = Var(\frac{\epsilon_i}{\overline{\epsilon}^*}) - 1$.

$$\begin{aligned}\theta^* &= \theta + \overline{\epsilon}^* - 1 + \left(\overline{\epsilon}^* - \frac{1}{\zeta}\overline{\gamma}^*\right)\left[1 + (1-\sigma)\, Var^*\left(\frac{\epsilon_i}{\overline{\epsilon}^*}\right)\right], \\ &= \theta + \overline{\epsilon}^* - 1,\end{aligned}$$

where in the second equality we have used Equation (2.35).

We will now pin down the rate of growth of aggregate real consumption g^*. Substituting from $\gamma_i = \zeta\epsilon_i$ for $i \in \mathcal{I}^*$ in Equation (2.41) we find:

$$\frac{1}{\zeta}\left(H - L^*\right) = g^* A^* \sum_{i\in\mathcal{I}^*} \epsilon_i \Lambda_i^*. \tag{B.14}$$

Then, taking the asymptotic limit of Equation (2.28), we find

$$\left(r^* + \gamma_i^S g^*\right)\Lambda_i^* = \Omega_i^* \frac{L^*}{\zeta A^*}. \tag{B.15}$$

Summing over $i \in \mathcal{I}^*$ for which $\gamma_i^S = \gamma_i = \zeta\epsilon_i$ and substituting for $r^* = \theta^* g^* + \rho$, we reach

$$\rho\left(\sum_{i\in\mathcal{I}^*}\Lambda_i^*\right) + g^*\left[\theta^*\left(\sum_{i\in\mathcal{I}^*}\Lambda_i^*\right) + \zeta\left(\sum_{i\in\mathcal{I}^*}\epsilon_i\Lambda_i^*\right)\right] = \frac{L^*}{\zeta\, A^*}.$$

Summing this expression from Equation (B.14), we find

$$\frac{H}{\zeta A^*} = \rho\left(\sum_{i\in\mathcal{I}^*}\Lambda_i^*\right) + g^*\left[\theta^*\left(\sum_{i\in\mathcal{I}^*}\Lambda_i^*\right) + (\zeta+1)\left(\sum_{i\in\mathcal{I}^*}\epsilon_i\Lambda_i^*\right)\right],$$

which pins down the economy-wide rate of growth as

$$g^* = \frac{1/\zeta A^* - \rho}{\theta^* + (\zeta+1)\left\langle\epsilon_i\right\rangle^*},$$

where $A^* \equiv \frac{1}{H}\sum_{i\in\mathcal{I}^*}\Lambda_i^*$ is the asymptotic value of assets owned by the households and $\left\langle\epsilon_i\right\rangle^* \equiv \frac{\sum_{i\in\mathcal{I}^*}\epsilon_i\Lambda_i^*}{\sum_{i\in\mathcal{I}^*}\Lambda_i^*}$ is the average of income elasticity parameters as weighted by the sector's asymptotic shares of corporate assets. Similarly, we can derive the expression for the asymptotic share of employment in the production sector

$$\frac{L^*}{H} = \frac{\rho\zeta A^* + \theta^*/\left\langle\epsilon_i\right\rangle^* + \zeta}{\theta^*/\left\langle\epsilon_i\right\rangle^* + 1 + \zeta}.$$

Finally, for the real interest rate we find

$$\begin{aligned} r^* &= \rho + \frac{\theta^* / \langle \epsilon_i \rangle^*}{\theta^* / \langle \epsilon_i \rangle^* + 1 + \zeta} \left(\frac{1}{\zeta A^*} - \rho \right), \\ &= \left(\frac{\theta^* / \langle \epsilon_i \rangle^*}{\theta^* / \langle \epsilon_i \rangle^* + 1 + \zeta} \right) \frac{1}{\zeta A^*} + \left(\frac{1 + \zeta}{\theta^* / \langle \epsilon_i \rangle^* + 1 + \zeta} \right) \rho. \end{aligned}$$

The positivity of g^* and the transversality condition $g^* < r^*$ together imply the following constraints on the discount rate ρ

$$\frac{1}{\zeta A^*} \left(1 - \frac{\theta^* + (\zeta + 1) \langle \epsilon_i \rangle^*}{1 + (\zeta + 1) \langle \epsilon_i \rangle^*} \right) < \rho < \frac{1}{\zeta A^*}.$$

Since, as we see below $\overline{\epsilon}^* \geq \langle \epsilon \rangle^*$, Equation (2.43) is stronger than the condition above as a restriction on ρ. Therefore, if condition (2.43) is satisfied the CGP is unique.

Let us now calculate $\overline{\epsilon}^*$. Substituting from Equation (B.15) for Ω_i^* we find

$$\begin{aligned} \overline{\epsilon}^* &= \frac{\sum_{i \in \mathcal{I}^*} (r^* + \zeta \epsilon_i) \Lambda_i^* \epsilon_i}{\sum_{i \in \mathcal{I}^*} (\rho + \theta^* g^* + \zeta \epsilon_i) \Lambda_i^*}, \\ &= \frac{r^* \left(\sum_{i \in \mathcal{I}^*} \Lambda_i^* \epsilon_i \right) + \zeta \left(\sum_{i \in \mathcal{I}^*} \Lambda_i^* \epsilon_i^2 \right)}{r^* \left(\sum_{i \in \mathcal{I}^*} \Lambda_i^* \right) + \zeta \left(\sum_{i \in \mathcal{I}^*} \Lambda_i^* \epsilon_i \right)}, \\ &= \langle \epsilon_i \rangle^* \frac{r^* + \zeta \frac{\langle \epsilon_i^2 \rangle^*}{\langle \epsilon_i \rangle^*}}{r^* + \zeta \langle \epsilon_i \rangle^*}, \\ &= \langle \epsilon_i \rangle^* \left(1 + \frac{\frac{Var \langle \epsilon_i \rangle^*}{\langle \epsilon_i \rangle^*}}{r^* + \zeta \langle \epsilon_i \rangle^*} \right), \end{aligned}$$

where we have defined the variance of the income elasticity parameters under the distribution implied by sectoral shares in corporate assets $Var \langle \epsilon_i \rangle^* \equiv \frac{\sum_{i \in \mathcal{I}^*} \left(\epsilon_i - \langle \epsilon_i \rangle^* \right)^2 \Lambda_i^*}{\sum_{i \in \mathcal{I}^*} \Lambda_i^*}$.

Combining these equations and letting $\varepsilon \equiv \frac{\overline{\epsilon}^*}{\langle \epsilon_i \rangle^*} - 1$ we find a system of two equations

$$\begin{aligned} \varepsilon &= \frac{Var \langle \epsilon_i \rangle^* / \langle \epsilon_i \rangle^*}{r^* + \zeta \langle \epsilon_i \rangle^*}, \\ r^* &= \frac{1}{\zeta A^*} - \frac{1 + \zeta}{\varepsilon + \frac{\theta + \langle \epsilon_i \rangle^* - 1}{\langle \epsilon_i \rangle^*} + 1 + \zeta} \left(\frac{1}{\zeta A^*} - \rho \right), \end{aligned}$$

that together determine (ε, r^*). If condition (2.43) is satisfied, this system of equations has a unique solution. The quadratic equation always has a solution. The first equation defines

r^* as a decreasing function of ε that goes from ∞ to 0 as ε goes from 0 to the maximum of $Var\left\langle \epsilon_i \right\rangle^* / \zeta \left(\left\langle \epsilon_i \right\rangle^* \right)^2$. The second equation defines r^* as an increasing function of ε that goes from a $\frac{1}{\zeta A^*} - \frac{1+\zeta}{\frac{\theta + \left\langle \epsilon_i \right\rangle^* - 1}{\left\langle \epsilon_i \right\rangle^*} + 1 + \zeta} \left(\frac{1}{\zeta A^*} - \rho \right)$ to $\frac{1}{\zeta A^*}$ as ε goes from 0 to ∞.[2] This completes the characterization of the CGP. □

B.3 Schumpeterian Model

Consider the following production function for the final good producers in each sector to replace the production function in Equation (2.10)

$$Y_i = \left(\int_0^1 Q_{iv}^{\frac{1}{\zeta+1}} X_{iv}^{\frac{\zeta}{\zeta+1}} dv \right)^{\frac{\zeta+1}{\zeta}},$$

where each sector has a unit interval of varieties v, each with a variety-specific quality Q_{iv}. The demand for the intermediate good v in sector i is given by

$$X_{iv} = Q_{iv} \left(\frac{P_i}{P_{iv}} \right)^{1+\zeta} Y_i,$$

which again suggests $P_{iv} = 1$ for all varieties in all sectors at all times. The resulting price index for goods in sector i is given by $P_i = Q_i^{-\frac{1}{\zeta}}$, where we have defined Q_i as the average quality of products in sector i, that is, $Q_i \equiv \int_0^1 Q_{iv} dv$. Note that this parallels the result in the expanding varieties model where Q_i, average sectoral quality, has replaced the number of intermediate goods N_i. Similarly, the profits for a given producer is given by

$$\Pi_{iv} = \frac{1}{\zeta} X_{iv} = \frac{1}{\zeta} \frac{Q_{iv}}{Q_i} \Omega_i Y,$$

which again parallels Equation (2.11).

Potential entrants invest in raising the quality of the available intermediate goods through investing in R&D, which targets a specific variety. When a potential entrant hires Z_{iv} R&D workers to improve on the quality of variety v in sector i with current quality Q_{iv},

[2] Simple algebra shows that this system indeed can be reduced to a quadratic equation.

it succeeds at a flow rate

$$\frac{S_i(\boldsymbol{Q})}{\eta_i}\frac{Z_{iv}}{Q_{iv}},$$

in which case the quality of the variety improves to $Q_{iv}\Gamma_i$ and the entrant takes over the current monopolist. We assume that $\Gamma_i \geq \left(\frac{\zeta+1}{\zeta}\right)^{\zeta}$ so that the monopolistic pricing of the current incumbent does not allow the previous incumbent to enter the market. The Bellman equation for the net present value of the monopoly rights on variety v in sector i is given by $rV_{iv} - \dot{V}_{iv} = \Pi_{iv} - \frac{S_i}{\eta_i}\frac{Z_{iv}}{Q_{iv}}V_{iv}$. Due to constant returns to scale, the value function will be linear in quality, that is $V_{iv} = V_i Q_{iv}$ for some V_i. Similary, the rate of creative destruction is going to be the same for all intermediate goods within a sector. Define $Z_i \equiv Q_i\frac{Z_{iv}}{Q_{iv}}$ such that this rate for sector i is $\frac{Z_i}{Q_i}$ and the total number of R&D workers in sector i $\int_0^1 Z_i\frac{Q_{iv}}{Q_i}dv = Z_i$. Since Q_{iv} is a jump process, its derivative is zero and we have $\left(r + \frac{S_i}{\eta_i}\frac{Z_i}{Q_i}\right)V_i - \dot{V}_i = \frac{1}{\zeta}\frac{\Omega_i Y}{Q_i}$ where Z_{iv} is the number of R&D workers hired for any intermediate variety in sector i. The free entry condition is now given by

$$1 \geq \frac{S_i}{\eta_i/\Gamma_i}V_i.$$

The rate of growth of quality in sector i is $\dot{q}_i = \frac{\Gamma_i - 1}{\eta_i}\frac{S_i}{Q_i}Z_i$. [3] Therefore, we can rewrite the free entry condition as

$$\dot{q}_i = \left(1 - \frac{1}{\Gamma_i}\right)\frac{Z_i}{Q_i V_i}.$$

In Equation (2.23), we instead have $H \cdot A = \sum_i \int V_{iv}dv = \sum_i Q_i V_i$.

