<table>
<thead>
<tr>
<th>Citation</th>
<th>Jaffe, Sonia Patricia. 2015. Essays on Heterogeneity in Markets and Games. Doctoral dissertation, Harvard University, Graduate School of Arts &amp; Sciences.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Citable link</td>
<td><a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:17467179">http://nrs.harvard.edu/urn-3:HUL.InstRepos:17467179</a></td>
</tr>
<tr>
<td>Terms of Use</td>
<td>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 <a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA">http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA</a></td>
</tr>
</tbody>
</table>
Essays on Heterogeneity in Markets and Games

A dissertation presented
by
Sonia Patricia Jaffe
to
The Department of Economics
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy
in the subject of
Economics

Harvard University
Cambridge, Massachusetts
April 2015
Abstract

Many markets exhibit substantial heterogeneity – e.g. in ability, in preferences, in products, in strategies. Allowing for this (sometimes multi-dimensional) heterogeneity can change both the theoretical predictions of models and the results of empirical analyses. This dissertation consists of three essays on markets and games with different forms of heterogeneity.

The first chapter introduces preference heterogeneity and multi-dimensional skill heterogeneity into the analysis of labor markets. In matching markets, agents have heterogeneous preferences over potential partners, so welfare depends on which agents are matched to each other in equilibrium. Taxes in matching markets can generate inefficiency by changing who is matched to whom, even if the number of workers at each firm is unaffected. This “allocative distortion” is not evident in traditional models of income taxation that do not allow for workers to have multi-dimensional preference and productivity heterogeneity. For markets in which workers refuse to match without a positive wage, higher taxes decrease match efficiency. However, in more balanced matching markets where transfers may flow in either direction, such as the student–college market, lowering taxes may decrease match efficiency because an agent can transfer enough to “buy” an inefficient partner (only to be “bought back” when taxes are lowered further). Simulations show that, in matching markets, traditional deadweight loss estimates based on the change in taxable income can be substantially biased in either direction.

The second chapter builds on the merger analysis literature that recognizes the impor-
tance of product heterogeneity. Both merger simulations and the more recent “first-order” approach to merger analysis recognize that because product heterogeneity can vary, market concentration is not always a good measure of the competitiveness of the market. We derive approximations of the expected changes in prices and welfare generated by a merger, using information local to the pre-merger equilibrium. We extend the pricing pressure approach of recent work to allow for non-Bertrand conduct, adjusting the diversion ratio and incorporating the change in anticipated accommodation. To convert pricing pressures into quantitative estimates of price changes, we multiply them by the merger pass-through matrix. Pass-through rates can vary by industry and the same pricing pressure can lead to very different price changes, depending on the pass-through rates. Weighting the price changes by quantities gives the change in consumer surplus.

The third chapter uses data on thousands of players who play a game over a hundred times to do a with-in player analysis of play and allow for heterogeneity in the mixed strategies that players use. We use of data from a Facebook application where users play a simultaneous move, zero-sum game – rock-paper-scissors – with varying information to provide empirical insights into whether play is consistent with extant theories. We report three major insights. First, we observe that many employ strategies consistent with Nash, at least some of the time. Second, players predictably respond to incentives in the game. For example, out of equilibrium, players strategically use information on previous play of their opponents, and they are more strategic when the payoffs for such actions increase. Third, experience matters: players with more experience use information on their opponents more efficiently than less experienced players, and are more likely to win as a result. We also explore the degree to which the deviations from Nash predictions are consistent with various non-equilibrium models. We find that both a level-$k$ framework and a quantal response model have explanatory power: whereas one group of people employ strategies that are close to $k_1$, there is also a set of people who use strategies that resemble quantal response.
## Contents

Abstract ................................................................. iii  
Acknowledgments ......................................................... vii 

### 1 Taxation in Matching Markets

1.1 General Model ......................................................... 4  
1.2 Proportional Taxation .................................................. 9  
   1.2.1 Possible Inefficiencies of Tax-Reduction in General Markets ........ 10  
   1.2.2 Efficiency of Tax-Reduction in Wage Markets .......................... 16  
   1.2.3 Deadweight Loss .................................................. 22  
1.3 Lump Sum Taxation .................................................... 23  
   1.3.1 Lump Sum Taxation of Transfers ..................................... 24  
   1.3.2 Lump Sum Taxation of Matches ..................................... 30  
1.4 Discussion ............................................................ 31  
   1.4.1 Importance of Taxation in Different Markets .......................... 31  
   1.4.2 Subsidies ......................................................... 36  
   1.4.3 Structural Properties ............................................. 38  
1.5 Conclusion ............................................................ 39  

### 2 The First-Order Approach to Merger Analysis

2.1 Background on the First-Order Approach ................................ 44  
2.2 Generalized Pricing Pressure .......................................... 46  
   2.2.1 The model ....................................................... 46  
   2.2.2 Incentives created by a merger .................................... 48  
   2.2.3 Marginal cost efficiencies ....................................... 49  
   2.2.4 Specific Contexts and Examples ................................... 50  
2.3 Price Changes .......................................................... 56  
2.4 Role of Pass-Through .................................................. 62  
   2.4.1 Pre-merger, post-merger and merger pass-through .................... 63  
   2.4.2 Calculation and approximation of merger pass-through ............... 64  
2.5 Welfare Changes ....................................................... 66
2.6 Applying the formula ........................................ 68
    2.6.1 Simplifying the formula .................................. 69
    2.6.2 Approximateness ........................................ 70
    2.6.3 At which stage should our tools apply? ................. 71
    2.6.4 Other applications ...................................... 71
2.7 Conclusion .................................................. 72

3 Understanding Behavior in Strategic Settings: Evidence from a Million
   Rock-Paper-Scissors Games 74
   3.1 An Introduction to Roshambull .............................. 78
   3.2 Model of the game .......................................... 83
   3.3 Players respond to information .............................. 84
   3.4 Level-k behavior ........................................... 90
       3.4.1 Reduced-form evidence for level-k play ............... 92
       3.4.2 Maximum likelihood estimation of a structural model of level-k thinking 95
       3.4.3 Cognitive Hierarchy .................................. 98
       3.4.4 Naive level-k strategies .............................. 99
       3.4.5 Comparison to the literature ......................... 100
       3.4.6 When are players’ throws consistent with \( k_1 \)? .... 101
   3.5 Quantal Response ........................................... 104
   3.6 Likelihood Comparison ..................................... 106
   3.7 Conclusion ................................................ 107

References 109

Appendix A Appendix to Chapter 1 116
   A.1 Proofs Omitted from the Text .............................. 116
   A.2 Simulation Details ......................................... 126

Appendix B Appendix to Chapter 2 128
   B.1 Deriving GePP ............................................. 128
   B.2 Cournot ..................................................... 128
   B.3 Conjectural variations examples ........................... 129
   B.4 Taylor Series Error Term .................................. 131
   B.5 Calculating \( \frac{\partial \theta}{\partial P} \) ............................... 132

Appendix C Appendix to Chapter 3 134
Acknowledgments

I'd like to thank my co-authors – Scott Kominers, Glen Weyl, Dimitris Batzilis, Steven Levitt, John List and Jeffrey Picel – for pushing me and allowing me to push them. I am deeply grateful to my dissertation committee – David Cutler, Roland Fryer, Edward Glaeser, Jerry Green, Nathan Hendren, and David Laibson – for their wisdom and support. I have also benefited greatly during my graduate research from the advice and feedback of Gary Becker, Steven Durlauf, Adam Jaffe, Scott Kominers, Steven Levitt, John List, and Glen Weyl.

This research benefited from the helpful comments of Tim Bresnahan, Colin Camerer, Cory Capps, Dennis Carlton, Pierre-André Chiappori, Raj Chetty, Vince Crawford, William Diamond, Arnaud Dupuy, Joe Farrell, Michal Fabinger, Alexander Frankel, Alfred Galichon, Lars Hansen, James Heckman, Teck Ho, Sabrina Howell, Stephanie Hurder, Peter Klibanoff, Fuhito Kojima, Kai-Uwe Kuhn, Jonathan Libgober, Robert McCann, Serge Moresi, Stephen Morris, Kevin Murphy, Yusuke Narita, Charles Nathanson, Derek Neal, Alexandru Nichifor, Kathryn Peters, Toan Phan, Guillaume Pouliot, Philip Reny, Markus Reisinger, Alvin Roth, Steve Salop, Larry Samuelson, Florian Scheuer, Dick Schmalensee, James Schummer, Carl Shapiro, Robert Shimer, Aloysius Siow, Tayfun Sönmez, Jean Tirole, Daichi Ueda, Rakesh Vohra, Rui Wang, Greg Werden and numerous seminar and conference participants.

Thanks to Eric Andersen, Dhiren Patki, Toan Phan, Daichi Ueda, and Rui Wang for their excellent research assistance and to the Roshambull team for sharing their data. I am also grateful for the financial support I received from the Harvard University Economics Department, a NSF Graduate Research Fellowship, a Terence M. Considine Fellowship in Law and Economics, a Bradley Fellowship, the Lab for Economic Applications and Policy, the Institute for Quantitative Social Sciences, and the Becker Friedman Institute at the University of Chicago.

Many thanks to my friends, colleagues, my brother, and especially my parents; you have been an invaluable source of support and distraction during the past six years.
Chapter 1: Taxation in Matching Markets

High taxes might not cause CEOs to quit or reduce their hours, but raising taxes might affect the sorting of CEOs across firms. Taxes reduce large transfers more than small ones; hence, high taxes can diminish the extent to which productivity differences are reflected in wages. Thus, under high taxes, CEOs may choose the firms they like working for instead of the ones at which they are most productive. This inefficient allocation of CEOs to firms may not be apparent in the set of workers employed (or the hours that they work) if there is horizontal heterogeneity in the market – i.e. if CEOs disagree about the desirability of different jobs and firms disagree about the desirability of CEOs. Analogous re-sorting of workers can happen at all levels of employment so long as there is not perfect correlation between idiosyncratic job productivity and idiosyncratic tastes.

We analyze the impact of taxation on matching in markets with flexible preference heterogeneity. In these markets, match efficiency depends crucially on the assignment of agents to partners – not just on the set of agents who are matched in equilibrium. A classical economic intuition holds, but with a caveat: raising taxes always increases equilibrium deadweight loss in markets where agents on one side of the market do not match unless they receive positive wages; however, raising taxes can decrease the deadweight loss in fully general matching markets. Lowering taxes may decrease match efficiency because an agent

---

1Co-authored with Scott Kominers

2Most labor markets have wages flowing from firms to workers, though there may be internships that workers would pay to get. There are other, more balanced, matching markets where it may be more reasonable to think of transfers flowing in both directions. Most students pay for college, but a few are given
can transfer enough to “buy” an inefficient partner and then when taxes are lowered further he is “bought back” by his efficient partner.

Even in markets where taxes have a monotonic effect on welfare, there can be deadweight loss on the allocative margin – inefficiencies in the assignment of match partners – even if there are neither intensive nor extensive margin effects. Moreover, the change in taxable income is not a sufficient statistic for the welfare loss from taxation.

In our framework, agents have heterogeneous rankings of potential match partners and may make transfer payments to their partners. Transfers may be “taxed,” causing some of each payment to be taken from the agents.\footnote{We do not explicitly model the central authority that collects the tax. Our welfare analysis focuses on total match utility, implicitly assuming that the social value of tax revenue equals the private value.} In the case of a proportional tax $\tau$, an agent receives fraction $(1 - \tau)$ of the amount his partner gives up (see Section 1.2);\footnote{In Section 1.3 we look at lump sum transfers, where a fixed amount $f$ is subtracted from (non-zero) transfers living-stipends or free room and board in order to induce them to attend.}. Taxation lowers the value of transfers, causing agents to prefer match partners that provide higher individual-specific match utilities over those offering higher transfers; for example, with taxation, a worker may switch to a firm he happens to enjoy, but where he is less productive. The tax reduces the firms’ ability to compensate workers for the disutility of jobs where they are more productive, thereby distorting away from efficient matching.

The matching distortion we identify differs from the well-known effects of taxation on intensive and extensive labor supply; it affects the allocation of workers to firms without necessarily changing the provision of labor and it is not fully captured by the elasticity of taxable income. Also, matching distortions arise even in markets with frictionless search, and thus differ from the well-known effects of search costs on matching efficiency and of taxation
on search behavior.\(^5\)

Although our results are presented in the language of labor markets, they also have implications for other matching markets. Some transfers may be non-monetary and therefore may not be valued equally by givers and receivers: colleges may offer free housing to scholarship students, which may cost them more to provide than students’ willingness to pay. Marriage markets also often have in-kind transfers: it may be the case that a woman may value receiving a gift less than it costs her husband (in time and money) to give it.\(^6\) Taxation can be reinterpreted as representing the frictions or loss on in-kind transfers. In college admissions and marriage markets positive transfers may flow in both directions. Our non-monotonicity result implies that it is non-trivial to predict the sign of efficiency consequences of a reduction in transfer frictions.

Of the vast literature on taxation, our work is most closely related to the research on the effect of taxation on workers’ occupational choices (e.g., Parker (2003); Sheshinski (2003); Powell and Shan (2012); Lockwood et al. (2013)). However, this prior work only reflects part of the matching distortion because it does not model the two-sidedness of the market. If workers and firms both have heterogeneous preferences over match partners, then matching distortions can reduce productivity even without causing an aggregate shift in workers from one firm (or industry) to another.

Our approach is also related to the literature on taxation in Roy models (e.g., Rothschild and Scheuer (2012); Boadway et al. (1991)). The utility that a manager or firm in our model derives from a worker could be thought of the productivity of the worker in that firm or sector. However, most Roy models assume that workers earn their marginal product. When preferences are heterogeneous, firms may keep some of the productivity surplus – if a worker...


\(^6\)A similar idea is modeled by Arcidiacono et al. (2011), who treat sexual activity as an imperfect transfer from women to men in the context of adolescent relationships.
is more productive at one firm than at any other, that firm need not pay the worker his full productivity in equilibrium. Explicitly modeling firms allows for the possibility of taxation affecting the share of output that workers receive as wages.

Our model of matching with imperfect transfers provides a link between the canonical models of matching with and without transfers: Absent taxation, our framework is equivalent to matching with perfect transfers (e.g., Koopmans and Beckmann (1957); Shapley and Shubik (1971); Becker (1974)); under 100% taxation, it corresponds to the standard model of matching without transfers (e.g., Gale and Shapley (1962); Roth (1982)). Thus, the intermediate tax levels we consider introduce a continuum of models between the two existing, well-studied extremes.

While prior work has analyzed frameworks that can embed our intermediate transfer models (Crawford and Knoer (1981); Kelso and Crawford (1982); Quinzii (1984); Hatfield and Milgrom (2005)), it has focused on the structure of the sets of stable outcomes within (fixed) models and has not looked at how the efficiency of stable outcomes changes across transfer models. It is therefore unable to analyze the effect of taxation. Legros and Newman (2007) do examine outcome changes across transfer models, but they use one-dimensional agent types and therefore have limited preference heterogeneity.

The remainder of the chapter is organized as follows: Section 1.1 introduces our general model. Sections 1.2 and 1.3 analyze the cases of proportional and lump sum taxation, respectively. Section 1.4 discusses structural properties common to both models. Section 1.5 concludes. All proofs are presented in Appendix A.1.

1.1 General Model

Before introducing our models of taxation, we describe our underlying matching framework.

7This is true even if firms are price-takers. The idiosyncratically high productivity at a firm means that firm’s presence increases the surplus in the market; the firm gets to keep some of the marginal surplus.
We refer to agents on one side of the market as *managers*, denoted \( m \in M \); we refer to agents on the other side *workers*, denoted \( w \in W \). Our notation and language are also consistent with modeling marriage markets.

Each agent \( i \in M \cup W \) derives utility from being matched to agents on the other side of the market. We denote these *match utilities* by \( \alpha^Y_m \) and \( \gamma^m_w \), with \( \alpha^Y_m \) denoting the utility \( m \in M \) obtains from matching with the set of workers \( Y \subseteq W \) and \( \gamma^m_w \) denoting the utility \( w \in W \) obtains from matching with manager \( m \in M \). Without loss of generality, we normalize the utility of being unmatched (an agent’s reservation value) to 0, setting \( \alpha^m_m = \gamma^w_w = 0 \) for all \( m \in M \) and \( w \in W \). In the labor market context, \( \alpha^Y_m \) may be the productivity of the set of workers \( Y \) when employed by manager \( m \) and \( \gamma^m_w \) may be the utility or disutility worker \( w \) gets from working for \( m \).

Note that it is possible for workers to disagree about the relative desirabilities of potential managers and for managers to disagree about the relative values of potential workers. We impose no structure on workers’ match utilities and only impose enough structure on managers’ preferences to ensure the existence of equilibria. For example, the match utilities could be random draws or may result from an underlying utility function in which agents have multi-dimensional types and preferences. To ensure existence, we assume that managers’ preferences satisfy the standard Kelso and Crawford (1982)/Hatfield and Milgrom (2005) *substitutability* condition: the availability of new workers cannot make a manager want to hire a worker he would otherwise reject.

A *matching* \( \mu \) is an assignment of agents such that each manager is either matched to himself (unmatched) or matched to a set of workers who are matched to him. Denoting the

---

8Although it may seem that \( \alpha^Y_m \) should be positive and \( \gamma^m_w \) should be negative, for our general analysis we do not make sign assumptions. That is, we allow for the possibility of highly demanded internships and for counterproductive employees.

9Substitutability plays no role in our analysis other than ensuring, through appeal to previous work (Kelso and Crawford (1982)), that equilibria exist. Thus, we leave the formal discussion of the substitutability condition to the Appendix.
power set of \( W \) by \( \varphi(W) \), a matching is then a mapping \( \mu \) such that

\[
\mu(m) \in (\varphi(W) \cup \{m\}) \quad \forall m \in M,
\]

\[
\mu(w) \in (M \cup \{w\}) \quad \forall w \in W,
\]

with \( w \in \mu(m) \) if and only if \( \mu(w) = m \).

We allow for the possibility of (at least partial) transfers between matched agents. We denote the transfer from \( m \) to \( w \) by \( t^{m \to w} \in \mathbb{R} \); if \( m \) receives a positive transfer from \( w \), then \( t^{m \to w} < 0 \). A transfer vector \( t \) identifies (prospective) transfers between all manager–worker pairs, not just between those pairs that are matched. We also include in the vector \( t \) “transfers” \( t^{i \to i} \) for all agents \( i \in M \cup W \), with the understanding that \( t^{i \to i} = 0 \). For notational convenience, we denote by \( t^{m \to Y} \) the total transfer from manager \( m \) to workers in \( Y \):

\[
t^{m \to Y} \equiv \sum_{w \in Y} t^{m \to w}.
\]

In the presence of taxation, a worker might not receive an amount equal to that which his match partner gives up; in general, a (weakly increasing) transfer function \( \xi(\cdot) \) converts managers’ transfer payments into the amounts that workers receive, post-tax, with \( \xi(t^{m \to w}) \leq t^{m \to w} \). For all our transfer functions, we use the convention that \( \xi(t^{w \to w}) = 0 \) for all \( w \in W \). With these conventions, we have the following lemma.

**Lemma 1.** For a given matching \( \mu \) and transfer vector \( t \), the sum of transfers managers pay to their match partners equals the sum of the transfers paid by workers’ match partners,

\[
\sum_{m \in M} t^{m \to \mu(m)} = \sum_{m \in M} \sum_{w \in \mu(m)} t^{m \to w} = \sum_{w \in W} t^{\mu(w) \to w} \leq \sum_{w \in W} \xi(t^{\mu(w) \to w}). \tag{1.1}
\]

An *arrangement* \([\mu; t]\) consists of a matching and a transfer vector.\(^\text{10}\) We assume that agent utility is quasi-linear in transfers and that agents only care about their own match

---

\(^\text{10}\)Here we use the term “arrangement” instead of “outcome” for consistency with the matching literature (e.g., Hatfield et al. (2013)), which uses the latter term when the transfer vector only includes transfers between agents who are matched to each other.
partner(s). With these assumptions, the utility values of arrangement \([\mu; t]\) for manager \(m \in M\) and worker \(w \in W\) are

\[
\begin{align*}
  u^m([\mu; t]) &\equiv \alpha^\mu_m(m) - t^{m \rightarrow \mu(m)}, \\
u^w([\mu; t]) &\equiv \gamma^\mu_w(w) + \xi(t^{\mu(w) \rightarrow w}).
\end{align*}
\]

Note that the both the match utilities and the transfers may be either positive or negative. The utility of a worker \(w \in W\) depends on the transfer function \(\xi(\cdot)\).

Our analysis focuses on the arrangements that are stable, in the sense that no agent wants to deviate.

**Definition.** An arrangement \([\mu; t]\) is stable given transfer function \(\xi(\cdot)\) if the following conditions hold:

1. Each agent (weakly) prefers his assigned match partner(s) (with the corresponding transfer(s)) to being unmatched, that is,

\[
  u^i([\mu; t]) \geq 0 \quad \forall i \in M \cup W.
\]

2. Each manager (weakly) prefers his assigned match partners (with the corresponding transfers) to any alternative set of workers (with the corresponding transfers), that is,

\[
  u^m([\mu; t]) = \alpha^\mu_m(m) - t^{m \rightarrow \mu(m)} \geq \alpha^\mu_Y - t^{m \rightarrow Y}, \quad \forall m \in M \text{ and } Y \subseteq W;
\]

and each worker (weakly) prefers his assigned match partner (with the corresponding transfer) to any alternative manager (with the corresponding transfer), that is,

\[
  u^w([\mu; t]) = \gamma^\mu_w(w) + \xi(t^{\mu(w) \rightarrow w}) \geq \gamma^\mu_m + \xi(t^{m \rightarrow w}), \quad \forall w \in W \text{ and } m \in M.
\]

A matching \(\mu\) is stable given transfer function \(\xi(\cdot)\) if there is some transfer vector \(t\) such that the arrangement \([\mu; t]\) is stable given \(\xi(\cdot)\); in this case \(t\) is said to support \(\mu\) (given \(\xi(\cdot)\)).

Arguments of Kelso and Crawford (1982) show that the stability concept we use is
equivalent to the other standard stability concept of matching theory, which rules out the possibility of “blocks” in which groups of agents jointly deviate from the stable outcome (potentially adjusting transfers).\textsuperscript{11} The assumption of substitutable preferences ensures that at least one stable arrangement always exists.\textsuperscript{12}

In analyzing stable arrangements we focus on the \textit{total match utility} of the match $\mu$, defined as

$$\mathcal{U}(\mu) \equiv \sum_{m \in M} \alpha_{\mu(m)}^m + \sum_{w \in W} \gamma_{\mu(w)}^w.$$ 

We do not model the institution imposing the tax; as we focus on match utility, our analysis is most relevant to the case where the social value of tax revenue equals the private value.