[3] To see this, note that

$$\begin{aligned} Q_i(t+dt) &= \left(\int_0^1 (\Gamma_i - 1)Q_{iv} \times \frac{S_i}{\eta_i}\frac{Z_i}{Q_i}dv\right)dt, \\ &= \frac{\Gamma_i - 1}{\eta_i}S_i Z_i dt, \end{aligned}$$

where in the second line we have used the fact that R&D employment is the same for all varieties within the same sector.

Appendix C

Appendix to Chapter 3

C.1 Facts

In this section, I provide an overview of the set of facts on the distribution and dynamics of firm size and productivity that the model aims to explain. I also discuss how the existing theories of firm dynamics generate these facts.

1. Firm size and productivity distributions exhibit Pareto tail behavior In the US data, the logarithm of the complementary cumulative distribution function (tail distribution) of firm size is linear in the logarithm of firm size in its right tail (Axtell, 2001; Luttmer, 2010).[1] There is also evidence that the distribution of total factor productivity of firms across different countries has a Pareto tail (Aoyama *et al.*, 2010; Mizuno *et al.*, 2012).[2] Among the theories of firm heterogeneity and dynamics that endogenize the firm size distribution, the models of Luttmer (2007), Luttmer (2011), Arkolakis (2016), and the current paper match with this fact. In contrast, the benchmark model of Klette and Kortum (2004) and its empirical extension by Lentz and Mortensen (2008) both yield considerably thinner right

[1] Geerolf (2014) provides evidence for the Pareto tail in the distribution of firm size in France and constructs an alternative, static, matching-based model that can rationalize this phenomenon.

[2] This is indeed a standard assumption in many models of firm heterogeneity and trade (see, e.g., Helpman *et al.*, 2004; Chaney, 2008).

tails. Luttmer (2010) gives an overview of different approaches and necessary assumptions for generating a Pareto tail of firm size.

2. Size and productivity are correlated Following Olley and Pakes (1996), a large body of empirical work has utilized the covariance of size and productivity as a measure of aggregate industry productivity (see Bartelsman *et al.*, 2013, for cross-country evidence). Under the baseline of the workhorse innovation model of Klette and Kortum (2004), firm size and productivity are independent of each other due to the undirectedness of the innovation process. Therefore, we need to introduce further heterogeneity among firms in terms of the quality of their innovations (e.g., Lentz and Mortensen, 2008). In contrast, the productivity-based class of models readily explains the correlation between size/productivity, since the variations in size derive from productivity heterogeneity in the presence of decreasing returns to scale or imperfect competition.

3. Growth is negatively correlated with size for small firms and uncorrelated with size for large firms Based on early firm-level data, Gibrat and subsequent researchers found that firm growth does not depend on size (e.g., Hart and Prais, 1956; Simon and Bonini, 1958). Using US manufacturing data, Hall (1987) and Evans (1987) revisited Gibrat's law accounting for the role of selection and measurement error, and concluded that the relationship is negative for small firms, but approximately holds for larger firms.[3] Models of firm dynamics match relying on different mechanisms to explain this fact. In the current paper, the faster growth of smaller firms is driven by stronger selection.[4] Other recent papers such as Akcigit and Kerr (2012) and Acemoglu *et al.* (2014) deliver this prediction by introducing further heterogeneity in the types and quality of innovations across firms.

[3] Davis *et al.* (1996) used US census data to make the case that controlling for age, this relationship does not hold even among smaller firms (see also Haltiwanger *et al.*, 2013; Decker *et al.*, 2014; Haltiwanger, 2014). However, debate around this point still continues, as some recent studies challenge the conclusions of Davis *et al.* (1996) based on other datasets and different empirical specifications (Neumark *et al.*, 2011; Huber *et al.*, 2013). See also the discussions about the details of these specifications in (Akcigit and Kerr, 2012).

[4] An example of an alternative approach is the model by Jovanovic and MacDonald (1994a), in which all firms look within the same pool of available ideas. Firms that are ahead have further exhausted the space of potential new ideas and have less to gain from their investments.

4. The extensive margin of innovation intensity is correlated with productivity and size The evidence on this point mainly comes from the early studies of R&D, size, and productivity (see Griliches, 2000; Hall, 1996, for surveys of this literature). The earlier evidence based on manufacturing firms generally emphasized the correlation between the extensive margin of innovation investments (e.g., whether or not firms report any R&D expenditures) and productivity or size (e.g., Cohen and Klepper, 1996). In the US retail sector, Doms *et al.* (2004) find a pronounced correlation between firm size and productivity and intensive measures of investment in IT. In the presence of very small fixed costs of innovation, the model in this paper delivers this fact through the linkage between selection and innovation incentives. This is because larger and more productive firms have higher incentives for innovation as they have a longer expected lifetime in the industry. However, focusing on R&D expenditure and patents as measures of innovation investments and outputs, respectively, Akcigit and Kerr (2012) find a negative relationship between innovation intensity and size among a subsample of manufacturing firms that they refer to as "innovative firms," i.e., those that do report R&D expenditures over an extended period of time and also file for patents. In explaining the extensive margin of innovation and considering a broader concept of innovation activity, the current paper could be complementary to more refined theories of firm-level R&D and patenting behavior (Akcigit and Kerr, 2012).

5. Likelihood of exit is negatively correlated with productivity and size The negative relationship between size and likelihood of exit is nearly universal. Foster *et al.* (2001) summarize the evidence that shows that low productivity predicts exit even controlling for a variety of different firm characteristics. The current model explains this fact through the assumption on the fixed costs, which gives rise to the endogenous exit of low-productivity firms. In the workhorse model of Klette and Kortum (2004), since size and productivity are independent we do not necessarily find a correlation between productivity and size. Once again, introducing additional heterogeneity in terms of innovation quality within this framework can produce the correlation between size and productivity in the cross section

(e.g., Lentz and Mortensen, 2008). More recent work has combined competition through the input markets (as in the current paper) with the heterogeneity in innovative types within the workhorse framework (Acemoglu *et al.*, 2013).

6. Entrants are smaller, less productive, and less dispersed than incumbents The evidence is widespread in studies of comprehensive firm-level data from industries across the world. Seminal studies reporting the evidence include, among others, Dunne *et al.* (1988) and Disney *et al.* (2003).

C.2 Derivations and Proofs

C.2.1 Section 3.2

Dynamic Industry Equilibrium Let $v_t(\theta)$ denote the value of a firm with productivity θ at time t. Due to the presence of fixed costs, firms endogenously exit the industry once their productivity falls below an exit threshold. Let $\theta_{o,t}$ denote the exit threshold of incumbent firms at time t. The value function then satisfies a Hamilton-Jacobi-Bellman (HJB) partial differential equation for all productivity states $\theta > \theta_{o,t}$ above this threshold (see Equation (C.3) below). The solution to the HJB equation yields the value function $[v_t(\cdot)]_t$ as well as an incumbent firm's path of optimal innovation and exit policies $\varsigma^* \equiv [\gamma_t^*(\cdot), \theta_{o,t}]_{t=0}^{\infty}$.

Define the cumulative distribution of entrant productivity (before entry) at time t as

$$F_{e,t}^{Pre}(\theta) \equiv F_a\left[M_t\left(\left(e^{-\gamma_e}\theta_{o,t}\tilde{\theta}_t^{\eta}\right)^{1/(1-\eta)}\right)/\overline{M}_t\right].$$

Let $\lambda_{e,t} \geq 0$ denote the rate of entry at time t, defined as the flow of entrants divided by the total mass of currently active firms in the industry. Standard arguments show that along any equilibrium path, the following zero profit condition holds

$$c_{e,t} \geq v_{e,t}, \text{ and } \lambda_{e,t}(c_{e,t} - v_{e,t}) = 0, \qquad \forall t \geq 0, \tag{C.1}$$

where the costs of entry are given by $c_{e,t} \equiv \psi_e w_t$ and the expected net presented value $v_{e,t}$ by

$$v_{e,t} \equiv \int v_t(\theta)\, dF_{e,t}(\theta).$$

Now consider such an optimal policy ς^* and define the mass density of firms as $m_t(\theta) = \frac{\partial}{\partial\theta} M_t(\theta)$. We can show that the industry measure has to satisfy a Kolmogorov Forward Equation (KFE) (see Equation (C.5) in Appendix C.2.1). Along with boundary conditions $M_t(\theta_{o,t}) = 0$ and $\overline{M}_t < \infty$ (for all t), this partial differential equation characterizes the path of industry measure $[M_t(\cdot)]_t$ and the rate of turnover (exit) $\lambda_{o,t}$ satisfying

$$\overline{M}_t \lambda_{o,t} \equiv \frac{1}{2}\sigma^2 \frac{\partial}{\partial\theta}\left(\theta^2 m_t(\theta)\right)\Big|_{\theta=\theta_{o,t}}, \tag{C.2}$$

for a given initial industry measure M_o. Note that the rate of growth of the total mass of firms in the industry is given by the difference between rate of (successful) entry $\lambda_{e,t}$ and exit $\lambda_{o,t}$, such that $\dot{\overline{M}}_t/\overline{M}_t = \lambda_{e,t} - \lambda_{o,t}$. If $\gamma_e > -\eta \log\left(\tilde{\theta}_t/\theta_{o,t}\right)$, then entrants do not exit immediately, the gross rate of entry equals the successful rate of entry $\lambda_{o,t}$, and the post-entry distribution of entrants is the same as the pre-entry one: $F_{e,t}(\theta) \equiv F^{Pre}_{e,t}(\theta)$. Otherwise, the gross rate of entry is greater than the rate of successful entry by a factor of $\left[1 - F^{Pre}_{e,t}(\theta_{o,t})\right]^{-1}$. In this case, the post-entry distribution of entrants is given by

$$F_{e,t}(\theta) \equiv \frac{F^{Pre}_{e,t}(\theta) - F^{Pre}_{e,t}(\theta_{o,t})}{1 - F^{Pre}_{e,t}(\theta_{o,t})}.$$

In the dynamic industry equilibrium, firms take the future paths of wages, interest rate, and aggregate sales as given, rationally expect the behavior of other firms, and decide on their optimal strategies.[5] The aggregation of all these decisions generates the path of the industry measure $[M_t(\cdot)]_{t=0}^{\infty}$ that contains all the information regarding the state of the industry.

Definition 12. (Industry Equilibrium) Given an initial industry measure M_0, a time path

[5]Due to the assumption of a continuum of firms, there is no aggregate uncertainty in this model. If we allow for a finite number of firms, then aggregate uncertainty emerges, following the arguments presented by Gabaix (2011) and Carvalho and Grassi (2015).

of wages $[w_t]_{t=0}^{\infty}$, interest rate r , and a time path of industry aggregate sales $[N_t]_{t=0}^{\infty}$, an industry tuple $\mathcal{I} \equiv [M_t(\cdot), \gamma_t^*(\cdot), \theta_{o,t}, \lambda_{e,t}]_{t=0}^{\infty}$ characterizes a dynamic industry equilibrium if

1. The policy function $\varsigma \equiv [\gamma_t^*(\cdot), \theta_{o,t}]_{t=0}^{\infty}$ constitutes a solution to the HJB equation (C.3) (see the derivation below),

2. Rate of entry $\lambda_{e,t}$ satisfies (C.1) for the corresponding value function,

3. The measure $M_t(\cdot)$ satisfies the KFE (C.5) (see the derivation below) with the corresponding boundary conditions.

General Equilibrium I introduce the market clearing conditions and use Definition 12 for the industry equilibrium to define the general equilibrium of this economy.