\textbf{Definition.} We say that a matching $\hat{\mu}$ is \textit{efficient} if it maximizes total match utility among all possible matchings, i.e. if $\mathcal{U}(\hat{\mu}) \geq \mathcal{U}(\mu)$ for all matchings $\mu$.\textsuperscript{13}

Some of our analysis focuses on markets in which workers have nonpositive valuations for matching, so that they will only match if paid positive “wage” transfers. Formally, we say that a market is a \textit{wage market} if

$$\gamma_{mw}^m \leq 0$$

for all $w \in W$ and $m \in M$; it is a \textit{strictly positive wage market} if the inequality in (1.2) is \textit{strict} for all $w \in W$ and $m \in M$. The existence of internships notwithstanding, most labor markets can be reasonably modeled as wage markets.

\textsuperscript{11}Our stability concept is defined in terms of arrangements; the block-based definition is defined only in terms of a matching and the transfers \textit{between matched partners}. Kelso and Crawford (1982) used the term \textit{competitive equilibrium} for the former concept and used the \textit{core} to refer to the latter.

\textsuperscript{12}Results of Kelso and Crawford (1982) guarantee the existence of a stable arrangement in our framework. Details are provided in the Appendix.

\textsuperscript{13}An alternative welfare measure would be \textit{total agent utility}, i.e. total match utility minus total tax revenue. However, while government expenditures may not always be valued dollar-for-dollar, including government revenue in welfare is typically considered a better approximation than assigning it no value (Mas-Colell et al., 1995). Moreover, total agent utility depends on the transfer vector; as there are frequently many transfer vectors supporting a given stable match, total agent utility is not typically well-defined, even fixing a given stable match and/or tax function. The possibility of non-monotonicities can easily be shown to extend to agent utility.
For simplicity, we set our illustrative examples in *one-to-one matching markets*, in which each manager matches to at most one worker. For such markets, we abuse notation slightly by only specifying match utilities for manager–worker pairs and writing \( w \) in place of the set \( \{w\} \) (e.g., \( \alpha^w_m \) is denoted \( \alpha^w_{m} \)).

### 1.2 Proportional Taxation

First we analyze proportional (linear) taxation systems, of the type used in some US states and dozens of countries around the world. These taxes take the form of a fixed percentage deduction of each agent’s income. Formally, under proportional tax \( \tau \), if an agent pays \( p \), then his partner receives \((1 - \tau)p\). The associated transfer function \( \xi^\text{prop}_\tau(\cdot) \) is

\[
\xi^\text{prop}_\tau(t^m \rightarrow w) \equiv \begin{cases} 
(1 - \tau)t^m \rightarrow w & t^m \rightarrow w \geq 0 \\
\frac{1}{(1-\tau)}t^m \rightarrow w & t^m \rightarrow w < 0.
\end{cases}
\]

Figure 1.1 illustrates the transfer function \( \xi^\text{prop}_\tau(\cdot) \) for different tax rates \( \tau \).

![Figure 1.1: Transfer function \( \xi^\text{prop}_\tau(\cdot) \).](image)

If an arrangement \([\mu; t]\) or matching \( \mu \) is stable given \( \xi^\text{prop}_\tau(\cdot) \), then we say it is *stable under tax* \( \tau \). We analyze how the set of stable matchings changes as \( \tau \) decreases from 1 to 0.

The case \( \tau = 1 \) corresponds to the standard Gale and Shapley (1962) setting in which
transfers are not allowed\textsuperscript{14}, so inefficient matchings may be stable. When $\tau = 0$, by contrast, only efficient matchings are stable (see, e.g., Shapley and Shubik (1971); Hatfield et al. (2013)). Given this, one might expect that as the tax rate $\tau$ decreases, the match utilities of stable matchings should always (weakly) increase. Unfortunately, a simple example shows that this is not true in general.

### 1.2.1 Possible Inefficiencies of Tax-Reduction in General Markets

\textbf{(a) Match Utilities}

\begin{align*}
\alpha_{m_1}^{w_1} \gamma_{m_1}^{w_1} &= (0, 200) \\
\alpha_{m_1}^{w_2} \gamma_{m_2}^{w_1} &= (100, -8)
\end{align*}

\textbf{(b) Matching without Transfers ($\tau = 1$)}

\begin{align*}
(0, 200) \\
(100, -8)
\end{align*}

\textbf{(c) Matching with Perfect Transfers ($\tau = 0$)}

\begin{align*}
(101, 99) \\
(100, -8) \\
\mu = 0
\end{align*}

\textbf{(d) Matching with Tax ($\tau = .8$)}

\begin{align*}
(40, 0) \\
(50, 2) \\
\xi_{\tau}(t) = -200 \\
\xi_{\tau}(t) = 10
\end{align*}

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{example1.png}
\caption{Example 1 – Non-monotonicity under a proportional tax on transfers.}
\end{figure}

Note: Utilities, net of transfers, are above the lines (manager’s, worker’s). Possible supporting transfers (when applicable) are below the lines. Solid lines indicate the stable matching.

**Example 1.** Consider a one-to-one market with one manager, $M = \{m_1\}$, two workers, $W = \{w_1, w_2\}$, and match utilities as pictured in Figure 1.2a. Worker $w_1$ receives high utility

\textsuperscript{14}When $\tau = 1$, the set of stable matchings is the same as in the case that transfers are not allowed. The associated arrangements are not exactly the same, however, because the supporting transfer vectors need not be equal to $0$. However, if $\mu$ is stable when $\tau = 1$, then there is a transfer vector $t$ supporting $\mu$ such that $t^{m \rightarrow w} = 0$ for all $m \in M$ and $w \in \mu(m)$; the arrangement $[\mu; t]$ therefore replicates the utilities that arise under $\mu$ when transfers are not allowed.
from matching with $m_1$. Manager $m_1$ is indifferent towards worker $w_1$ and receives moderate utility from matching with $w_2$. Worker $w_2$ has a mild preference for being unmatched, rather than matching with $m_1$.

We can think of $w_1$ as an intern who would not be very productive in working for $m_1$, but would learn a lot; $w_2$ represents a normal worker, who is productive but does not like working. With this interpretation, the tax represents a proportional income tax – which $m_1$ must also pay if the intern $w_1$ bribes him in exchange for a job. Alternatively, we may interpret the example in a marriage context: $w_1$ is an unremarkable woman who really wants to get married; $w_2$ is a highly desirable woman who prefers to remain single; and $m_1$ is the last man on Earth. In that case, the tax reflects the extent to which it is difficult to transfer utility between individuals within a couple.

As illustrated in Figure 1.2b, when $\tau = 1$ (or when transfers are not allowed), the only stable matching $\hat{\mu}$ has $\hat{\mu}(m_1) = w_1$. This happens to be the efficient matching; therefore, it is also stable when $\tau = 0$, as shown in Figure 1.2c. This matching yields total match utility $\mathcal{T}(\hat{\mu}) = 200$.

Figure 1.2d shows that for $\tau = .8$, an inefficient matching $\tilde{\mu}$, for which $\tilde{\mu}(m_1) = w_2$, is stable. This matching generates a total match utility $\mathcal{T}(\tilde{\mu}) = 92$. Even if $w_1$ transfers 200 – his maximal utility of matching – to $m_1$, there is a transfer $m_1$ can offer to $w_2$ that is sufficient to attract $w_2$, while still providing $m_1$ more utility than he would obtain from matching with $w_1$ (and receiving $(1 - .8)(200) = 40)$.

Not only is an inefficient matching stable under tax $\tau = .8$, but the efficient matching $\hat{\mu}$ is not stable under this tax. Indeed, the efficient matching $\hat{\mu}$ is unstable under any tax $\tau \in (.6,.9)$. For that range, $(100 - 200(1 - \tau))(1 - \tau) - 8 > 0$, so that the maximum $m_1$ can transfer to $w_2$ while still preferring $w_2$ to $w_1$ is sufficient to outweigh the disutility $w_2$ gets from matching to $m_1$.\footnote{Note that here total agent utility (match utility minus government revenue), like total match utility, can be non-monotonic. When $\tau = 1$, total agent utility is 200 (assuming they do not burn money). When $\tau = .8$ it is 52.}
While Example 1 may appear quite specialized, simulations suggest that non-monotonicities in the total match utility of stable matches as a function of $\tau$ can be relatively common. We examine simulations of a one-to-one market with twenty agents on each side of the market and match utilities independently and identically distributed according to a uniform distribution on $[-.5,.5]$. We vary the tax rate, $\tau$, from 0 to .99 in increments of .01. For each tax rate, we find the manager-optimal stable arrangement and calculate the total match utility.\footnote{If there are multiple stable arrangements, the manager-optimal arrangement is the one preferred by all managers. See Section 1.4 for a discussion of opposition of worker and manager interests when there are multiple stable arrangements.} Non-monotonicities in the total match utility of stable matchings appear in over half of the markets (55%).\footnote{There may be additional non-monotonicities that we do not observe because we cannot vary $\tau$ continuously. However, the non-monotonicities we fail to observe necessarily occur over very small ranges of $\tau$, as we observe all non-monotonicities that persist over values of $\tau$ with a range of .01 or more.}

Figure 1.3 plots the total match utility as a function of the tax rate in nine randomly-selected simulation markets with non-monotonicities. These nine markets are fairly representative, in that they have relatively small losses from non-monotonicity, mostly occurring at high tax rates. Nevertheless the non-monotonicities in our simulation markets can be dramatic. Figure 1.4 presents a simulation market in which, just as in Example 1, the efficient matching is stable under full taxation ($\tau = 1$) but is unstable under a range of tax rates between 0 and 1.

Table 1.1 summarizes the non-monotonocities arising in our simulations. Row 1 shows the fraction of markets that have non-monotonicities in a given tax rate range. While the majority of non-monotonicities occur at very high tax rates, 10% of our simulation markets have non-monotonicities at tax rates below 50%. Row 2 gives the (normalized) average size of the non-monotonicities in each tax rate range. Again, we see that non-monotonicities are most significant for high tax rates. Row 3 incorporates information on the persistence of non-monotonicities by computing the fraction of the deadweight loss from taxation that is due to a non-monotonicity. This is relatively high for lower tax rates because there is less
Figure 1.3: Total match utility of a stable match in nine simulated markets.

Note: The markets presented were randomly-selected from the set of simulated markets exhibiting non-monotonicities. Each market is one-to-one and has 20 agents on each side of the market, with match utilities independently and identically distributed according to a uniform distribution on \([-0.5, 0.5]\). For each tax rate total match utility is calculated for the manager-optimal stable arrangement.

Overall, our simulations suggest non-monotonicities in the tax rate are not just artifacts of example selection. However, they also suggest that non-monotonicities are relatively rare at more realistic tax rates (\(\tau \in [0, 0.5]\)) and tend not to persist over large ranges of \(\tau\).\(^{18}\)

Although Example 1 and the simulations show that total match utility of stable matchings may decrease when the tax rate falls, an arrangement that is stable under a tax rate \(\hat{\tau}\) must improve the utility of at least one agent, relative to an arrangement that is stable under a tax rate \(\tilde{\tau} > \hat{\tau}\).

**Proposition 1.** Suppose that \([\hat{\mu}; \hat{t}]\) is stable under tax \(\hat{\tau}\), and that \([\tilde{\mu}; \tilde{t}]\) is stable under tax

\(^{18}\)Increasing the sample size does not appear to decrease the frequency or importance of non-monotonicities.
Note: The market pictured is one-to-one and has 20 agents on each side of the market, with match utilities independently and identically distributed according to a uniform distribution on $[-.5,.5]$. For each tax rate total match utility is calculated for the manager-optimal stable arrangement.