Definition 13. (General Equilibrium) Consider an industry tuple $\mathcal{I}$ and a path of population $[N_t]_{t=0}^{\infty}$ such that

$$\frac{N_t}{\overline{M}_t} > \psi_f (1 + \psi \lambda_{e,t}),$$

for all t. A path of wages, consumption, and per-capita assets $[w_t, q_t, a_t]_{t=0}^{\infty}$ and the industry $\mathcal{I}$ together form a dynamic general equilibrium if we have

1. Given the path of wages $[w_t]_{t=0}^{\infty}$, interest rate $r_t = r$, and aggregate sales (and population) N_t, the industry constitutes a dynamic industry equilibrium,

2. The paths of wages, per-capita consumption, and per-capita assets satisfy

$$\begin{aligned} w_t &= \frac{\rho}{\rho+1} \frac{N_t}{N_t - \overline{M}_t \psi_f (1 + \psi \lambda_{e,t})}, \\ q_t &= \frac{\rho}{\rho+1} \frac{\overline{\theta}_t}{w_t}, \\ a_t &= \frac{1}{N_t} \int dM_t(\theta) \left[v_t(\theta) + \psi_e w_t \lambda_{e,t} \right], \end{aligned}$$

for all t, where $\overline{\theta}_t$ satisfies (3.7), and a_t defined above satisfies the transversality condition.

Derivation of the HJB and the KFE Equations The value function for a firm with productivity θ at time t has to satisfy the Bellman equation

$$r\, v_t(\theta)\, dt = \max_{\varsigma} \left[\pi_t(\theta, \gamma; M_t) - \psi_f w_t\right] dt + \mathbb{E}_t\left[dv_t(\theta)\right],$$

where π_t denotes the instantaneous profit function as in Equation (3.4) and $\varsigma \equiv (\gamma_t(\cdot), \theta_{o,t})$ is a tuple of firm policy choice. Using Itô's lemma, we find

$$r\, v_t(\theta) - \frac{\partial v_t(\theta)}{\partial t} = \max_{\gamma, \theta_o} \pi_t(\theta, \gamma) - \psi_f w_t + \gamma_\theta \theta \frac{\partial v_t(\theta)}{\partial \theta} + \frac{\sigma^2}{2}\theta^2 \frac{\partial^2 v_t(\theta)}{\partial \theta^2}. \tag{C.3}$$

In addition, we have the standard value matching and smooth pasting boundary conditions $v_t(\theta_{o,t}) = \frac{\partial}{\partial \theta} v_t(\theta_{o,t})$ at the exit threshold $\theta_{o,t}$.

Next, I derive the evolution of the industry measure. Let $M_t^b(\theta)$ denote the mass of firms with profitability states below θ at time t that belong to cohort b, and $m_t^b(\cdot)$ denote the corresponding density mass function. For a given choice of optimal firm policies $\varsigma = (\gamma_t(\cdot), \theta_{o,t})$, the KFE characterizes the evolution of the measure over time as

$$\frac{\partial m_t^b(\theta)}{\partial t} = -\frac{\partial}{\partial \theta}\left[\gamma_t^*(\theta)\ \theta\, m_t^b(\theta)\right] + \frac{\sigma^2}{2}\frac{\partial^2}{\partial^2 \theta}\left[\theta^2 m_t^b(\theta)\right],$$

subject to the boundary conditions $\lim_{t \downarrow b} M_t^b(\theta) = M_t^t(\theta) = \overline{M}_t \lambda_{e,t} F_{e,t}(\theta)$ and $m_t^b(\theta_{o,t}) = 0$, $\forall t, \forall b,\ t > b$.

We can now integrate the mass over all cohorts to find the industry-wide distribution of firm efficiencies

$$M_t(\theta) = \int_{-\infty}^{t} M_t^b(\theta)\, db, \qquad m_t(\theta) = \int_{-\infty}^{t} m_t^b(\theta)\, db.$$

Applying the same integration to the KFE above

$$\begin{aligned} \frac{\partial m_t(\theta)}{\partial t} &= m_t^t(\theta) + \int_{-\infty}^{t} \left\{ -\frac{\partial}{\partial \theta}\left[\gamma_t^*(\theta)\ \theta\, m_t^b(\theta)\right] + \frac{\sigma^2}{2}\frac{\partial^2}{\partial^2 \theta}\left[\theta^2 m_t^b(\theta)\right] \right\} db, \\ &= \overline{M}_t \lambda_{e,t} f_{e,t}(\theta) - \frac{\partial}{\partial \theta}\left[\gamma_t(\theta)\ \theta m_t(\theta)\right] + \frac{\sigma^2}{2}\frac{\partial^2}{\partial^2 \theta}\left[\theta^2 m_t(\theta)\right], \end{aligned} \tag{C.4}$$

where I passed the integral through the derivative in the second line. This leads us to the

following partial differential equation for the industry measure

$$\frac{\partial M_t(\theta)}{\partial t} = \overline{M}_t \left[\lambda_{e,t} F_{e,t}(\theta) - x_t\right] - \gamma_t^*(\theta)\, \theta m_t(\theta) + \frac{\sigma^2}{2}\frac{\partial}{\partial \theta}\left[\theta^2 m_t(\theta)\right], \tag{C.5}$$

with boundary conditions $M_t(\theta_{o,t}) = m_t(\theta_{o,t}) = 0$.

To derive the evolution of the total mass of firms, we integrate the equation above to find

$$\dot{M}_t(\theta) = \overline{M}_t \lambda_{e,t} F_{e,t}(\theta) - \left[\gamma_t(\theta)\, \theta m_t(\theta)\right]_{\theta t}^{\theta} + \frac{\sigma^2}{2}\frac{\partial}{\partial \theta}\left[\theta^2 m_t(\theta)\right]_{\theta_{o,t}}^{\theta},$$

which implies

$$\dot{\overline{M}}_t = \overline{M}_t \lambda_{e,t} - \frac{\sigma^2}{2}\frac{\partial^2}{\partial^2 \theta}\left[\theta^2 m_t(\theta)\right]\Big|_{\theta=\theta_{o,t}}.$$

Derivations of the SCGP HJB and KFE Equations I will first show that firm policy functions are stationary functions of z_t along a stationary constant growth path. This will establish that the mapping between firm employment and efficiency z_t remains constant over time. Therefore, the search for a stationary distribution of firm employment corresponds to a search for a stationary distribution of efficiency z_t. I then characterize the necessary conditions that this stationary distribution needs to satisfy.

Let us begin by deriving the value function as a function of the efficiency state. Define function $V_t(z)$ such that

$$v_t(\theta) = \psi_f w_t V_t\left(\log\left(\frac{1}{(1+\rho)\,\psi_f w_t}\frac{N_t}{\overline{M}_t}\left(\frac{\theta}{\overline{\theta}_t}\right)^{\rho}\right)\right).$$

Taking the derivative of this equation with respect to time, we find

$$\frac{1}{\psi_f w_t}\frac{\partial v_t(\theta)}{\partial t} = \frac{\dot{w}_t}{w_t} V_t + \frac{\partial V_t}{\partial t} + \left[\frac{\dot{N}_t}{N_t} - \frac{\dot{\overline{M}}_t}{\overline{M}_t} - \frac{\dot{w}_t}{w_t} - \rho\frac{\dot{\overline{\theta}}_t}{\overline{\theta}_t}\right]\frac{\partial V_t}{\partial z}.$$

Similarly, the first and second partial derivative with respect to θ yield:

$$\begin{aligned}
\frac{1}{\psi_f w_t}\frac{\partial v_t(\theta)}{\partial \theta} &= \frac{\rho}{\theta}\frac{\partial V}{\partial z}, \\
\frac{1}{\psi_f w_t}\frac{\partial^2 v_t(\theta)}{\partial^2 \theta} &= \left(\frac{\rho}{\theta}\right)^2\frac{\partial^2 V}{\partial z^2} - \frac{\rho}{\theta^2}\frac{\partial V}{\partial z}.
\end{aligned}$$

Substituting for these expressions in the HJB equation (C.3) yields:

$$\left(r-\frac{\dot{w}_t}{w_t}\right)V_t(z)-\frac{\partial V_t}{\partial t} = \max_{\gamma} e^z J(\gamma)-1+\left[\rho\left(\gamma-\frac{1}{2}\sigma_\theta^2-\frac{\dot{\bar{\theta}}_t}{\bar{\theta}_t}\right)-\frac{\dot{w}_t}{w_t}+g_N-\frac{\dot{\overline{M}}_t}{\overline{M}_t}\right]\frac{\partial V_t}{\partial z}+\frac{\rho^2}{2}\sigma^2\frac{\partial^2 V_t}{\partial z^2}.$$

Along as a stationary constant growth path, in addition to the conditions in Definition 5, we need the nominal wage rate to converge to a constant w , and hence, $\dot{w}_t/w_t \to 0$. In addition, let us assume for now that $\dot{\overline{M}}_t/\overline{M}_t \to g_N$. We then find

$$r\,V(z) = \max_{\gamma} e^z J(\gamma)-1+\rho\left(\gamma-\frac{1}{2}\sigma^2-g_\theta\right)\frac{\partial V}{\partial z}+\frac{\rho^2}{2}\sigma^2\frac{\partial^2 V}{\partial z^2},$$

which implies that the optimal investment strategy is also a stationary function of the efficiency state

$$\begin{aligned}\gamma_t(\theta) &= \gamma^*(z_t(\theta)),\\ z_t(\theta_{o,t}) &= z_o.\end{aligned}$$

The derivations above suggest that along a stationary constant growth path, we can consider the following form for the industry measure

$$\begin{aligned}M_t(\theta) &= \overline{M}e^{g_M t}H(z_t(\theta)),\\ m_t(\theta) &= \overline{M}e^{g_M t}\frac{\rho}{\theta}h(z_t(\theta)).\end{aligned}$$

Once again, we calculate the derivatives:

$$\frac{\partial M_t}{\partial t} = g_M\,\overline{M}e^{g_M t}H+(g_N-g_M-\rho g_\theta)\,\overline{M}e^{g_M t}h.$$

and

$$\begin{aligned}\frac{\partial}{\partial\theta}\left[\theta^2 m_t(\theta)\right] &= \rho\overline{M}e^{g_M t}\frac{\partial}{\partial\theta}\left[\theta h(z_t(\theta))\right],\\ &= \rho\overline{M}e^{g_M t}\left[h(z)+\rho\,h'(z)\right].\end{aligned}$$

Substituting these expressions in the Kolmogorov forward equation (C.5) and dividing both

sides by $\overline{M}_t = \overline{M}e^{g_{Mt}}$, we find

$$g_M H = \lambda_e F_e(z) - x - \left(\rho\left(\gamma^*(z) - \frac{1}{2}\sigma^2 - g_\theta\right) + g_N - g_M\right) h(z) + \frac{\rho^2}{2}\sigma^2 h'(z),$$

where $\lambda_o \equiv \frac{\rho^2}{2}\sigma^2 h'_+(z_o)$. If an equilibrium with stationary employment exists, then the industry measure has to be of the form above where H is the solution to this boundary value problem.

Now, we can see why condition $g_M = g_N$ is necessary for the existence of a stationary constant growth path. We have

$$\begin{aligned}
\overline{\theta}_t^\rho &= \frac{1}{\overline{M}_t}\int \theta^\rho J(\gamma_t(\theta))\, dM_t(\theta), \qquad \text{(C.6)}\\
&= \int h(z_t(\theta))\, \theta^\rho J(\gamma^*(z_t(\theta)))\, \frac{\rho d\theta}{\theta},\\
&= \left((1+\rho) w \psi_f \frac{\overline{M}_t}{N_t}\overline{\theta}_t^\rho\right)\int h(z) e^z J(\gamma^*(z))\, dz,
\end{aligned}$$

where I have used the fact that $dz = \rho d\theta/\theta$ and $\theta^\rho = (1+\rho) w\psi_f \frac{\overline{M}_t}{N_t}\overline{\theta}_t^\rho e^z$. This implies

$$\frac{N_t}{\overline{M}_t} = (1+\rho) w\psi_f \int h(z) e^z J(\gamma^*(z))\, dz.$$

Finally, consider the industry-wide stock of knowledge defined by

$$\tilde{\theta}_t^\vartheta = \frac{1}{\overline{M}_t}\int \theta^\vartheta\, dM_t(\theta).$$

Along the stationary constant growth path, we can write this as

$$\begin{aligned}
\tilde{\theta}_t^\vartheta &= \int h(z_t(\theta))\, \theta^\vartheta \frac{\rho d\theta}{\theta},\\
&= \left((1+\rho) w \psi_f \frac{\overline{M}_t}{N_t}\right)^{\frac{\vartheta}{\rho}} \overline{\theta}_t^\vartheta \int h(z)\, e^{\vartheta z/\rho} dz.
\end{aligned}$$

We then have

$$\begin{aligned} e^{z_t(\theta_e)} &= \frac{1}{(1+\rho)\, w\psi_f} \frac{N_t}{\overline{M}_t} \left[\frac{\tilde{\theta}_t^{\eta} \theta_a^{1-\eta} e^{\gamma_e}}{\bar{\theta}_t} \right]^{\rho}, \\ &= \left[\frac{1}{(1+\rho)\, w\psi_f} \frac{N_t}{\overline{M}_t} \left(\frac{\tilde{\theta}_t}{\bar{\theta}_t} \right)^{\rho} \right]^{\eta} \times \left[\frac{1}{(1+\rho)\, w\psi_f} \frac{N_t}{\overline{M}_t} \left(\frac{\theta_a}{\bar{\theta}_t} \right)^{\rho} \right]^{1-\eta} \times e^{\gamma_e}, \\ &= e^{\eta\, \tilde{z} + (1-\eta) z_t(\theta_a) + \rho\gamma_e}, \end{aligned}$$

where, in the last line, I have defined

$$e^{\tilde{z}} \equiv \left(\int h(z)\; e^{\vartheta z/\rho} dz \right)^{\frac{\rho}{\vartheta}}.$$

Proofs of Propositions First, I will provide a generalization of Proposition 6 for the case in which the evolution of efficiencies is given by Equation (3.15) along with an additional exogenous death that arrives according to a Poisson process with rate δ.

Proposition. *See Proposition 6.*

Proof. I map the problem into a two-dimensional first order differential equation. Use the definitions (3.27) and (3.28) and note that

$$\begin{aligned} V''(z)\, e^{-z} &= u_2'(z) + u_2(z), \\ u_1'(z) &= u_2(z) - u_1(z). \end{aligned}$$

Now, we can rewrite Equation (3.17) as

$$(r+\delta)\, u_1(z) = \max_{\gamma} J(\gamma) + \left(\rho\gamma + \frac{\rho^2}{2}\sigma^2 - g \right) u_2(z) + \frac{\rho^2}{2}\sigma^2 u_2'(z),$$

which gives the FOC for the choice of firm growth

$$-J'(\gamma^*(z)) = u_2(z). \tag{C.7}$$

Therefore, we can write the differential equation as $\frac{d}{dz}\boldsymbol{u}(z) = \boldsymbol{Q}(\boldsymbol{u}(z), z)$ such that $\boldsymbol{u} \equiv$

$(u_1, u_2)^t$ and $\boldsymbol{Q}$ is given by

$$u_1' = Q_1(\boldsymbol{u}) = u_2 - u_1, \tag{C.8}$$

$$u_2' = Q_2(\boldsymbol{u}) = \frac{2}{\sigma^2}\left[(r+\delta)u_1 - J(\gamma^*(z)) - \left(\rho\gamma^*(z) - g + \frac{\rho^2}{2}\sigma^2\right)u_2\right], \tag{C.9}$$

where γ^* is a function of $u_2(z)$ as above. This is a boundary value problem with boundary condition $u_2(z_o) = 0$, which ensures $V'(z_o) = 0$. To find the other boundary conditions, we need to first study the asymptotic behavior of the problem and the existence of a stable equilibrium point for this system of differential equations.

What are the necessary conditions for the existence of an economically meaningful solution to the problem? Let us examine the behavior of the largest firms in the industry. Since the likelihood of exit and the option value of exit shrink to zero for the largest firms, we expect the value function to be linear in e^z for these firms. For a solution satisfying these limiting conditions, we need to ensuare that

$$\lim_{z\to\infty} u_1(z) = \lim_{z\to\infty} u_2(z) = u^*. \tag{C.10}$$

This implies $\lim_{z\to\infty}\gamma^*(z) = \gamma^{**}$ where $J'(\gamma^{**}) = -u^*$. Since $J(\cdot)$ is concave, this uniquely pins down the investment of the largest firms as a function of the asymptotical marginal value u^*. Equation (C.9) and $\lim_{z\to\infty} u_2'(z) = 0$ and then imply:

$$\left[r + \delta + g - \left(\rho\gamma^{**} + \frac{\rho^2}{2}\sigma^2\right)\right]u^* - J(\gamma^{**}) = 0.$$

Together, these conditions give

$$u^* = = \frac{J(\gamma^{**})}{r+\delta+g-\left(\rho\gamma^{**}+\frac{\rho^2}{2}\sigma^2\right)} = -J'(\gamma^{**}),$$

where the second equality follows from the first order condition corresponding to the firm's choice of process innovation rate. We can simplify this relation to

$$-\frac{J'(\gamma^{**})}{J(\gamma^{**})} = \frac{1}{\varpi - \rho\gamma^{**}}, \tag{C.11}$$

where I have defined $\varpi \equiv r + \delta + g - \frac{1}{2}\rho^2\sigma^2$.

To establish conditions that ensure this equation has a solution, note that the function $\mathsf{A}(\gamma) = J(\gamma) + (\varpi - \rho\gamma) J'(\gamma)$ is monotonically decreasing. Since we have $\mathsf{A}(0) = J(0) = 1$, the condition for there to be a unique equilibrium is that $\mathsf{A}(\overline{\gamma}) < 0$, which yields condition

$$\varpi - \rho\hat{\gamma} > -\frac{J(\overline{\gamma})}{J'(\overline{\gamma})},$$

which gives us condition (3.18). Therefore, a unique γ^{**} exists.

Now, in addition to condition $u_2(z_o) = 0$, we have the boundary conditions for large firms in Equation (C.10). Together, these three boundary conditions would allow us to fully characterize the solution to the 2nd order differential equation with a free boundary value z_o. Note that since we established the existence of γ^{**}, we have shown that the differential equation has a fixed point.

Next, we should check the stability of the fixed point. Assume that a solution exists and that around the equilibrium point $(u_1, u_2)^t = (u^*, u^*)^t$, it behaves as follows

$$u_1(z) = u^* + Be^{-\beta z}, \qquad u_2(z) = u^* - (\beta - 1) Be^{-\beta z}, \qquad \text{for large } z, \tag{C.12}$$

which automatically satisfies (C.8). Letting $\gamma^*(z) = \gamma^{**} + \Delta\gamma(z)$ where the second term is small for large z, from Equation (C.7) we find $\Delta\gamma(z) = \frac{\beta-1}{J''(\gamma^{**})} Be^{-\beta z}$. The behavior of Equation (C.9) around the equilibrium point can be expressed as

$$u_2'(z) \approx \frac{2}{\rho^2\sigma^2}\left[r + \delta - \left(\rho\gamma^{**} - g + \frac{1}{2}\rho^2\sigma^2\right)(\beta - 1)\right] Be^{-\beta z}. \tag{C.13}$$

We find the following quadratic equation that characterizes β

$$\frac{1}{2}\rho^2\sigma^2\beta(\beta - 1) + \left(\rho\gamma^{**} - g + \frac{1}{2}\sigma^2\right)(\beta - 1) - (r + \delta) = 0.$$

The condition for uniqueness is that this equation should have one positive and one strictly negative root, which is ensured by $r + g + \delta > \rho\gamma^{**} + \frac{1}{2}\rho^2\sigma^2$.[6] Since $\gamma^{**} < \overline{\gamma}$, this condition

[6] Another way to see this is to note that we can simply linearize the system around the infinite limit of z. The linearized equation near $(u_1, u_2)^t = (u^*, u^*)^t$ is given by $\boldsymbol{u}'(z) = \Delta\boldsymbol{Q}\,\boldsymbol{u}(z)$ with $\Delta\boldsymbol{Q} \equiv \begin{bmatrix} -1 & 1 \\ Q_{21} & Q_{22} \end{bmatrix}$, where $Q_{21} = \frac{2(r+\delta)}{\sigma^2}$, $Q_{22} = -\frac{2(\rho\gamma^* - g)}{\rho^2\sigma^2} - 1$. If the determinant $-(Q_{22} + Q_{11}) < 0$ is negative, there will be two eigenvalues one positive and one negative. The condition for the determinant to be negative is

is subsumed by Equation (3.18). This establishes the existence of a stable, fixed point for the system of differential equations.

To derive the approximation in Section **??**, simply extend the form of the solution given by Equations C.12. The smooth pasting condition requires $u_2(z_o) = 0$ and we find $u^* = (\beta - 1) Be^{-\beta z_o}$, while the value matching condition gives us $u_1(z_o) = \frac{1}{r} = u^* \left(1 + \frac{1}{\beta - 1}\right) e^{z_o}$. □

Proposition. *See Proposition 7.*

Proof. I will peform the derivations under the assumption that γ_e is large enough so that all entry attempts are successful. The derivation in the other case is also very similar.

As with the strategy used in the case of the HJB equation, I begin by mapping the 2nd order differential equation into a first-order one defined in the two-dimensional space of $\boldsymbol{h}(z) = (h_1(z), h_2(z))' \equiv (H(z), h(z))'$. We can rewrite the stationary KFE (3.20) as

$$h_1'(z) = \boldsymbol{\mathcal{P}}_1(\boldsymbol{h}) = h_2, \tag{C.14}$$

$$h_2'(z) = \boldsymbol{\mathcal{P}}_2(\boldsymbol{h}, z) = \frac{2}{\rho^2\sigma^2}\left[\lambda_o - \lambda_e F_a\left(\frac{h_1(z) - \eta\tilde{z} - \rho\gamma_e}{1-\eta}\right) + (g_M + \delta) h_1(z) + (\rho\gamma^*(z) - g) h_2(z)\right] \tag{C.15}$$

If a stationary distribution exists, the system of differential equations above along with boundary conditions $h_1(z_o) = h_2(z_o) = 0$ and $\lim_{z\to\infty} h_1(z) = 1$ together pin down the solution and the unknown parameter λ_o.

Once again, to ensure that such a solution exists and is unique, we first examine the behavior of the system of differential equations around its equilibrium point $\boldsymbol{h} = (1, 0)^t$ where $\frac{d}{dz}\boldsymbol{h}(z) = 0$. I linearize the system around this point to find

$$\Delta\boldsymbol{\mathcal{P}} \equiv \begin{bmatrix} 0 & 1 \\ \mathcal{P}_{21} & \mathcal{P}_{22} \end{bmatrix},$$

where, assuming $\eta > 0$ or $f_a(1) = 0$, we have

$$\mathcal{P}_{21} = \frac{2(g_M + \delta)}{\rho^2\sigma^2}, \qquad \mathcal{P}_{22} = \frac{2(\rho\gamma^* - g)}{\rho^2\sigma^2}.$$

$r + g + \delta > \rho\gamma^{**} + \frac{1}{2}\rho^2\sigma^2$.