$\tilde{\tau}$, with $\tilde{\tau} > \hat{\tau}$. Then, $[\tilde{\mu}; \tilde{t}]$ (under tax $\tilde{\tau}$) cannot Pareto dominate $[\hat{\mu}; \hat{t}]$ (under tax $\hat{\tau}$).\footnote{We say that an arrangement $[\tilde{\mu}; \tilde{t}]$ under tax $\tilde{\tau}$ Pareto dominates arrangement $[\hat{\mu}; \hat{t}]$ under tax $\hat{\tau}$ if
\[
\begin{align*}
\alpha_{m}^{\tilde{\mu}}(m) - \hat{t}_{m}^{\tilde{\mu}} & \geq \alpha_{m}^{\hat{\mu}}(m) - \hat{t}_{m}^{\hat{\mu}}(m) & \forall m \in M, \\
\gamma_{w}^{\tilde{\mu}}(w) + \xi_{\tilde{\tau}}^{\text{prop}}(\hat{t}_{\tilde{\mu}}^{\hat{\mu}}(w) \rightarrow w) & \geq \gamma_{w}^{\hat{\mu}}(w) + \xi_{\hat{\tau}}^{\text{prop}}(\hat{t}_{\hat{\mu}}^{\hat{\mu}}(w) \rightarrow w) & \forall w \in W,
\end{align*}
\]}

To see the intuition behind Proposition 1, we consider the case in which $\tilde{\tau} = 1$ and choose $\tilde{t} = 0$: If $[\tilde{\mu}; \tilde{t}]$ (under tax $\tilde{\tau} = 1$) Pareto dominates $[\hat{\mu}; \hat{t}]$ (under tax $\hat{\tau}$), then every manager $m \in M$ (weakly) prefers $\tilde{\mu}(m)$ to $\hat{\mu}(m)$ with the transfer $\hat{t}_{m}^{\tilde{\mu}}(m)$. But then, because $[\tilde{\mu}; \tilde{t}]$

\[19\text{To see this, we first note that under tax } \tilde{\tau} = 1, \text{ an arrangement with transfers of 0 among match partners Pareto dominates any other arrangement associated to the same matching. Thus, the transfers between match partners under } [\tilde{\mu}; \tilde{t}] \text{ can be assumed to be 0. Then, the comparison between } [\tilde{\mu}; \tilde{t}] \text{ (under tax } \tilde{\tau} = 1) \text{ and } [\hat{\mu}; \hat{t}] \text{ (under tax } \hat{\tau}) \text{ amounts to a comparison of agents' match utilities under } \tilde{\mu} \text{ and their total utilities under } [\tilde{\mu}; \tilde{t}]\]
Table 1.1: Summary of the non-monotonicities arising in simulated markets

<table>
<thead>
<tr>
<th>Range of $\tau$</th>
<th>$[0, .25)$</th>
<th>$[.25, .5)$</th>
<th>$[.5, .75)$</th>
<th>$[.75, 1)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fraction of markets with non-monotonicity</td>
<td>0.006</td>
<td>0.088</td>
<td>0.190</td>
<td>0.394</td>
</tr>
<tr>
<td>Avg size of non-monotonicity, as fraction of range</td>
<td>0.021</td>
<td>0.066</td>
<td>0.111</td>
<td>0.140</td>
</tr>
<tr>
<td>Fraction of deadweight loss from taxation due to non-monotonicity</td>
<td>0.076</td>
<td>0.070</td>
<td>0.051</td>
<td>0.027</td>
</tr>
</tbody>
</table>

Note: The table summarizes 500 simulations of one-to-one matching markets with 20 agents on each side of the market. All agents’ match utilities are independently and identically distributed according to a uniform distribution on $[-.5, .5]$. We vary the tax rate, $\tau$, from 0 to .99 in increments of .01. For each tax rate, we find the manager-optimal stable arrangement and calculate the total match utility. Row 1 presents the fraction of markets that have non-monotonicities in a given tax rate range. Row 2 presents the average size of non-monotonicities within each range, normalized as a fraction of the (within-market) gap between the highest and lowest total stable match utilities calculated for any tax rate. Row 3 presents the average fraction of taxation deadweight loss that is due to non-monotonicity, across all markets. The deadweight loss from non-monotonicity is computed for each tax rate $\tau$ as the difference between the highest total match utility for a tax rate $\hat{\tau} \geq \tau$ and the total match utility under tax rate $\tau$; it is divided by the total deadweight loss from taxation at tax rate $\tau$, which is computed as the difference in total match utility between the efficient matching and the matching stable under tax rate $\tau$.

is stable under tax $\hat{\tau}$,

$$\hat{\alpha}_m^{\hat{\mu}(m)} \geq \hat{\alpha}_m^{\mu(m)} - \hat{t}_m \rightarrow \hat{\mu}(m) \geq \hat{\alpha}_m^{\tilde{\mu}(m)} - \hat{t}_m \rightarrow \tilde{\mu}(m),$$

so every $m$ must be offering a weakly positive transfer to $\tilde{\mu}(m)$ under $\hat{t}$ (that is, $\hat{t}_m \rightarrow \tilde{\mu}(m) \geq 0$). An analogous argument shows that each worker $w \in W$ must be offering a weakly positive transfer to $\tilde{\mu}(w)$ under $\hat{t}$ (that is, $\xi_{\hat{\tau}}^{\text{prop}}(\hat{t}_w \rightarrow \tilde{\mu}(w)) \leq 0$). Moreover, Pareto dominance implies that at least one manager or worker must be paying a strictly positive transfer. But then, that agent must pay a strictly positive transfer and receive a weakly positive transfer – impossible.
1.2.2 Efficiency of Tax-Reduction in Wage Markets

The preceding discussion shows that in general markets, decreasing the tax rate on transfers may decrease the total match utility of stable matchings. Our next result shows that in wage markets, these non-monotonicities do not arise – decreasing the tax rate in a wage market always makes (weakly) more efficient matchings stable.\footnote{The non-monotonicities described in Section 1.2.1 arise from transfers flowing in both directions, either simultaneously or across equilibria. As transfers are an equilibrium phenomenon, requiring that transfers flow in one direction does not directly correspond to conditions on the primitives of the market. However, the wage market condition we use in Theorem 1 is a sufficient condition on primitives to guarantee that transfers to flow in one direction, and thus is sufficient to rule out non-monotonicity. All the results in this section hold in any market where transfers always (across stable arrangements and tax rates) flow in one direction.}

In wage markets, payments flow from managers to workers; hence, any stable matching can be supported by a non-negative transfer vector.\footnote{There may be a supporting transfer vector where some off-path transfers (transfers between unmatched agents) are negative, but in that case there is always another supporting transfer vector that replaces those negative transfers with 0s. Our results only require the existence of a non-negative supporting transfer vector.} Thus, the transfer function $\xi^\text{prop}_\tau(\cdot)$ takes the simpler form

$$\xi^\text{prop}_\tau(t^{m\to w}) = (1 - \tau)t^{m\to w} \geq 0.$$

As all positive transfers are paid from managers to workers, there cannot be a scenario in which, as in Example 1, when the tax is reduced, a manager can transfer enough to get a worker he prefers ($w_2$), but when the tax falls more, a different worker ($w_1$) can “buy back” the manager. Our next result shows that this intuition extends to wage markets more generally.

**Theorem 1.** In a wage market with proportional taxation, a decrease in taxation (weakly) increases the total match utility of stable matchings. That is, if in a wage market, matching $\hat{\mu}$ is stable under tax $\hat{\tau}$, matching $\tilde{\mu}$ is stable under tax $\tilde{\tau}$, and $\hat{\tau} < \tilde{\tau}$, then

$$\mathcal{I}(\hat{\mu}) \geq \mathcal{I}(\tilde{\mu}).$$
To prove Theorem 1, we let \( \hat{t} \geq 0 \) and \( \tilde{t} \geq 0 \) be transfer vectors supporting \( \hat{\mu} \) and \( \tilde{\mu} \) respectively. The stability of \([\hat{\mu}; \hat{t}]\) under tax \( \hat{\tau} \) implies that

\[
\alpha_{m}^{\hat{\mu}(m)} - \hat{t}_{m}^{\hat{m} \rightarrow \hat{\mu}(m)} \geq \alpha_{m}^{\tilde{\mu}(m)} - \tilde{t}_{m}^{\tilde{m} \rightarrow \tilde{\mu}(m)},
\]

(1.3)

\[
\gamma_{w}^{\hat{\mu}(w)} + (1 - \hat{\tau})\hat{t}_{w}^{\hat{w} \rightarrow \hat{\mu}(w)} \geq \gamma_{w}^{\tilde{\mu}(w)} + (1 - \hat{\tau})\tilde{t}_{w}^{\tilde{w} \rightarrow \tilde{\mu}(w)}.
\]

(1.4)

Summing (1.3) and (1.4) across agents, applying Lemma 1, and regrouping terms, we find that

\[
\mathfrak{T}(\hat{\mu}) - \mathfrak{T}(\tilde{\mu}) = \sum_{m \in M} \left( \alpha_{m}^{\hat{\mu}(m)} - \alpha_{m}^{\tilde{\mu}(m)} \right) + \sum_{w \in W} \left( \gamma_{w}^{\hat{\mu}(w)} - \gamma_{w}^{\tilde{\mu}(w)} \right) \\
\geq \hat{\tau} \sum_{m \in M} \left( \hat{t}_{m}^{\hat{m} \rightarrow \hat{\mu}(m)} - \hat{t}_{m}^{\tilde{m} \rightarrow \tilde{\mu}(m)} \right) .
\]

(1.5)

Intuitively, since the tax change has a larger effect on larger transfers, if we had

\[
\sum_{m \in M} \left( \hat{t}_{m}^{\hat{m} \rightarrow \hat{\mu}(m)} - \hat{t}_{m}^{\tilde{m} \rightarrow \tilde{\mu}(m)} \right) < 0,
\]

then lowering the tax from \( \tilde{\tau} \) to \( \hat{\tau} \) would increase workers’ relative preference for \( \tilde{\mu} \) over \( \hat{\mu} \). Since \( \hat{\mu} \) is stable under the lower tax \( \hat{\tau} \), the difference in (1.5) must thus be positive; this implies Theorem 1.

Theorem 1 shows that the non-monotonicities observed in fully general markets in Section 1.2.1 do not arise in wage markets. To gain insight into how quickly non-monotonicity disappears as a market’s structure becomes closer to that of a wage market, we return to our simulation environment. We begin with simulations in a setting identical to that used in Section 1.2.1: one-to-one markets with all match utilities independently and identically distributed according to a uniform distribution on \([-0.5, 0.5]\). We next consider one-to-one markets with match utilities slightly imbalanced across the market: managers’ match utilities are independently and identically distributed according to a uniform distribution on \([-0.45, 0.55]\), while workers’ match utilities are independently and identically distributed according to a uniform distribution on \([-0.55, 0.45]\). We repeat this process, adjusting the match utility means
by .05 each time, to generate a series of markets ranging from our original symmetric markets to wage markets with manager utilities distributed uniformly on [0, 1] and worker utilities distributed uniformly on [−1, 0]. Figure 1.5 shows the how the fraction of markets with non-monotonocities in the match utility of the manager-optimal stable arrangement changes as the mean manager utility varies. Even fairly asymmetrical markets have moderate rates of non-monotonicities.

![Figure 1.5: Fraction of simulated markets where the total match utility is non-monotonic in the tax rate, τ.](image)

Note: For each mean manager utility level, we report the fraction of the 500 simulated markets that have a non-monotonicity in total match utility of the manager-optimal arrangement as the tax rate increases from 0 to 1. All simulated markets are one-to-one and have 20 agents on each side of the market.

Although total match utility in wage markets increases as the tax is reduced, individual utility may be non-monotonic. For example, pursuant to a tax decrease, a manager $m$ may be made worse off because his match partner is now able to receive more from some other

---

23To reduce noise in the simulation process, we use a single set of 500 baseline markets and repeatedly shift each match utility by .05 in the appropriate direction.
manger: In this circumstance, $m$ might lose his match partner to his competitor; even if $m$’s match is unchanged, his total utility may decrease because he is forced to increase his transfer to compensate for a competitor’s increased offer.

Individual managers’ match utilities may decrease with a decrease in the tax rate, but the sum of workers’ match utilities must decrease.

**Proposition 2.** In a wage market with proportional taxation, if a matching $	ilde{\mu}$ is stable under tax $\tilde{\tau}$, and a matching $\hat{\mu}$ is stable under tax $\hat{\tau} < \tilde{\tau}$, then workers’ aggregate match utility must be (weakly) higher under $\tilde{\mu}$ than under $\hat{\mu}$. That is,

$$\sum_{w \in W} \gamma_{\tilde{\mu}}(w) \geq \sum_{w \in W} \gamma_{\hat{\mu}}(w).$$

The logic is that in order for a less efficient match to be stable at the higher tax rate, it must be that workers prefer that match – and managers cannot lure workers to a more efficient match because of the high tax. As the tax rate decreases, managers’ ability to make transfers to workers increases, and so the weight put on their match utilities relative to workers’ match utilities increases. Absent taxation, stable matches maximizes the sum of match utilities with equal weight on managers and workers; under taxation, stable matches still maximize the sum of match utilities, but with different weights.

**Proposition 3.** In a wage market with proportional taxation, if a matching $\tilde{\mu}$ is stable under tax $\tilde{\tau}$, then $\tilde{\mu}$ is a matching that maximizes the sum of worker match utilities plus $(1 - \tilde{\tau})$ times the sum of manager match utilities:

$$\tilde{\mu} \in \arg \max_{\{\mu\}} \left[ (1 - \tilde{\tau}) \sum_{m \in M} \alpha_{\mu}(m) + \sum_{w \in W} \gamma_{\mu}(w) \right].$$

Note that Propositions 2 and 3 do not imply that lower taxes necessarily make workers worse off: workers might receive transfers sufficiently high as to more than compensate for their lower match utilities.

Finally, we show that if two distinct matchings $\hat{\mu}$ and $\tilde{\mu}$ are both stable under tax $\tau$, then either managers and workers must disagree as to which matching is preferred, or both groups
must be indifferent between the two matchings. This is a consequence of the following more general result.

**Proposition 4.** In a wage market with proportional taxation, if two distinct matchings  and  are both stable under tax , then

\[
\sum_{w \in W} \left( \gamma_{\hat{\mu}}(w) - \gamma_{\tilde{\mu}}(w) \right) = (1 - \tau) \sum_{m \in M} \left( \alpha_{\hat{\mu}}(m) - \alpha_{\tilde{\mu}}(m) \right). \tag{1.6}
\]

Thus, if the managers are not indifferent in aggregate between  and , then the only tax rate under which both  and  can be stable is

\[
\tau = 1 + \frac{\sum_{w \in W} \left( \gamma_{\hat{\mu}}(w) - \gamma_{\tilde{\mu}}(w) \right)}{\sum_{m \in M} \left( \alpha_{\hat{\mu}}(m) - \alpha_{\tilde{\mu}}(m) \right)}. \tag{1.7}
\]

For \( \tau \) as defined in (1.7) to be less than 1, the fraction in (1.7) must be negative, so that managers and workers in aggregate disagree about which matching they prefer.

In order for there to be multiple values of \( \tau \) at which two given matchings are both stable, it must be that both managers and (following (1.6)) workers are indifferent between those two matchings.

**Corollary 1.** In a wage market with proportional taxation, if there is more than one tax under which two distinct matchings  and  both are stable, then \( T(\tilde{\mu}) = T(\hat{\mu}) \).

Corollary 1 implies that for generic match utilities, there is at most one value of \( \tau \) at which two matchings  and  are both stable; in this case, since there are finitely many matchings, there is a unique stable matching under almost every tax \( \tau \).

At a tax rate \( \tau \) under which two distinct matches  and  are stable, we can renormalize the utilities and use results from matching with transfers to draw conclusions about match utilities and revenue.\(^{24}\) If we multiply all worker utilities and transfers by \( \frac{1}{1 - \tau} \), then workers’ comparison of options is unchanged – and therefore the stable matches are unaffected – but

\(^{24}\)This re-normalization, which was not in the original draft of this work, was independently originated by Ismael Mourifie and Aloysius Siow, and Arnaud Dupuy and Alfred Galichon.
the new utility function

\[ u_t^w([\mu, t]) \equiv \frac{1}{1 - \tau} u([\mu, t]) = \frac{1}{1 - \tau} \gamma_{w}^{\mu(w)} + t_{\mu(w) \rightarrow w}^{\mu(w)} \]

is quasi-linear. When worker and manager utilities are quasi-linear in the transfer, then results of Hatfield et al. (2013) show that if two matches, \( \hat{\mu} \) and \( \tilde{\mu} \) are stable, any transfer vector \( t \) that supports one also supports the other. Moreover, for a given transfer vector, \( t \), all agents are indifferent between the two arrangements:

\[ \alpha_{\tilde{\mu}(m)}^{\tilde{\mu}(m)} - t^{m \rightarrow \tilde{\mu}(m)} = \alpha_{\hat{\mu}(m)}^{\hat{\mu}(m)} - t^{m \rightarrow \hat{\mu}(m)}, \tag{1.8} \]

\[ \frac{1}{1 - \tau} \gamma_{w}^{\tilde{\mu}(w)} + t^{w \rightarrow \tilde{\mu}(w)} = \frac{1}{1 - \tau} \gamma_{w}^{\hat{\mu}(w)} + t^{w \rightarrow \hat{\mu}(w)} \tag{1.9} \]

Multiplying (1.9) by \((1 - \tau)\) we get:

\[ \gamma_{w}^{\tilde{\mu}(w)} + (1 - \tau) t^{w \rightarrow \tilde{\mu}(w)} = \gamma_{w}^{\hat{\mu}(w)} + (1 - \tau) t^{w \rightarrow \hat{\mu}(w)}, \tag{1.10} \]

Summing (1.8) and (1.10) across agents gives the difference in total match utility:

\[ T(\tilde{\mu}) - T(\hat{\mu}) = \tau \sum_{m \in M} t^{m \rightarrow \tilde{\mu}(m)} - \tau \sum_{m \in M} t^{m \rightarrow \hat{\mu}(m)}. \]

These results are summarized in the following proposition.

**Proposition 5.** In a wage market, if two matches \( \tilde{\mu} \) and \( \hat{\mu} \) are both stable under tax rate \( \tau \), then:

1. Any transfer vector that supports either \( \tilde{\mu} \) or \( \hat{\mu} \) also supports the other.

2. For any transfer vector \( t \) that supports \( \tilde{\mu} \) and \( \hat{\mu} \) all agents are indifferent between \([\tilde{\mu}, t]\) and \([\hat{\mu}, t]\)

3. The difference in total match utility between \( \tilde{\mu} \) and \( \hat{\mu} \) equals the difference in revenue \([\tilde{\mu}, t]\) and \([\hat{\mu}, t]\) for every transfer vector \( t \) supporting \( \tilde{\mu} \) and \( \hat{\mu} \).

Unfortunately, the third part of Proposition 5 – that changes in revenue sometimes correspond to changes in utility – is very limited. As the tax rate changes, transfers will
change even when the underlying match does not change (so there is no change in total match utility). Also, even at the tax rate where multiple matches are stable, there may be multiple supporting transfer vectors and the revenue between \([\hat{\mu}, \hat{t}]\) and \([\tilde{\mu}, \tilde{t}]\) does not tell us anything about the difference in total match utility between \(\hat{\mu}\) and \(\tilde{\mu}\).

### 1.2.3 Deadweight Loss

In addition to causing some workers not to work, taxation generates deadweight loss by changing the matching of workers to firms. Thus, workers’ decisions on where to work affect managers’ productivities and the opportunities available to other workers. These externalities mean that, unlike in the framework of Feldstein (1999), the deadweight loss cannot be calculated from the change in taxable income.\(^{25}\)

Using the Feldstein (1999) formula,

\[
\frac{d\text{DWL}}{d\tau} = \tau \frac{d\text{Taxable Income}}{d\tau},
\]

can generate substantial bias in our setting. Figure 1.6 shows the average actual deadweight loss and the average estimated deadweight loss for simulated markets. Each market, has 25 agents on each side. Match utilities are drawn i.i.d. from a lognormal distribution. The mean and variance are based on estimates of the wage distribution from the work of Woodcock (2008), one of the few papers on wage hedonics that attempts to estimate match-specific effects in addition to worker and firm fixed-effects. As the figure shows, the Feldstein (1999) estimated deadweight loss can be very different from the actual deadweight loss and for some tax values is actually negative. At the manager-optimal stable arrangements, managers pay workers the minimum necessary; when the tax increases, they still need to pay workers a comparable post-tax wage; hence, the pre-tax wage actually increases, leading to a negative estimate of the deadweight loss.

\(^{25}\)See, e.g. Chetty (2009a) for other conditions under which the Feldstein (1999) formula does not hold.
Figure 1.6: Actual and Estimated Deadweight Loss at the Manager-Optimal Stable Arrangement in 192 simulated markets.

Note: The graphs show the average across 192 simulations of one-to-one matching markets with 25 agents on each side of the market. All agents’ match utilities are independently and identically log-normally distributed. See Appendix A.2 for details. We vary the tax rate, \( \tau \), from 0 to 1 in increments of .02. For each tax rate, we find the manager-optimal stable arrangement and calculate the actual deadweight loss (relative to \( \tau = 0 \)) and the deadweight loss estimated based on the formula \( \frac{d\text{DWL}}{d\tau} = \tau \frac{dTaxable\ Income}{d\tau} \).

job a lot more. This second effect is not captured by the change in taxable income, leading the estimate to be potentially biased upward. Adding correlation to the match utilities or looking at the worker optimal stable match shows that in some markets, for some tax rates, the Feldstein (1999) estimated deadweight loss overstates the deadweight loss from taxation.

1.3 Lump Sum Taxation

While not typically phrased in the exact language of taxation, lump sum taxes are present throughout labor markets. They might take the form of fixed costs per employee (e.g., employee health care costs) or costs of entering employment (e.g., licensing requirements). In
the marriage market context, lump sum taxes can take the form of marriage license fees or tax penalties for marriage.

1.3.1 Lump Sum Taxation of Transfers

We first consider a lump sum tax that is levied only on (nonzero) transfers between match partners. Such a lump sum tax on transfers, $f$, corresponds to the transfer function

$$
\xi_{lump}^f(t^m \rightarrow w) \equiv \begin{cases} 
  t^m \rightarrow w - f & t^m \rightarrow w \neq 0 \\
  t^m \rightarrow w & t^m \rightarrow w = 0.
\end{cases}
$$

Figure 1.7 shows this transfer function. Under this tax structure, the case $f = 0$ corresponds to the standard (Shapley and Shubik (1971)) model of matching with transfers and the case $f = \infty$ corresponds to (Gale and Shapley (1962)) matching without transfers.