For the stationary distribution to exist, this matrix needs to have a positive and a negative eigenvalue which is smaller than -1. The eigvenvalues are

$$\frac{\rho\gamma^* - g}{\rho^2\sigma^2} \pm \sqrt{\left(\frac{\rho\gamma^* - g}{\rho^2\sigma^2}\right)^2 + \frac{2\left(g_M + \delta\right)}{\rho^2\sigma^2}}.$$

The product of the two is negative so long as $g_M + \delta > 0$. For the eigenvalue to be less than -1, we have the condition

$$g_M + \delta > \rho\gamma^* - g + \frac{1}{2}\rho^2\sigma^2.$$

To see this, note that the Pareto tail index is given by

$$\zeta \equiv \sqrt{\left(\frac{\rho\gamma^* - g}{\rho^2\sigma^2}\right)^2 + \frac{2\left(g_M + \delta\right)}{\rho^2\sigma^2}} - \frac{\rho\gamma^* - g}{\rho^2\sigma^2} > 1, \tag{C.16}$$

suggesting

$$\left(\frac{\rho\gamma^* - g}{\rho^2\sigma^2}\right)^2 + \frac{2\left(g_M + \delta\right)}{\rho^2\sigma^2} > \left(\frac{\gamma^* - g}{\sigma^2} + 1\right)^2,$$

and

$$\begin{aligned}\frac{2\left(g_M + \delta\right)}{\rho^2\sigma^2} &> \left(\frac{\rho\gamma^* - g}{\rho^2\sigma^2} + 1\right)^2 - \left(\frac{\rho\gamma^* - g}{\rho^2\sigma^2}\right)^2 \\ &= 2\left(\frac{\rho\gamma^* - g}{\rho^2\sigma^2} + \frac{1}{2}\right).\end{aligned}$$

□

Proposition. *See Proposition 8.*

Proof. I use Lemma 8 below that decomposes average productivity growth everywhere along the path. Consider a stationary balanced growth path, along which $F_t\left(\theta\right) = H\left(z_t\left(\theta\right)\right)$. We also have$J_t\left(\theta\right) = J\left(\gamma^*\left(z_t\left(\theta\right)\right)\right)$, which implies

$$\begin{aligned}\varepsilon^{J,1}_{\theta,t} &= \frac{\theta J_t'}{J_t} = \rho\frac{J'}{J} = \rho\varepsilon^J_1, \\ \varepsilon^{J,2}_{\theta,t} &= \frac{\theta^2 J_t''}{J_t} = \frac{\theta^2}{J_t}\left(\frac{\rho}{\theta}J'\right)' = \rho^2\frac{J''}{J} - \rho\frac{J'}{J} = \rho^2\left[\left(\varepsilon^J_1\right)^2 + \varepsilon^J_2\right] - \rho\varepsilon^J_1, \\ \frac{\dot{J}_t}{J_t} &= -\rho g_\theta\frac{J'}{J} = -\rho g_\theta\varepsilon^J_1,\end{aligned}$$

where, with some abuse of notation, I have defined $J(z) \equiv J(\gamma^*(z))$. We then find

$$\begin{aligned}\rho g_\theta &= \lambda_o \left(\frac{\mathbb{E}_F\left[e^z J\right]}{\mathbb{E}_H\left[e^z J\right]} - \frac{e^{z^o}}{\mathbb{E}_H\left[e^z J\right]} \right) - g_M \left(1 - \frac{\mathbb{E}_F\left[e^z J\right]}{\mathbb{E}_H\left[e^z J\right]} \right) + \mathbb{E}_\nu \left[\gamma^* \left(1 + \varepsilon_1^J \right) \right] \\ &\quad + \frac{\sigma^2}{2} \left(\rho(\rho - 1) + \rho^2 \mathbb{E}_\nu \left[2\varepsilon_1^J + \left(\varepsilon_1^J \right)^2 + \varepsilon_2^J \right] - \rho \mathbb{E}\left[\varepsilon_1^J \right] \right).\end{aligned}$$

□

Lemma 8. *(Evolution of Sectoral State Variables) Assume that (1) Gibrat's law holds for the largest firms,* $\lim_{\theta\to\infty} \gamma_t(\theta) = \gamma^{**} < \infty$*, and (2) the distribution of firm size has a finite expectation at all times, i.e.,* $m_t(\theta) = O(\theta^{-1-\rho})$ *for all* t*. Consider optimal firm incumbent and entrant policies, that is, the paths of investment policies and exit threshold* $(\gamma_t^*(\cdot), \theta_{o,t})$ *and the rates of entry* $\lambda_{e,t}$*. Assume that the distribution of entrants is given by* $f_{e,t}(\cdot)$*. Let* $f_t(\cdot)$ *denote the distribution of firm productivity* θ *at time* t*, and let* $\mathbb{E}_{\nu,t}$ *denote expectation with respect to the share of each type in sales at time* t*.*[7] *The evolution of the sector state variables satisfies the following equations*

$$\frac{\dot{\overline{M}}_t}{\overline{M}_t} = \lambda_{e,t} - \lambda_{o,t}, \tag{C.17}$$

$$\begin{aligned}\frac{\dot{\overline{\theta}}_t}{\overline{\theta}_t} &= \frac{\lambda_{o,t}}{\rho} \left[\left(\frac{\overline{\theta}_{e,t}}{\overline{\theta}_t} \right)^\rho - \left(\frac{\theta_{o,t}}{\overline{\theta}_t} \right)^\rho \right] - \frac{\lambda_{e,t} - \lambda_{o,t}}{\rho} \left[1 - \left(\frac{\overline{\theta}_{e,t}}{\overline{\theta}_t} \right)^\rho \right] + \mathbb{E}_{\nu,t} \left[\gamma_{\theta,t} \left(1 + \frac{1}{\rho} \epsilon_{\theta,t}^{J,1} \right) \right] \\ &\quad + (\rho - 1) \frac{\sigma^2}{2} \left(1 + \frac{1}{\rho - 1} \mathbb{E}_{\nu,t} \left[2\varepsilon_{\theta,t}^{j,1} + \frac{\varepsilon_{\theta,t}^{J,2}}{\rho} \right] \right) + \frac{1}{\rho} \mathbb{E}_{\nu,t} \left[\frac{\dot{J}_t}{J_t} \right],\end{aligned} \tag{C.18}$$

where $\lambda_{o,t}$ *as in Equation (C.2), I have defined and* $J_t(\theta) \equiv J(\gamma_{\theta,t}(\theta))$ *and the corresponding elasticities* $\varepsilon_{\theta,t}^{J,1} \equiv \frac{\theta \partial J_t}{J_t \partial \theta}$ *and* $\varepsilon_{\theta,t}^{J,2} \equiv \frac{\theta^2 \partial^2 J_t}{J_t \partial \theta^2}$*.*

Proof. Let us calculate the integral in the definition of the aggregate productivity in Equation

[7] Note that that the notation in this proof is slightly different from the one in the main text, to allow for recylcing variable names.

(C.6) using the KFE (C.4):

$$\begin{aligned}
\frac{d}{dt}\left(\overline{M}_t\,\overline{\theta}_t^{\rho}\right) &\overset{(1)}{=} \overline{M}_t\lambda_{e,t}\int d\theta\, f_{e,t}\,\theta^{\rho}J_t - \int d\theta\,\frac{\partial}{\partial\theta}\left[\gamma_t\,\theta m_t\right]\theta^{\rho}J_t + \frac{\sigma^2}{2}\int d\theta\,\frac{\partial^2}{\partial^2\theta}\left[\theta^2 m_t\right]\theta^{\rho}J_t + \int d\theta\, m_t\theta^{\rho}\dot{J}_t,\\
&\overset{(2)}{=} \overline{M}_t\lambda_{e,t}\,\mathbb{E}_{f_{e,t}}\left[\theta^{\rho}J_t\right] - \left[\gamma_t\, m_t\theta^{1+\rho}J_t\right]_{\theta_{o,t}}^{\infty} + \rho\int d\theta\,\gamma_t\, m_t\theta^{\rho}J_t\left(1+\frac{\theta J_t'}{\rho J_t}\right)\\
&\qquad + \frac{\sigma^2}{2}\left[\frac{\partial}{\partial\theta}\left(\theta^2 m_t\right)\theta^{\rho}J_t\right]_{\theta_t^o}^{\infty} - \frac{\rho\sigma^2}{2}\int d\theta\,\frac{\partial}{\partial\theta}\left(\theta^2 m_t\right)\left(\theta^{\rho-1}J_t + \frac{1}{\rho}\theta^{\rho}J_t'\right) + \overline{M}_t\,\mathbb{E}_{f_t}\left[\theta^{\rho}\dot{J}_t\right],\\
&\overset{(3)}{=} \overline{M}_t\lambda_{e,t}\,\mathbb{E}_{f_{e,t}}\left[\theta^{\rho}J_t\right] + \overline{M}_t\,\mathbb{E}_{f_t}\left[\gamma_{\theta,t}\theta^{\rho}J_t\left(\rho+\varepsilon_{\theta,t}^{J,1}\right)\right] - \frac{\sigma^2}{2}\theta_{o,t}^{\rho}\frac{\partial}{\partial\theta}\left(\theta^2 m_t\right)\Big|_{\theta_{o,t}} + \overline{M}_t\,\mathbb{E}_{f_t}\left[\theta^{\rho}\dot{J}_t\right]\\
&\qquad -\frac{\rho\sigma^2}{2}\left[\theta^2 m_t\left(\theta^{\rho-1}J_t + \frac{1}{\rho}\theta^{\rho}J_t'\right)\right]_{\theta_{o,t}}^{\infty}\\
&\qquad\quad +\rho\left(\rho-1\right)\frac{\sigma^2}{2}\int d\theta\, m_t\theta^{\rho}J_t\left(1+\frac{2}{\rho-1}\frac{\theta J_t'}{J_t} + \frac{1}{\rho\left(\rho-1\right)}\frac{\theta^2 J_t''}{J_t}\right)\\
&\overset{(4)}{=} \overline{M}_t\lambda_{e,t}\,\mathbb{E}_{f_{e,t}}\left[\theta^{\rho}J_t\right] + \overline{M}_t\,\mathbb{E}_{f_t}\left[\gamma_{\theta,t}\theta^{\rho}J_t\left(\rho+\varepsilon_{\theta,t}^{J,1}\right)\right] - \frac{\sigma^2}{2}\theta_{o,t}^{\rho}\frac{\partial}{\partial\theta}\left(\theta^2 m_t\right)\Big|_{\theta_{o,t}} + \overline{M}_t\,\mathbb{E}_{f_t}\left[\theta^{\rho}\dot{J}_t\right]\\
&\qquad\qquad + \frac{\sigma^2}{2}\,\mathbb{E}_{f_t}\left[\theta^{\rho}J_t\left(\rho\left(\rho-1\right)+2\rho\varepsilon_{\theta,t}^{J,1}+\varepsilon_{\theta,t}^{J,2}\right)\right]
\end{aligned}$$

where in equality (1), I have substituted for $\frac{\partial m_t}{\partial t}$ from Equation (C.5), in equality (2), used the chain rule to expand the integrals, and in equality (3), used the fact that $m_t\left(\theta\right) = O\left(\theta^{-(1+\rho)}\right)$ to drop the boundary terms involving $\theta\to\infty$. Finally, in equality (4) I have again used $m_t\left(\theta\right) = O\left(\theta^{-(1+\rho)}\right)$ and $m_t\left(\theta_{o,t}\right) = 0$ to drop the term involving boundary values.

Dividing the two sides of the last equation by $\overline{M}_t\overline{\theta}_t^{\rho}$, we find

$$\begin{aligned}
\frac{\dot{\overline{M}}_t}{\overline{M}_t} + \rho\frac{\dot{\overline{\theta}}_t}{\overline{\theta}_t} &= \lambda_{e,t}\left(\frac{\overline{\theta}_{e,t}}{\overline{\theta}_t}\right)^{\rho} - \lambda_{o,t}\left(\frac{\theta_{o,t}}{\overline{\theta}_t}\right)^{\rho} + \rho\mathbb{E}_{s,t}\left[\gamma_{\theta,t}\left(1+\epsilon_{\theta,t}^{J,1}\right)\right]\\
&\qquad + \frac{\sigma^2}{2}\left(\rho\left(\rho-1\right) + \mathbb{E}_{s_t}\left[2\rho\varepsilon_{\theta,t}^{J,1}+\varepsilon_{\theta,t}^{J,2}\right]\right) + \mathbb{E}_{s_t}\left[\frac{\dot{J}_t}{J_t}\right].
\end{aligned}$$

Using $\dot{\overline{M}}_t/\overline{M}_t = \lambda_{e,t} - \lambda_{o,t}$ gives us (C.18). □

C.2.2 Section 3.3

Proposition. *The result generalizes Proposition 9 of the main text, by assuming a generalized industry-wide stock of knowledge*

$$\tilde{\theta}_t^{\vartheta} = \frac{1}{\overline{M}_t} \int \theta^{\vartheta} m_t(\theta)\, d\theta, \qquad \vartheta \neq 0,$$

and additional shocks to the productivity of entrants, with a log-Normal distribution.