![Figure 1.7: Transfer function $\xi_{lump}^f(\cdot)$.](image)

We say that an arrangement or matching is stable under lump sum tax $f$ if it is stable given transfer function $\xi_{lump}^f(\cdot)$.

A lump sum tax on transfers has an extensive margin effect that makes being unmatched more attractive relative to matching with a transfer. In non-wage markets, a lump sum tax on transfers can have significant effects on market outcomes. An alternative approach to lump sum taxation, which we discuss in the next section, imposes a flat fee on all matches.

Since it is difficult to observe transfers in non-wage markets, such as marriage markets, it is somewhat
tax on transfers can also encourage matchings in which transfers are unnecessary. As our
next example illustrates, this second distortion can cause the total match utility of stable
matchings to be non-monotonic in the size of the lump sum tax.

Consider a one-to-one market with two managers – $M = \{m_1, m_2\}$ – two workers –
$W = \{w_1, w_2\}$ – and match utilities as pictured in Figure 1.8a. Worker $w_1$ likes $m_1$ – who
has a strong preference for $w_2$ – but $w_2$ prefers $m_2$. When transfers are not allowed (or when
there is a high lump sum tax on transfers, $f \geq 18$), the only stable matching is the matching
$\mu_1$ in which $\mu_1(m_1) = w_1$ and $\mu_1(m_2) = w_2$, as shown in Figure 1.8b. This matching yields
total match utility of $T(\mu_1) = 22$.

Example 2. When the lump sum tax is lowered to $f = 12$, only the matching $\mu_2$ is stable,
where $\mu_2(m_1) = w_2$ and $w_1$ and $m_2$ are unmatched; this matching gives a total match utility
$T(\mu_2) = 19$, as shown in Figure 1.8d. When $f = 12$, the tax is low enough that $m_1$ can
convince $w_2$ to match with him, but not low enough for $w_1$ to hold onto $m_1$ when he has the
option of matching with $w_2$ (or $m_2$ to hold onto $w_2$). Lowering the lump sum tax from 20
to 12 decreases the total match utility of the stable matching and decreases the number of
agents matched.

Just as in Section 1.2, we use simulations to confirm that Example 2 is not an exceptional
case. We return to the 500 randomly drawn one-to-one markets presented in Section 1.2, and
consider lump sum taxes varying from 0 to 1 in increments of .01. We find that match utility
is non-monotonic in the lump sum tax in 61% of our simulated markets.

Figure 1.9 plots the total match utility of the manager-optimal stable match as a function
of the tax rate in nine randomly-selected simulation markets with non-monotonicities under

---

28 To see this, consider the case of balanced one-to-one matching markets. In such markets, lump sum
taxes on transfers promotes pairing $(m, w)$ in which the match utility $\alpha_m^w + \gamma_m^w$ is evenly distributed between
the two partners ($\alpha_m^w \approx \gamma_m^w$), so that transfers are unnecessary.
(a) Match Utilities

\[(\alpha_{w_1}^{m_1}, \gamma_{w_1}^{m_1}) = (1, 10)\]

\[(\alpha_{w_2}^{m_1}, \gamma_{w_1}^{m_2}) = (20, -1)\]

\[(\alpha_{w_2}^{m_1}, \gamma_{w_2}^{m_2}) = (10, 1)\]

(b) Matching without Transfers

\[(f = \infty)\]

\[(1, 10)\]

\[(20, -1)\]

(c) Matching with Perfect Transfers

\[(f = 0)\]

\[(11, 0)\]

\[(10, 9)\]

\[(1, 10)\]

\(t = -10, 10, 9\)

(d) Matching with Lump Sum Tax

\[(f = 12)\]

\[(-1, 0)\]

\[(18, -2)\]

\[(-10, -1)\]

\(t = 18, 10, 2\)

\(t = 18, 10, 2\)

Figure 1.8: Example 2 – Non-monotonicity under a lump sum tax on transfers.

Note: Utilities, net of transfers, are above the lines (manager’s, worker’s). Possible supporting transfers (when applicable) are below the lines. Solid lines indicate the stable matching.
Figure 1.9: Total match utility of a stable match in nine simulated markets.

Note: The markets presented were randomly-selected from the set of simulated markets exhibiting non-monotonicities. Each market is one-to-one and has 20 agents on each side of the market, with match utilities independently and identically distributed according to a uniform distribution on $[-.5,.5]$. For each tax rate, total match utility is calculated for the manager-optimal stable arrangement.

In all markets, the total match utility is unchanged for lump sum taxes above .5 because this is the maximum individual match utility. (In equilibrium there are no transfers paid when the tax is .5, so increasing the lump sum tax on transfers above .5 has no effect.)

In strictly positive wage markets, all matchings require a transfer, so a lump sum tax on transfers does not distort agents’ preferences among match partners – for a given transfer vector, if a worker prefers manager $m_1$ to $m_2$ without a tax, then that worker also prefers $m_1$ to $m_2$ under a lump sum tax. Thus, in strictly positive wage markets, the matching distortion of the lump sum tax is only on the extensive margin – the decision of whether to match – under a higher lump sum tax, fewer agents find matching desirable. This intuition is captured in the following lemma, where we use $(\mu)$ to denote the number of workers
Lemma 2. In strictly positive wage markets, reduction in a lump sum tax on transfers (weakly) increases the number of workers matched in stable matchings. That is, if matching $\tilde{\mu}$ is stable under lump sum tax $\tilde{f}$, matching $\hat{\mu}$ is stable under lump sum tax $\hat{f}$, and $\hat{f} < \tilde{f}$, then

$$\#(\hat{\mu}) \geq \#(\tilde{\mu}).$$

In non-wage markets, the conclusion of Lemma 2 is not true, in general, because distortion among match partners can dominate the extensive margin effect, as in Example 2.

As lump sum taxes do not distort among match partners in strictly positive wage markets, they can only reduce the efficiency of stable matchings in such markets by reducing the number of workers matched. This observation, when combined with Lemma 2, gives the following result.

Theorem 2. In strictly positive wage markets, a reduction in a lump sum tax on transfers (weakly) increases the total match utility of stable matchings. That is, if $\tilde{\mu}$ is stable under lump sum tax $\tilde{f}$, $\hat{\mu}$ is stable under lump sum tax $\hat{f}$, and $\hat{f} < \tilde{f}$, then

$$T(\hat{\mu}) \geq T(\tilde{\mu}).$$

Theorem 2 indicates that in strictly positive wage markets, match utility increases monotonically as lump sum taxation decreases.

Just as in the case of proportional taxation, non-monotonicity disappears as a market’s structure becomes closer to that of a wage market. Using the same set of simulation markets described in Section 1.2.2, we analyze how the fraction of markets with non-monotonicities changes as we move from symmetric markets to wage markets. Figure 1.10 shows the results. We see that a substantial amount of market asymmetry is needed before the fraction of markets with non-monotonicities drops below 50%.

In strictly positive wage markets, we can also bound the total match utility loss from a given lump sum tax.
Figure 1.10: *Fraction of simulated markets where the total match utility is non-monotonic in the lump sum tax, $f$.*

Note: For each mean manager utility level, we report the fraction of the 500 simulated markets that have a non-monotonicity in total match utility as the lump sum increases from 0 to 1. All simulated markets are one-to-one and have 20 agents on each side of the market.

**Proposition 6.** In a strictly positive wage market, let $\hat{\mu}$ be an efficient matching, and let $\tilde{\mu}$ be stable under lump sum tax on transfers $\tilde{f}$. Then,

$$0 \leq \mathcal{U}(\hat{\mu}) - \mathcal{U}(\tilde{\mu}) \leq \tilde{f} \cdot (\#(\hat{\mu}) - \#(\tilde{\mu})).$$

The intuition for Proposition 6 is that since the workers unmatched under a lump sum tax of $\tilde{f}$ have negative surplus from matching under that lump sum tax, their surplus from matching could not be more than $\tilde{f}$. So the change in total utility is less than the change in the number of unmatched workers times a maximum surplus of $\tilde{f}$ per worker.

Finally, we can show that, for a fixed limit on the number of workers matched in the presence of a lump sum tax, stable matchings in strictly positive wage markets must generate the maximal match utility possible.
Proposition 7. In a strictly positive wage market, a matching \( \bar{\mu} \) can be stable under a lump sum tax on transfers only if

\[
\bar{\mu} \in \arg \max \left \{ \mu : \#(\mu) \leq \#(\bar{\mu}) \right \} \left[ T(\mu) \right].
\]

Proposition 7 shows that a lump sum tax is an efficient way for a market designer to limit the number of matches (in strictly positive wage markets): the matchings stable under lump sum taxation have maximal utility, given the tax’s implied limit on the number of agents matched. Analogously, if a market designer wants to encourage matches, a lump-sum subsidy will maximize total match utility for a given (subsidy-induced) lower bound on the number of agents matched. For example, this suggests that if a government wants to use tuition subsidies to encourage people to go to school, then uniform tuition subsidies are more efficient than subsidies proportional to the cost of tuition.

1.3.2 Lump Sum Taxation of Matches

Some fee structures tax all pairings, rather than just those that include nonzero transfers. Such flat fees for matching can also be interpreted in the language of taxation: they correspond to the transfer function

\[
\xi_f(t^{m\rightarrow w}) \equiv t^{m\rightarrow w} - f.
\]

Figure 1.11 shows this transfer function for different levels of \( f \).

Unlike lump sum taxes on transfers, flat fees for matching never distort among match partners – even in non-wage markets. Flat fees for matching only have extensive margin effects, and thus markets with such fees are similar to strictly positive wage markets with lump sum taxes on transfers.\(^{29}\) As we show in the Appendix, the conclusions of Lemma 2, Theorem 2, and Propositions 6 and 7 always hold in markets with flat fees for matching.

\(^{29}\)Indeed, in strictly positive wage markets, lump sum taxation of transfers is equivalent to lump sum taxation of matchings because workers never match without receiving a strictly positive transfer.
1.4 Discussion

1.4.1 Importance of Taxation in Different Markets

To get some intuition for the relationship between the deadweight loss and certain market properties, we run simulations based on distribution estimates of Woodcock (2008). See Appendix A.2 for details.

Proposition 3 indicates that, in the presence of proportional taxation, increasing worker utility (i.e. decreasing the onerousness of work) has a bigger effect on welfare than increasing manager utility (increasing productivity). However, in general, increasing manager utility will increase wages, while increasing worker utility will have a smaller or possibly negative effect on wages. So if taxes are implemented to meet a revenue requirement, a higher tax rate will be needed to raise the same revenue in markets where work is less onerous relative to markets where workers are more productive. So the theory does not tell us whether, for a given revenue requirement, welfare will be higher in a market with higher manager utilities or higher worker utilities.