Proof. The proof proceeds in the following steps. I will first discretize the probelm. I then set up the social planner's problem in the discretized space, and derive the first order conditions. Finally, I combine the first order conditions and take the continuous time limit to characterize the solution.

Discretization I discretize the space of time and states following the random walk construction of Brownian motion processes (Dixit and Pindyck, 1994). Using Itô's lemma, we find

$$d\log\theta_t = \left(\gamma_t - \frac{1}{2}\sigma^2\right) dt + \sigma d\mathcal{W}_t.$$

Abusing some notation to recycle variable names, for the sake of this proof I will define $t \equiv j\Delta t$, $\rho \ln \theta_t \equiv z\Delta h$, where $\Delta h = \rho\sigma\sqrt{\Delta t}$, to map the stochastic process for the evolution of profitability states θ_t into a random walk over a space $z \in \mathbb{Z}$ and $j \in \mathbb{N} \cup \{0\}$. The probability of moving from z by ± 1 between j and $j+1$ is given by

$$\begin{aligned}
\mu_j^+(z) &= \frac{1}{2}\left[1 + \left(\gamma_j(z) - \frac{1}{2}\sigma^2\right)\frac{\sqrt{\Delta t}}{\sigma}\right], \\
\mu_j^-(z) &= \frac{1}{2}\left[1 - \left(\gamma_j(z) - \frac{1}{2}\sigma^2\right)\frac{\sqrt{\Delta t}}{\sigma}\right],
\end{aligned}$$

where I have let $\gamma_j(z) \equiv \gamma_t(\theta_t)$.

Problem Define $M_j(z) \equiv \Delta h \sum_{z'=z_{o,j}}^{z} m_j(z')$ and let $\tilde{\mathsf{z}}_j$ be defined through

$$e^{\tilde{\mathsf{z}}_j} \equiv \left(\Delta h \sum_{z=z_{o,j}}^{\infty} \frac{m_j(z)}{\overline{M}_j} e^{\frac{\vartheta}{\rho} z\Delta h}\right)^{\frac{\rho}{\vartheta}}, \tag{C.19}$$

corresponding to the industry-wide index of knowledge. The social planner's problem can be written as that of maximizing

$$\Delta t\, N_j \sum_{j=0}^{\infty} e^{-j\, r \Delta t} \log q_j,$$

subject to the constraints

$$N_j q_j = \left(\Delta h \sum_{z=z_{o,j}}^{\infty} m_j(z)\; e^{\frac{z\,\Delta h}{1+\rho}}\, l_j(z)^{\frac{\rho}{1+\rho}} \right)^{\frac{1+\rho}{\rho}}, \tag{C.20}$$

$$N_j = L_{e,j} + \Delta h \sum_{z=z_{o,j}}^{\infty} m_j(z) \left[\psi_f + l_j(z) + \bar{i}_j \right],$$

$$\overline{M}_j \bar{i}_j = \Delta h \sum_{z=z_{o,j}}^{\infty} m_j(z)\; l_j(z)\, \varphi\left(\gamma_j(z)\right), \tag{C.21}$$

$$m_{j+1}(z) = \frac{L_{e,j}}{\psi_e} \Delta t\, f_j(z) + m_j(z+1)\, \mu_j^-(z+1) + m_j(z-1)\; \mu_j^+(z-1),$$

$$f_j(z) = \frac{\Delta h}{\rho} \sum_{z_a=z_{o,j}}^{\infty} f_e\left(\frac{z - \eta \tilde{\mathrm{z}}_j - (1-\eta)\, z_a}{\rho} \Delta h \right) \frac{m_j(z_a)}{\overline{M}_j} f_a\left(\frac{M_j(z_a)}{\overline{M}_j} \right), \tag{C.22}$$

where $\bar{i}_j$ denotes the average innovation workers of the incumbent firms. Let the Lagrange multipliers corresponding to the three constraints above be denoted by $\Delta t\, \chi_j e^{-j\, r \Delta t}$, $\Delta t\, \omega_j e^{-j\, r \Delta t}$, $\Delta t\, \omega_j^i e^{-j\, r \Delta t}$, $\Delta h\, \nu_{j+1}(z)\, e^{-j\, r \Delta t}$, respectively. The social planner decides on the per-capita consumption q_j, on the exit threshold at time $j\Delta t$, which I denote by $z_{o,j}$, on the total number of innovation workers in the entrepreneurial work $L_{e,j}$ and the average number of innovation workers in incumbent firms $\bar{i}_j$ at each point in time, the number of production workers assigned to each firm $l_j(z)$, the innovation investments of firms $\gamma_j(z)$, and the mass of firms in each state $m_j(z)$.

First Order Conditions (I): consumption, incumbent innovation and production The FOC with respect to q_j gives $\frac{N_j}{q_j} = \chi_j N_j$, which yields $q_j \chi_j = 1$. The FOC with respect to the average number of incumbent innovation workers $\bar{i}_j$ is given by $\omega_j^i = \omega_j$. First, the FOC

with respect to $l_j(z)$ gives

$$\left(\frac{N_j q_j e^{z\Delta h}}{l_j(z)}\right)^{\frac{1}{1+\rho}} \chi_j = \omega_j\left(1+\varphi\left(\gamma_j(z)\right)\right). \tag{C.23}$$

Substituting $\chi_j = 1/q_j$, using the definition of function $J(\gamma) \equiv (1+\varphi(\gamma))^{-\rho}$, and letting Θ_j be defined through

$$\overline{M}_j \Theta_j^{\rho} = \Delta h \sum_{z=z_j^o}^{\infty} m_j(z)\, e^{z\,\Delta h}\, J\left(\gamma_j(z)\right),$$

we can substitute for $l_j(z)$ in Equation (C.20) to find $\chi_j \Theta_j \overline{M}_j^{\frac{1}{\rho}} = \omega_j$.

First Order Condition (II): The FOC with respect to $\gamma_j(z)$ we find

$$\begin{aligned}
\frac{1}{2}\frac{\sqrt{\Delta t}}{\sigma}\left[\nu_{j+1}(z+1) - \nu_{j+1}(z-1)\right] &= \Delta t\, l_j(z)\ \omega_j\, \varphi'\left(\gamma_j(z)\right), &\text{(C.24)}\\
&= \Delta t\, \chi_j \left(\frac{\chi_j}{\omega_j}\right)^{\rho} N_j q_j\, e^{z\Delta h}\left(1+\varphi\left(\gamma_j(z)\right)\right)^{-(1+\rho)}\ \varphi'\left(\gamma_j(z)\right), \\
&= -\Delta t\, \frac{N_j}{\overline{M}_j}\frac{e^{z\Delta h}}{\Theta_j} J'\left(\gamma_j(z)\right), &\text{(C.25)}
\end{aligned}$$

where, in the second line, I have again substitute for $l_j(z)$ from Equation (C.20).

First Order Conditions (III): social free entry The FOC with respect to $L_{e,j}$ gives

$$\omega_j \psi_e = \Delta h \sum_{z=z_{o,j}}^{\infty} f_j(z)\, \nu_{j+1}(z). \tag{C.26}$$

First Order Conditions (IV): incumbent value The FOC with respect to $m_{j+1}(z)$ is the most involved calculation. The first term in Equation (C.20) gives the flow social value of production for a firm z at time interval $j+1$:

$$\begin{aligned}
\Delta t \Delta h\, \chi_j \frac{1+\rho}{\rho}\left(N_j q_j\right)^{\frac{1}{1+\rho}}\ e^{\frac{z\Delta h}{1+\rho}}\ l_j(z)^{\frac{\rho}{1+\rho}} &= \Delta t \Delta h\, \chi_j \frac{1+\rho}{\rho} N_j q_j \left(\frac{\chi_j}{\omega_j}\right)^{\rho} e^{z\Delta h}\left(1+\varphi\left(\gamma_j(z)\right)\right)^{-\rho}, \\
&= \Delta t \Delta h \frac{1+\rho}{\rho}\frac{N_j}{\overline{M}_j}\frac{e^{z\Delta h}}{\Theta_j} J\left(\gamma_j(z)\right).
\end{aligned}$$

The second and third terms give the employment costs by

$$\begin{aligned}\Delta t\Delta h\left[\omega_j\psi_f+\omega_j l_j(z)\left(1+\varphi\left(\gamma_j(z)\right)\right)\right] &= \Delta t\Delta h\left[\omega_j\psi_f+l_j(z)\left(\frac{N_j q_j e^{z\Delta h}}{l_j(z)}\right)^{\frac{1}{1+\rho}}\chi_j\right],\\ &= \Delta t\Delta h\left[\omega_j\psi_f+\frac{N_j}{\overline{M}_j}\frac{e^{z\Delta h}}{\Theta_j}J\left(\gamma_j(z)\right)\right],\end{aligned}$$

where on the second line, I have substituted from Equation (C.23). The fourth term gives

$$\Delta h\left[e^{r\Delta t}\nu_j(z)-\frac{L_{e,j}}{\psi_e}\left(\Delta h\sum_{x=z_{o,j}}^{\infty}\frac{\partial f_j(x)}{\partial m_j(z)}\nu_{j+1}(x)\right)\Delta t+\mu_{j+1}^{+}(z)\,\nu_{j+1}(z+1)+\mu_{j+1}^{-}(z)\,\nu_{j+1}(z-1)\right].$$

The key term is the second term in the expression above, that captures the knowledge spillovers from incumbents to entrants. This spillover term is in turn composed of two components: first, the contribution of incumbent firms to the average stock of knowledge $\tilde{z}$, and second, the direct diffusion of ideas of incumbents to entrants through adoption $f_a(\cdot)$. Defining $H_j(z)\equiv\Delta h\sum_{z'\leq z}m_j(z)/\overline{M}_j$ and $h_j(z)\equiv m_j(z)/\overline{M}_j$ to simplify notation, we can computing the first term as

$$-\frac{\eta}{\rho}\frac{\partial\tilde{z}_j}{\partial m_j(z)}\frac{\Delta h^2}{\rho}\sum_{x=z_{o,j}}^{\infty}\sum_{z_a=z_{o,j}}^{\infty}\nu_{j+1}(x)\,f_e'\left(\frac{x-\eta\tilde{z}_j-(1-\eta)z_a}{\rho}\Delta h\right)h_j(z_a)\,f_a\left(H_j(z_a)\right).$$

Taking the derivative of expression (C.19), we find

$$\frac{\partial e^{\tilde{z}_j}}{\partial m_j(z)}=\frac{\Delta h}{\overline{M}_j}\frac{\rho}{\vartheta}\left[e^{\frac{\vartheta}{\rho}\left(z\Delta h-\tilde{z}_j\right)}-1\right]\times e^{\tilde{z}_j}.$$

Similarly, the second term may be computed as

$$\frac{\Delta h^2}{\rho}\sum_{x=z_{o,j}}^{\infty}\sum_{z_a=z_{o,j}}^{\infty}\nu_{j+1}(x)\,f_e\left(\frac{x-\eta\tilde{z}_j-(1-\eta)z_a}{\rho}\Delta h\right)\left[\frac{\partial h_j(z_a)}{\partial m_j(z)}f_a\left(H_j(z_a)\right)+h_j(z_a)\,\frac{\partial H_j(z_a)}{\partial m_j(z)}f_a'\left(H_j(z_a)\right)\right].$$

Calculating the derivatives within the square brackets yields

$$\begin{aligned}&\frac{1}{\overline{M}_j}\left[\left(\mathbb{I}\left(z_a=z_j\right)-\frac{m_j(z_a)}{\overline{M}_j}\Delta h\right)f_a\left(H_j(z_a)\right)+h_j(z_a)\left(\mathbb{I}\left(z_a\geq z_j\right)-\frac{M_j(z_a)}{\overline{M}_j}\right)\Delta h\,f_a'\left(H_j(z_a)\right)\right]\\ &\quad=\frac{1}{\overline{M}_j}\left[\left(\mathbb{I}\left(z_a=z\right)-h_j(z_a)\,\Delta h\right)f_a\left(H_j(z_a)\right)+h_j(z_a)\left(\mathbb{I}\left(z_a\geq z\right)-H_j(z_a)\right)\Delta h\,f_a'\left(H_j(z_a)\right)\right].\end{aligned}$$

Combining all these expressions, we can write the FOC with respect to $m_{j+1}(z)$ as

$$e^{r\Delta t}v_j(z) = \left[\frac{1}{\rho}\frac{N_{j+1}}{\overline{M}_{j+1}}\frac{e^{z\Delta h}}{\Theta_{j+1}^{\rho}}J\left(\gamma_j(z)\right) - \psi_f\,\omega_{j+1} + \lambda_{e,j}\left(\tilde{u}_{s,j}(z) + u_{s,j}(z)\right)\right]\Delta t + \mu_{j+1}^{+}(z)\,v_{j+1}(z+1) + \mu_{j+1}^{-}(z)\,v_{j+1}(z-1) \quad \text{(C.27)}$$

where $\lambda_{e,j} = L_{e,j}/\overline{M}_j\psi_e$ is the rate of entry, and the two spillover functions are given by

$$\tilde{u}_{s,j}(z) \equiv -\frac{\eta}{\vartheta}\left[e^{\frac{\vartheta}{\rho}\left(z\Delta h - \tilde{\mathbf{z}}_j\right)} - 1\right]\frac{\Delta h^2}{\rho}\sum_{x=z_{o,j}}^{\infty}\sum_{z_a=z_{o,j}}^{\infty}v_{j+1}(x)\,f_e'\left(\frac{x-\eta\tilde{\mathbf{z}}_j-(1-\eta)\,z_a}{\rho}\Delta h\right)h_j(z_a)\,f_a\left(H_j(z_a)\right), \quad \text{(C.28)}$$

$$\begin{aligned} u_{s,j}(z) \equiv\ & \frac{\Delta h}{\rho}\sum_{x=z_{o,j}}^{\infty}f_e\left(\frac{x-\eta\tilde{\mathbf{z}}_j-(1-\eta)\,z}{\rho}\Delta h\right)f_a\left(H_j(z)\right)\,v_{j+1}(x) \\ & -\frac{\Delta h^2}{\rho}\sum_{x=z_{o,j}}^{\infty}\sum_{z_a=z_{o,j}}^{\infty}f_e\left(\frac{x-\eta\tilde{\mathbf{z}}_j-(1-\eta)\,z_a}{\rho}\Delta h\right)h_j(z_a)\,f_a\left(H_j(z_a)\right)\,v_{j+1}(x) \\ & +\frac{\Delta h^2}{\rho}\sum_{z_a=z}^{\infty}\sum_{x=z_{o,j}}^{\infty}f_e\left(\frac{x-\eta\tilde{\mathbf{z}}_j-(1-\eta)\,z_a}{\rho}\Delta h\right)h_j(z_a)\,f_a'\left(H_j(z_a)\right)\,v_{j+1}(x) \\ & -\frac{\Delta h^2}{\rho}\sum_{x=z_{o,j}}^{\infty}\sum_{z_a=z_{o,j}}^{\infty}f_e\left(\frac{x-\eta\tilde{\mathbf{z}}_j-(1-\eta)\,z_a}{\rho}\Delta h\right)H_j(z_a)\,f_a'\left(H_j(z_a)\right)\,v_{j+1}(x). \end{aligned} \quad \text{(C.29)}$$

Limit of Continuous Time Now, let us take the limit of $\Delta t \to \infty$, define $\hat{\theta} \equiv \rho\log\theta$ and $\widehat{\tilde{\theta}}_t \equiv \rho\log\tilde{\theta}_t$, and with some further abuse of notation, let $v_{s,t}\left(\hat{\theta}\right)$ denote the social value of the a firm with productivity $e^{\hat{\theta}/\rho}$. I will first compute the limiting expressions for the summations in Equations (C.28) and (C.29). For l define the auxiliary function

$$\mathcal{V}_t\left(\hat{\theta}\right) \equiv \int_{\hat{\theta}_{o,t}}^{\infty}v_{s,t}(x)\,f_e\left(\frac{x-\eta\widehat{\tilde{\theta}}_t-(1-\eta)\,\hat{\theta}}{\rho}\right)\frac{dx}{\rho}.$$

Assuming $v_{s,t}\left(\hat{\theta}_{o,t}\right)=0$ and the right tail of $f_e\left(x\right)$ falls faster than e^{-x}, using integration by parts, we find

$$\begin{aligned}
\mathcal{V}_t'\left(\hat{\theta}\right) &= -\frac{1-\eta}{\rho}\int_{\hat{\theta}_{o,t}}^{\infty} v_{s,t}\left(x\right)\, f_e'\left(\frac{x-\eta\widehat{\tilde{\theta}}_t-\left(1-\eta\right)\hat{\theta}}{\rho}\right)\frac{dx}{\rho},\\
&= -\frac{1-\eta}{\rho}\left[v_{s,t}\left(x\right)\, f_e\left(\frac{x-\eta\widehat{\tilde{\theta}}_t-\left(1-\eta\right)\hat{\theta}}{\rho}\right)\right]_{x=\hat{\theta}_{o,t}}^{\infty}\\
&\qquad +\left(1-\eta\right)\int_{\hat{\theta}_{o,t}}^{\infty} v_{s,t}'\left(x\right)\, f_e\left(\frac{x-\eta\widehat{\tilde{\theta}}_t-\left(1-\eta\right)\hat{\theta}_a}{\rho}\right)\frac{dx}{\rho},\\
&= \left(1-\eta\right)\int_{\hat{\theta}_{o,t}}^{\infty} v_{s,t}'\left(x\right)\, f_e\left(\frac{x-\eta\widehat{\tilde{\theta}}_t-\left(1-\eta\right)\hat{\theta}_a}{\rho}\right)\frac{dx}{\rho}.
\end{aligned}$$

Using a similar argument, we can write the limit of the first spillover term may be found

$$\begin{aligned}
\tilde{u}_{s,t}\left(\hat{\theta}\right) &= -\frac{\eta}{\vartheta}\, e^{\frac{\vartheta}{\rho}\left(\hat{\theta}-\widehat{\tilde{\theta}}_t\right)}\int_{\hat{\theta}_{o,t}}^{\infty}\int_{\hat{\theta}_{o,t}}^{\infty} v_{s,t}\left(x\right)\, f_e'\left(\frac{x-\eta\widehat{\tilde{\theta}}_t-\left(1-\eta\right)\hat{\theta}_a}{\rho}\right) h_t\left(\hat{\theta}_a\right)\, f_a\left(H_t\left(\hat{\theta}_a\right)\right)\frac{dx}{\rho}\, d\hat{\theta}_a,\\
&= \eta\,\frac{\rho}{\vartheta}\, e^{\frac{\vartheta}{\rho}\left(\hat{\theta}-\widehat{\tilde{\theta}}_t\right)}\int_{\hat{\theta}_{o,t}}^{\infty}\int_{\hat{\theta}_{o,t}}^{\infty} v_{s,t}'\left(x\right)\, f_e\left(\frac{x-\eta\widehat{\tilde{\theta}}_t-\left(1-\eta\right)\hat{\theta}_a}{\rho}\right) h_t\left(\hat{\theta}_a\right)\, f_a\left(H_t\left(\hat{\theta}_a\right)\right)\frac{dx}{\rho}\, d\hat{\theta}_a.
\end{aligned}$$

For the second term, the limit is given by

$$\begin{aligned}
u_{s,t}\left(\hat{\theta}\right) &= f_a\left(H_t\left(\hat{\theta}\right)\right)\,\mathcal{V}_t\left(\hat{\theta}\right)-\int_{\hat{\theta}_{o,t}}^{\infty}\mathcal{V}_t\left(\hat{\theta}_a\right)\, h_t\left(\hat{\theta}_a\right)\, f_a\left(H_t\left(\hat{\theta}_a\right)\right)\, d\hat{\theta}_a\\
&\qquad +\int_{\hat{\theta}}^{\infty}\mathcal{V}_t\left(\hat{\theta}_a\right)\, h_t\left(\hat{\theta}_a\right)\, f_a'\left(H_t\left(\hat{\theta}_a\right)\right)\, d\hat{\theta}_a\\
&\qquad -\int_{\hat{\theta}_{o,t}}^{\infty}\mathcal{V}_t\left(\hat{\theta}_a\right)\, H_t\left(\hat{\theta}_a\right)\, f_a'\left(H_t\left(\hat{\theta}_a\right)\right)\, d\hat{\theta}_a.
\end{aligned}$$

Once again, using integration by part, we can simplify this expression to find

$$\begin{aligned}
u_{s,t}\left(\hat{\theta}\right) &= \mathcal{V}_t\left(\hat{\theta}\right)\, f_a\left(H_t\left(\hat{\theta}\right)\right)-\int_{\hat{\theta}_{o,t}}^{\infty}\mathcal{V}_t\left(\hat{\theta}_a\right)\, h_t\left(\hat{\theta}_a\right)\, f_a\left(H_t\left(\hat{\theta}_a\right)\right)\, d\hat{\theta}_a\\
&\qquad +\left[\mathcal{V}_t\left(\hat{\theta}_a\right)\, f_a\left(H_t\left(\hat{\theta}_a\right)\right)\right]_{\hat{\theta}_a=\hat{\theta}}^{\infty}-\int_{\hat{\theta}}^{\infty}\mathcal{V}_t'\left(\hat{\theta}_a\right)\, f_a\left(H_t\left(\hat{\theta}_a\right)\right)\, d\hat{\theta}_a\\
&\qquad -\left[\mathcal{V}_t\left(\hat{\theta}_a\right)\, H_t\left(\hat{\theta}_a\right)\, f_a\left(H_t\left(\hat{\theta}_a\right)\right)\right]_{x=\hat{\theta}_{o,t}}^{\infty}+\int_{\hat{\theta}_{o,t}}^{\infty}\frac{d}{d\hat{\theta}_a}\left[\mathcal{V}_t\left(\hat{\theta}_a\right)\, H_t\left(\hat{\theta}_a\right)\right]\, f_a'\left(H_t\left(\hat{\theta}_a\right)\right)\, d\hat{\theta}_a,\\
&= -\int_{\hat{\theta}}^{\infty}\mathcal{V}_t'\left(\hat{\theta}_a\right)\, h_t\left(\hat{\theta}_a\right)\, f_a\left(H_t\left(\hat{\theta}_a\right)\right)\, d\hat{\theta}_a+\int_{\hat{\theta}_{o,t}}^{\infty}\mathcal{V}_t'\left(\hat{\theta}_a\right)\, H_t\left(\hat{\theta}_a\right)\, f_a\left(H_t\left(\hat{\theta}_a\right)\right)\, d\hat{\theta}_a,\\
&= \int_{\hat{\theta}_{o,t}}^{\hat{\theta}}\mathcal{V}_t'\left(\hat{\theta}_a\right)\, f_a\left(H_t\left(\hat{\theta}_a\right)\right)\, d\hat{\theta}_a-\int_{\hat{\theta}_{o,t}}^{\infty}\mathcal{V}_t'\left(\hat{\theta}_a\right)\,\left[1-H_t\left(\hat{\theta}_a\right)\right]\, f_a\left(H_t\left(\hat{\theta}_a\right)\right)\, d\hat{\theta}_a,
\end{aligned}$$

where in the second equality, I have cancelled out the same terms that appear on the first three lines, and in the last equality, I have added and subtracted a term $\int_{\hat{\theta}_{o,t}}^{\infty} \mathcal{V}_t' f_a d\hat{\theta}_a$. We can now

$$\begin{aligned} u_{s,t}\left(\hat{\theta}\right) &= (1-\eta)\int_{\hat{\theta}_{o,t}}^{\infty}\int_{\hat{\theta}_{o,t}}^{\infty} v_{s,t}'\left(x\right) f_e\left(\frac{x-\eta\widehat{\tilde{\theta}}_t-(1-\eta)\hat{\theta}_a}{\rho}\right) f_a\left(H_t\left(\hat{\theta}_a\right)\right)\frac{dx}{\rho}\,d\hat{\theta}_a \\ &\quad -(1-\eta)\int_{\hat{\theta}_{o,t}}^{\infty}\int_{\hat{\theta}_{o,t}}^{\infty} v_{s,t}'\left(x\right) f_e\left(\frac{x-\eta\widehat{\tilde{\theta}}_t-(1-\eta)\hat{\theta}_a}{\rho}\right)\left[1-H_t\left(\hat{\theta}_a\right)\right] f_a\left(H_t\left(\hat{\theta}_a\right)\right)\frac{dx}{\rho}\,d\hat{\theta}_a. \end{aligned}$$