Figure 1.12 shows the results from three sets of simulations: a baseline market, a market
Figure 1.12: Welfare for Different Mean Utility Levels

Note: The graphs show the average across 100 simulations of one-to-one matching markets with 25 agents on each side of the market. All agents’ match utilities are independently and identically log-normally distributed. See Appendix A.2 for details. We vary the tax rate, $\tau$, from 0 to 1 in increments of 0.02 and calculate market welfare of the manager-optimal stable match.

with higher mean worker utility, and a market with higher mean manager utility. The latter two markets are calibrated such that the total welfare is equal, absent taxation. We find that, as expected, for every tax rate $\tau > 0$, average welfare is higher in the market with higher worker utility. However, Figure 1.13 shows that under a revenue requirement, average welfare is actually slightly higher in the market with higher manager utility. Thus, we see that whether improving productivity or enjoyability of work is better for welfare can depend on the specifics of the market and on whether tax rates or revenue are being held constant.$^{30}$

$^{30}$Since revenue follows a Laffer curve, each level of revenue can also be achieved by a higher tax rate with lower welfare. We focus on the lower tax rate for raising revenue. At the higher tax rate generating a given level of revenue, the two markets have similar levels of welfare.
Figure 1.13: Maximum Welfare Attainable for a Given Revenue Requirement, for Different Mean Utility Levels

Note: The graphs show the average across 100 simulations of one-to-one matching markets with 25 agents on each side of the market. All agents’ match utilities are independently and identically log-normally distributed. See Appendix A.2 for details. We vary the tax rate, $\tau$, from 0 to 1 in increments of .02 and calculate the revenue and market welfare of the manager-optimal stable match. For each market and each revenue level, we graph the maximal welfare attainable at a (manager-optimal) match that produces that level of revenue.

Heterogeneity

There are a few different ways we can think about varying the heterogeneity in a market: (1) varying the overall variance of match utilities (2) within each side of the market, varying the extent to which match utilities are determined by worker or manager fixed effects rather than idiosyncratic components and (3) across the two sides of the market, varying the correlation between worker and manager utility (for a given firm-worker pair). The effect of heterogeneity on the deadweight loss from taxation depends on the type of heterogeneity.

Increasing the variance in match utilities increases overall welfare in the market because it expands the right tail of match utilities. However, it decreases the fraction of welfare lost
due to taxation. Figure 1.14 shows the average deadweight loss and estimated deadweight loss for markets with three different variances of match utility. As the variance increases, the estimate based on taxable income does a better job at approximating the deadweight loss in the market.

![Figure 1.14: Actual and Estimated Deadweight Loss as a Fraction of Welfare for Different Variances of Match Utility](image)

Note: The graphs show the average across 95 simulations of one-to-one matching markets with 25 agents on each side of the market. All agents’ match utilities are log-normally distributed. We vary the tax rate, $\tau$, from 0 to 1 in increments of .02. For each tax rate, we find the manager-optimal stable arrangement and calculate the deadweight loss as a fraction of maximal welfare in the market.

Heterogeneity is also decreased when the match utilities within a side of the market are correlated. Some workers are more productive no matter where they work for and some firms are unpleasant for all workers. We run simulations varying the weight on the fixed-effect component versus the idiosyncratic component of match utilities. That is, we vary the correlation between $\alpha_{wm}$ and $\alpha_{wm}^\prime$ and between $\gamma_{wm}$ and $\gamma_{wm}^\prime$. Adding idiosyncrasies mechanically increases the maximum total match utility in a market because it means that workers are
more likely to have high match utility with at least one firm and that firm is less likely to be other workers’ first choice firm. Moreover, as shown in Figure 1.15, the fraction of welfare lost to taxation at a given tax rate is lower when match utilities are more idiosyncratic. This seems to be largely an extensive margin effect: when tax rates increase such that a firm cannot transfer enough to keep a given worker, there is likely to be another worker with lower productivity and lower disutility of work (less negative match utility) whom the firm can hire.

![Figure 1.15](image)

**Figure 1.15:** Deadweight Loss as a Fraction of Welfare for Different Weights on Manager and Worker Fixed Effects versus Idiosyncratic Components of Match Utility

Note: The graphs show the average across 192 simulations of one-to-one matching markets with 25 agents on each side of the market. All agents’ match utilities are log-normally distributed. No idiosyncratic component refers to markets where $\alpha_w = \beta_w$ and $\gamma_m = \beta_m$. Only Idiosyncratic Component refers to utilities being independently and identically distributed. The Intermediate and Idiosyncratic Component refers to markets that are a mixture of the two. See Appendix A.2 for details. We vary the tax rate, $\tau$, from 0 to 1 in increments of .02. For each tax rate, we find the manager-optimal stable arrangement and calculate the deadweight loss as a fraction of maximal welfare in the market.

The two preceding findings suggest that distortions from taxation will be smaller in markets with more variance (both idiosyncratic and overall) in productivity or tastes. If
human capital leads to more idiosyncratic ability – either across workers or within a worker across jobs – then taxes would cause less welfare loss in skilled labor markets relative to unskilled labor markets.

Lastly, we vary the correlation between how productive a worker is at a job and how onerous the job is for him, i.e. the correlation between $\alpha^w_m$ and $\gamma^m_w$. If this correlation is positive, the jobs where workers are most productive are the ones they like best. If it is negative, then the jobs where workers are most productive tend to be the ones that they dislike the most. Positive correlation tends to increase welfare in the market because when a worker-manager pair get an unusual good productivity draw, they also get a good worker utility draw, so maximum surplus across pairs tends to be higher. Positive correlation also decreases the negative effects of taxation. When workers like the jobs where they are productive, then higher tax rates cannot cause workers to move from jobs where they are more productive to ones that they prefer. Figure 1.16 shows the deadweight loss (as a fraction of welfare) for markets with different levels of taxation. This result suggests that the negative labor market effects of income taxation will be mitigated in markets where workers like jobs where they are more productive. In academia, for example, professors tend to directly enjoy the fruits of their productivity (high-quality research, published papers, successful students). If universities value professors’ productivity, then universities’ preferences will be correlated with professors’ preferences, so taxes should not cause a lot of misallocation of professors to universities.

### 1.4.2 Subsidies

Though we discuss taxation, most of our results either apply directly or translate easily to the case of negative taxes, i.e. subsidies. These also have important real-world applications such as the Earned Income Tax Credit in labor markets or federal education subsidies in the college market.

Increasing a proportional subsidy decreases total match utility (as well as worker match
Figure 1.16: Deadweight Loss as a Fraction of Welfare for Different Correlations of Manager and Worker Utility

Note: The graphs show the average across 192 simulations of one-to-one matching markets with 25 agents on each side of the market. All agents’ match utilities are log-normally distributed. The correlation refers to the correlation between $\alpha^w_m$ and $\gamma^m_w$. Within one side of the market, all match utilities are independently and identically distributed. See Appendix A.2 for details. We vary the tax rate, $\tau$, from 0 to 1 in increments of .02. For each tax rate, we find the manager-optimal stable arrangement and calculate the deadweight loss as a fraction of maximal welfare in the market.

utility). A matching stable under proportional subsidy $\tilde{\tau} < 0$ maximizes the sum of match utilities with weight $(1 - \tilde{\tau}) > 1$ on manager utilities. Increasing a lump sum subsidy increases the number of agents matched and decreases total match utility. A matching stable under lump-sum subsidy $\tilde{f} < 0$ maximizes total match utility for the number of agents that are matched. There will always be more agents matched under a lump sum subsidy than under a lump sum tax. Unfortunately, nothing can be said in general about total match utility under a tax relative to total match utility under a subsidy.
1.4.3 Structural Properties

The Effect of Very Small Taxes

Unlike in non-matching models of taxation, in our setting there is always a non-zero tax that does not generate distortions. To see this in the proportional tax setting, let $\hat{\mu}$ be an efficient matching. Our results show that if $\bar{\mu}$ is stable under $\bar{\tau}$, then

$$\bar{\tau} \geq \frac{\Sigma (\hat{\mu}) - \Sigma (\bar{\mu})}{\sum_{m \in M}(\alpha_{\hat{\mu}(m)} - \alpha_{\bar{\mu}(m)})}.$$  \hspace{1cm} (1.11)

For any inefficient matching $\tilde{\mu}$, there is a strictly positive minimum tax $\tau (\tilde{\mu})$ at which $\tilde{\mu}$ could possibly be stable. Since there are finitely many possible matchings, we can just take the minimum of this threshold across inefficient matchings,

$$\tau^* = \min_{\{\mu: \tau (\mu) < \tau (\hat{\mu})\}} \tau (\mu).$$

For $\tau < \tau^*$ only an efficient matching can be stable.\(^{32}\) The argument for the case of lump sum taxation is similar.

Structure of the Set of Stable Arrangements

Results of Kelso and Crawford (1982) and Hatfield and Milgrom (2005) imply that for any fixed $\tau$, or $f$, if there are multiple stable arrangements, then workers’ and managers’ interests are opposed. If all managers prefer $[\mu; t]$ to $[\hat{\mu}; \hat{t}]$, then all workers prefer $[\hat{\mu}; \hat{t}]$ to $[\mu; t]$. Moreover, there exists a manager-optimal (worker-pessimal) stable arrangement that the managers weakly prefer to all other stable arrangements and a worker-optimal (manager-pessimal) stable arrangement that all workers weakly prefer. In wage markets with proportional taxation, where there is generically a unique stable matching, this opposition of interests carries over to the set of supporting transfer vectors.

\(^{31}\) See Equation (A.30) of the Appendix.

\(^{32}\) One caveat is that if there are multiple efficient matchings (all of which are stable when $\tau = 0$), some of them may not be stable in the limit as $\tau \to 0$ or $f \to 0$. 

38
1.5 Conclusion

We analyze the matching distortion that arises when taxes on transfers affect the matching of workers to managers. In wage markets, matching distortions always decrease as taxes are reduced. In more balanced markets, such as marriage markets or student-college matching, where transfers can flow in either direction, the distortion may be non-monotonic in the amount of taxation or transfer frictions. Non-monotonicities can also occur with piece-wise linear (or curved) taxes because they create changes in the slope of the transfer function, similar to the change in slope that occurs at 0 in non-wage markets.

Matching distortions affect the allocative margin and can arise even without intensive or extensive margin effects. However, our framework allows for extensive margin effects and partially incorporates intensive margin effects, to the extent that changes in work hours are often achieved by changing jobs. An extension of our work would examine a fuller interaction of allocative and intensive margin effects in a model that allows for labor supply decisions within a job. It would also be valuable to analyze how the magnitude of the matching distortion depends on the variance and heterogeneity of agents’ preferences. Such work might inform the estimation of the losses due to matching distortions in real-world labor markets.

It is also natural to ask about revenue: How much revenue do different tax structures generate in matching markets? For a given revenue requirement, does a proportional tax generate more or less distortion than a lump sum tax?

The first challenge in answering questions about revenue is that for any stable match, there will generally be a lattice of possible supporting transfer vectors. For proportional taxation, revenue depends on the choice of supporting transfer vector. The easiest transfer vectors to think about are the maximal (worker-optimal) supporting transfer vectors and the minimal (manager-optimal) supporting transfer vectors, which, in wage markets, correspond to maximal and minimal revenue (given the match and tax rate).

Even after focusing on the extremal supporting transfers, addressing revenue questions requires adding structure on agents’ match utilities. Unfortunately, it is not clear whether
there is a natural structure to impose. Most papers in the matching literature that (unlike our work) do not allow for fully general match utilities assume that the match surplus is a function of one-dimensional agent types. This usually implies that agents agree on the ordinal ranking of agents on the other side of the market. This shifts the distortion to the extensive margin – at any tax level, the most desirable agents on each side will be matched. Moreover, in our framework, just assuming a structure for match surplus is insufficient because the pre-transfer split of match utility, rather than just the total surplus, affects match outcomes in the presence of taxation.