Putting everything together, we can combine the FOCs with respect to $\gamma_j(z)$ and $m_j(z)$ to find in the limit:

$$\begin{aligned} r\,v_{s,t}\left(\hat{\theta}\right)-\frac{\partial v_{s,t}\left(\hat{\theta}\right)}{\partial t} &= \max_{\gamma}\frac{1}{\rho}\frac{N_t}{\overline{M}_t}\frac{e^{\hat{\theta}}}{\overline{\theta}_t^{\rho}}J\left(\gamma\right)-\psi_f\,\omega_t+\lambda_{e,t}\left(\tilde{u}_{s,t}\left(\hat{\theta}\right)+u_{s,t}\left(\hat{\theta}\right)\right) \\ &\quad +\left(\gamma-\frac{\rho}{2}\sigma_\theta^2\right)\frac{\partial v_{s,t}\left(\hat{\theta}\right)}{\partial\hat{\theta}}+\frac{\rho^2\sigma_\theta^2}{2}\frac{\partial^2 v_{s,t}\left(\hat{\theta}\right)}{\partial\hat{\theta}^2}. \end{aligned}$$

Let us now assume that a social optimal equilibrium converges to a stationary constant growth path with growth rate $g_{\theta,s}$, where $\omega_t\to\omega$, $\overline{M}_t\to\overline{M}e^{g_N t}$, and $\overline{\theta}_t\to\overline{\theta}e^{g_\theta t}$. We can define $z_t\left(\hat{\theta}\right)\equiv\hat{\theta}-\rho g_{\theta,s}t+\log\left(\frac{1}{\rho\psi_f\omega}\frac{N_0}{\overline{M}}\overline{\theta}^{-\rho}\right)$ and transform $v_{s,t}\left(\hat{\theta}\right)=\omega\psi_f V\left(z_t\left(\hat{\theta}\right)\right)$. Similar to the construction of the market equilibrium path, this implies that $z_t\left(\hat{\theta}_{o,t}\right)$ and $z_t\left(\tilde{\theta}_t\right)$ also converge to constants $z_{o,s}$ and $\tilde{z}_s$ that are potentially different from the corresponding market equilibrium values.

$$r\,V_s\left(z\right)=\max_{\gamma}e^z J\left(\gamma\right)-1+\lambda_{e,s}\left(U_s\left(z\right)-\Psi_s\right)+\left(\gamma-\rho\left(g_{\theta,s}+\frac{1}{2}\sigma_\theta^2\right)\right)V_s'\left(z\right)+\frac{\rho^2}{2}\sigma_\theta^2\,V_s''\left(z\right),$$

where

$$\begin{aligned} U_s\left(z\right) &\equiv \eta\,\tilde{u}_s\frac{\rho}{\vartheta}\left[e^{\frac{\vartheta}{\rho}\left(z-\tilde{z}\right)}-1\right]+(1-\eta)\left(u_s\left(z\right)-\psi_s\right) \\ \tilde{u}_s &\equiv \int_{z_o}^{\infty}\mathcal{V}_e'\left(x\right)h_s\left(x\right)f_a\left(H_s\left(x\right)\right)dx, \\ u_s\left(z\right) &\equiv \int_{z_o}^{z}\mathcal{V}_e'\left(x\right)f_a\left(H_s\left(x\right)\right)dx, \\ \psi_s &\equiv \int_{z_o}^{\infty}\mathcal{V}_e'\left(x\right)\left[1-H_s\left(x\right)\right]f_a\left(H_s\left(x\right)\right)dx, \\ \mathcal{V}_e\left(z\right) &\equiv \int_{z_o}^{\infty}V\left(x\right)f_e\left(\frac{x-\eta\tilde{z}-(1-\eta)z}{\rho}\right)\frac{dx}{\rho}. \end{aligned}$$

In the particular case where γ_e has an atomic distribution, we have $\mathcal{V}_e(z) \equiv V(\eta\tilde{z} + (1-\eta)z + \rho\gamma_e)$.

□

C.2.3 Section 3.4

Specification of the Innovation Cost Function With the functional form assumed in Equation (3.39), the condition for the concavity of $J(\cdot)$ from Equation (3.5) gives

$$\kappa > (1+\rho)(1+\kappa)\frac{\varphi(\gamma)}{1+\varphi(\gamma)}.$$

It then follows that an upper bound for the rate of productivity growth given in Equation (3.40) ensures the convavity of function $J(\cdot)$. Throughout, I assume the value given by Equation (3.40) as the maximum attainable rate of productivity growth for firms.

In order to calibrate parameter κ, I rely on the estimates of the user cost elasticity of R&D that are available in the empirical literature. We can compute a counterpart for this elasticity based on the model by noting that $i = l\,\varphi(\gamma)$ and finding the elasticity of the innovation costs with respect to a proportional change $(1+\tau)$ in innovation costs. We have

$$\begin{aligned}\frac{1+\tau}{i}\frac{\partial i}{\partial(1+\tau)} &= \frac{1+\tau}{l}\frac{\partial l}{\partial(1+\tau)} + \frac{1+\tau}{\varphi(\gamma)}\frac{\partial\varphi(\gamma)}{\partial(1+\tau)},\\ &= \left[1-(1+\rho)\frac{\varphi}{1+\varphi}\right]\frac{1+\tau}{\varphi(\gamma)}\frac{\partial\varphi(\gamma)}{\partial(1+\tau)},\end{aligned}$$

where I have used the fact that $l^{-\frac{1}{1+\rho}} \propto 1+\varphi$. Focusing on the limit of large firms, we can compute the elasticity of $\varphi(\gamma^{**})$ with respect to $1+\tau$ using the condition

$$\frac{\rho}{\gamma^{**}}\frac{\varphi}{1+\varphi}(1+\tau)(1+\kappa) = \frac{1}{r^{**}},$$

where r^{**} is the effective discount rate, which is not influenced by this partial equilibrium analysis. We then find

$$\left[1-\frac{\varphi}{1+\varphi}-\frac{1}{1+\kappa}\right]\frac{1+\tau}{\varphi(\gamma)}\frac{\partial\varphi(\gamma)}{\partial(1+\tau)} = -1.$$

Combining the two expressions above, we have

$$\frac{1+\tau}{i}\frac{\partial i}{\partial\left(1+\tau\right)}=-\left(1+\kappa\right)\frac{1-\left(1+\rho\right)\frac{\varphi}{1+\varphi}}{\kappa-\frac{\varphi}{1+\varphi}}.$$

Two-Sector Extension of the Model For the exercise performed in Section 3.4, I include data from two sectors manufacturing and retail trade, where the employment in each sector grows at a different rate. Below, I generalize to account for two sectors and show that the calibration strategy of Section 3.4 is.

Define the per-capita real consumption at time t as a function of per-capita consumption of manufacturing goods $q_{\mathsf{M},t}$ and retail goods $q_{\mathsf{R},t}$ as follows

$$q_t=\left(q_{\mathsf{M},t}^{\frac{\varsigma-1}{\varsigma}}+q_{\mathsf{R},t}^{\frac{\varsigma-1}{\varsigma}}\right)^{\frac{\varsigma}{\varsigma-1}},$$

where $\varsigma\in(0,1)$ is the consumer's elasticity of substitution between the products of the two sector.[8]

$$P_t=\left(P_{\mathsf{M},t}^{1-\varsigma}+P_{\mathsf{R},t}^{1-\varsigma}\right)^{\frac{1}{1-\varsigma}},$$

where Once again, normalizing the per-capita expenditure of households to unity, we find sectoral expenditures to be

$$X_{\mathsf{M},t}=P_{\mathsf{M},t}q_{\mathsf{M},t}=\left(\frac{P_{\mathsf{M},t}}{P_t}\right)^{1-\varsigma},\qquad X_{\mathsf{R},t}=P_{\mathsf{R},t}q_{\mathsf{R},t}=\left(\frac{P_{\mathsf{R},t}}{P_t}\right)^{1-\varsigma}.$$

The relative price indices of the two sectors are given by

$$\frac{P_{\mathsf{M},t}}{P_{\mathsf{R},t}}=\frac{\overline{\theta}_{\mathsf{R},t}\overline{M}_{\mathsf{R},t}^{\frac{1}{\rho}}}{\overline{\theta}_{\mathsf{M},t}\overline{M}_{\mathsf{M},t}^{\frac{1}{\rho}}},$$

which gives us the following relationship between the per-capita expenditure between the two sectors

$$\frac{X_{\mathsf{M},t}}{X_{\mathsf{R},t}}=\left(\frac{\overline{\theta}_{\mathsf{R},t}}{\overline{\theta}_{\mathsf{M},t}}\right)^{1-\varsigma}\left(\frac{\overline{M}_{\mathsf{R},t}}{\overline{M}_{\mathsf{M},t}}\right)^{\frac{1-\varsigma}{\rho}}.$$

[8]Based on the available empirical evidence, the empirically relevant case is one in which the expenditure shares of sectors with falling prices also fall, suggesting that sectoral goods are gross complements (Comin *et al.*, 2015).

In the data, the number of firms grow faster in retail and the sectoral productivity grows faster in manufacturing. So long as the differences in the rates of growth in the number of firms and aggregate productivity between the two sectors satisfies

$$g_{\theta,\mathsf{M}} - g_{\theta,\mathsf{R}} > \frac{1}{\rho}\left(g_{M,\mathsf{R}} - g_{M,\mathsf{M}}\right),$$

we will find that the expenditure share of manufacturing asymptotically converges to zero at the rate

$$(1-\varsigma)\left[g_{\theta,\mathsf{M}} - g_{\theta,\mathsf{R}} - \frac{1}{\rho}\left(g_{M,\mathsf{R}} - g_{M,\mathsf{M}}\right)\right].$$

It then follows that we can construct a general equilibrium path where wages converge to a constant, the total expenditure in retail sector grows at the rate $g_{M,\mathsf{R}} = g_N$, the total expenditure in manufacturing grows at the rate given by

$$g_N - (1-\varsigma)\left[g_{\theta,\mathsf{M}} - g_{\theta,\mathsf{R}} - \frac{1}{\rho}\left(g_{M,\mathsf{R}} - g_{M,\mathsf{M}}\right)\right].$$

The mass of firms in each sector grows at the same rate as the rate of growth of the total expenditure of consumers in the sector.

This shows that the model is compatible with the exercise used in Section 3.4, where each sector is identified with a distinct rate of growth in the mass of firms (or alternatively, sectoral employment) and productivity. We can pin down the free parameter ς based on the observed rates of growth in the mass of firms and productivity.