Galichon and Salanié (2014) put enough structure on match utilities to get equations for matchings and surplus without assuming agents agree on the ordinal rankings of match partners. Jaffe is working with Galichon to adapt extensions of the Galichon and Salanié (2014) method to answer questions about deadweight loss and revenue in the presence of taxation. The complexity of this exercise arises not only because of imperfect transfers, but also from the resulting need to separately identify worker and manager match utilities.

Jaffe is also running lab experiments to understand the effects of transfer frictions on matching in a controlled setting. Explicitly dictating match utilities and manipulating tax rates, will allow her to see whether the availability of transfers (and taxes on those transfers) impacts the probability that a market reaches a stable match, and to analyze the transfers agents select.
Chapter 2: The First-Order Approach to Merger Analysis$^1$

Much recent theoretical and applied antitrust work has focused on using information local to the pre-merger equilibrium to predict the directional price impacts of mergers. This “first-order” approach adopts both the simplicity and transparency of approaches based on market definition and the firm grounding in formal economics of the market simulation approach. Section 2.1 gives a more extensive background, but the logic of this approach is intuitive: when firms 1 and 2 merge, firm 1 (and similarly, firm 2) has an incentive to raise its prices because of a new opportunity cost of selling its products – firm 1 now internalizes the profit lost by firm 2 when firm 1 lowers its price. For each extra unit firm 1 sells, the profit lost by firm 2 is the fraction of sales gained by firm 1 that are cannibalized from firm 2 (typically called the *diversion ratio*), multiplied by the profit-value of those sales (firm 2’s *mark-up*).

We extend this first-order approach in three ways. First, we develop a model that allows for non-Nash-in-prices oligopoly behavior. Second, we show how information on local pass-through rates can be used to convert directional indicators of “pricing pressure” into quantitative approximations to the price changes caused by a merger. Finally, we describe how these quantitative estimates of price changes can translate into approximations of impacts on consumer and social welfare.

$^1$Co-authored with Glen Weyl
\[ \Delta CS \approx - \mathbf{g}^T \cdot \mathbf{\rho}^T \cdot Q \]  

where the approximation error is proportional to the square of the merger’s effects on prices and to the curvature of the equilibrium conditions (third derivatives of supply and demand). The first term, \( \mathbf{g} \), is the Generalized Pricing Pressure (GePP), the change in pricing incentives, allowing for non-Bertrand conduct and non-constant-marginal-cost systems. Developed and explained more fully in Section 2.2, GePP is a vector that has entries of zero for all non-merging firms and – in the case when single-product firms 1 and 2 merge – a first entry of the form

\[ g = \frac{\text{Conjectured diversion ratio}}{\hat{D}_{12}} \cdot \frac{\text{Mark-up}}{(P_2 - mc_2)} - Q_1 \]

and an analogous second entry. The first term in Equation (2.2) extends the logic discussed above by replacing the Bertrand diversion ratio, \( D_{12} \), with the anticipated diversion ratio \( \hat{D}_{12} \). This ratio is the fraction of the units lost by good 1 that are gained by good 2 when firm 1 raises its price, holding fixed the price of good 2 but allowing all other prices to adjust as the merged firm expects.\(^2\) The price of good 2 is held fixed because, as a result of the merger, it is one of the quantities over which the merged firm optimizes. The second term in (2.2) is the quantity of good 1 multiplied by the (merger-induced) change in the inverse of the derivative of demand: now that the firms are merged, firm 1 no longer anticipates a reaction from firm 2 to a change in its price. If firm 2 were accommodating pre-merger (raising its price in response to a price increase by firm 1) then firm 1’s elasticity of demand will be

\[ \frac{1}{dP_1^2} - \frac{1}{dQ_1^1} \]

\[ \text{End of (accommodating) reactions} \]
higher post-merger.\textsuperscript{3}

The second term in Equation (2.1), $\rho$, is the \textit{merger pass-through} matrix, the rate at which the changes in opportunity cost created by the merger, the GePP, are passed through to changes in prices. As we show in Section 2.3, this matrix, which is a function of local second-order properties of the demand and cost systems, converts GePP into a quantitative approximation of the price effects of the merger. In Section 2.4 we discuss further the role of pass-through; we argue that in many relevant cases merger pass-through is close to both pre-merger and post-merger pass-through, reconciling divergent strains in recent literature on the relevant pass-through rate. In certain cases, exact merger pass-through may be identified from pre-merger pass-through. For other cases, the empirical and numerical results of Cheung (2011) and Miller et al. (2012) provide evidence for the validity of approximating merger pass-through with pre-merger pass-through rates.

The third term in Equation (2.1), $Q$, is the vector of quantities, which are the correct weights for aggregating the price changes into the change in consumer surplus, as we discuss in Section 2.5. A similar approach may be used to estimate social surplus impacts. This aggregation, which converts price changes into dollar measures of surplus, facilitates comparison with merger effects on consumer welfare not directly mediated by prices, such as changes in network size or product quality.

Section 2.6 discusses extensions and practical implications of our framework, including ways to simplify the formula given time and resource constraints. Our conclusion in Section 2.7 discusses directions for future research. A companion policy piece (Jaffe and Weyl, 2011) proposes a few potential reforms to the merger guidelines based on our analysis. Proofs not in the text are in the Appendix B.

\textsuperscript{3}For further discussion of the effects of accommodation see Section 2.2.2.2.4.
2.1 Background on the First-Order Approach

During the 1990s, merger simulation became widespread as a method for predicting the unilateral price effect of mergers. However, Shapiro (1996) and Crooke et al. (1999) argued that the effects of mergers predicted by simulations could differ by an order of magnitude or more based on properties of the curvature of demand not typically measured empirically. To address this concern, Werden (1996) pioneered the first-order approach by arguing that the “compensating marginal cost reductions” necessary to offset the anticompetitive effects of a merger could be calculated from local properties of the demand system. Such cost efficiencies would have to offset the change in first-order conditions created by the new opportunity cost of a sale due to the diversion of sales of a merger partner’s product. Shapiro (1996) observed that, regardless of functional form, merger effects appeared to be increasing in this “value of diverted sales,” which has come to be known as “Upward Pricing Pressure” (UPP).

Building on this work, antitrust officials in the United Kingdom, led by Peter Davis and Chris Walters, began to use UPP to evaluate mergers (Walters, 2007). Froeb et al. (2005) noted that functional forms which imply higher pass-through rates of cost efficiencies generated by the merger tend to also generate large unilateral merger effects on prices. They proposed an approach, based on Newton’s method, for conducting merger simulations in a computationally simpler manner whose first iteration only required information local to pre-merger prices. Building on the practical work in the UK and the theoretical analysis of Froeb et al. (2005), Farrell and Shapiro (2010a,b) translated these ideas into intuitive and widely accessible economic terms: they argued that the sign of UPP minus efficiency gains would indicate the direction of merger effects and put forward the measurement of UPP as a practical policy proposal for the evaluation of mergers.

Under the leadership of Farrell and Shapiro, the US incorporated UPP into the 2010 Merger Guidelines (United States Department of Justice and Federal Trade Commission, 2010). The UK followed close behind with an even more explicit incorporation of UPP (Competition Commission and Office of Fair Trading, 2010); the European Union is also
considering revising its merger guidelines.

Nevertheless, some objections have been raised against the use of UPP in analyzing mergers:


2. Schmalensee (2009) and Hausman et al. (2011) are skeptical of its assumption of default efficiencies and argue that providing only a directional indication of price effects is insufficient.


While these issues also arise to varying degrees in alternative approaches to merger analysis, they are still worth addressing. In this context, we make three contributions that, to the best of our knowledge, have not appeared in previous literature. First, we analyze pricing pressure in a model that is not limited to Nash-in-prices competition or constant marginal cost. Second, we formalize the folk wisdom that pass-through rates can be used to convert pricing pressure into an approximation of quantitative price impacts, thereby forgoing the need for default cost efficiencies. Finally, we present formulae for approximate changes in consumer welfare that allow for the aggregation of multiple price changes for multi-product firms.

Competition agencies’ increased openness to a range of simple tools with firm economic grounding (Shapiro, 2010) has sharpened the focus on the appropriateness of the first-order approach for policy and its soundness as a theoretical construct; the agencies’ increased interest in broadening the scope of analysis raises the relevance of extending UPP to non-Bertrand settings (see the Office of Fair Trading-comissioned report on the role of conjectural variations in merger policy (Majumdar et al., 2011)).
2.2 Generalized Pricing Pressure

In this section we adapt the Telser (1972) single-strategic-variable-per-product oligopoly model by formulating it in terms of prices (rather than quantities) and allowing for multi-product firms. As Telser shows, by including non-price behavior in anticipated reactions by other firms, this model encompasses most standard static oligopoly models – including Nash-in-Prices (Bertrand), Nash-in-Quantities (Cournot) and most supply function equilibria. In this framework, we derive a formula for the changes in pricing incentives firms face post-merger – the Generalized Pricing Pressure (GePP). In Subsection 2.2.3, we show how to incorporate efficiencies from the merger. In Subsection 2.2.4, we give two specific examples to illustrate how the formula works in the specific cases of Bertrand and Cournot; we also present an example that explores how the degree of accommodating reaction affects the size of GePP.

2.2.1 The model

Consider a market with \( N \) firms denoted \( i = (1, \ldots, n) \). Firm \( i \) produces \( m_i \) goods, and chooses a vector of prices \( P_i = (P_{i1}, P_{i2}, \ldots, P_{im_i}) \) from \( \mathbb{R}^{m_i} \). Following Telser (1972), we permit each firm to conjecture reactions by other firms: changes in those firms’ prices in response to changes in its own prices. This formulation is useful for two reasons:

First, it allows us to nest static oligopoly models where firms have a strategic variable other than price, such as quantity or a supply function shifter. If the strategic variable is not price, this is incorporated into a firm’s conjectures about other firms’ reactions to its price change. For example, as we illustrate in Subsection 2.2.4 below, Cournot competition is represented by a firm conjecturing that when it raises its price, other firms will raise their prices so as to hold fixed their quantities. This formulation encompasses many strategic contexts, but does restrict each firm to have a single strategic variable per product (as in

\[ \text{footnote}{In an earlier version of this work we considered the analysis for the case where any strategy (such as quantity) is chosen. This more general analysis is available in an online appendix.} \]
Werden and Froeb (2008)).

Second, these conjectured reactions allow for the possibility of non-static Nash behavior in the spirit of the conjectural variations of Bowley (1924). Despite their absence from the mainstream of industrial organization empirics and theory since the 1980s, recently there has been a resurgence in theoretical (Dockner, 1992; Cabral, 1995), empirical (Nevo, 1998; Ciliberto and Williams, 2011) and policy (Majumdar et al., 2011) interest in such non-Nash frameworks as a useful reduced form for the complexities of dynamic models of competition. Conjectured reactions can also result from tacit collusion in the industry pre-merger, which can potentially alter the effects of a merger.

These conjectures are modeled by letting a firm believe that when it changes its prices, $P_i$, its competitors will change their prices, $P_{-i}$, by $\frac{\partial P_{-i}}{\partial P_i}$. Therefore, the total effect of a change in one’s own price on a vector of interest is the sum of the direct (partial) effect and the indirect effect working through the effect on others’ prices:

$$\frac{dA}{dP_i} \equiv \frac{\partial A}{\partial P_i} + \frac{\partial A}{\partial P_{-i}} \frac{\partial P_{-i}}{\partial P_i}.$$  

In the case of a Bertrand equilibrium, we have $\frac{dA}{dP_i} = \frac{\partial A}{\partial P_i}$ since $\frac{\partial P_{-i}}{\partial P_i} = 0$.

**Pre-merger**

Firm $i$’s profit $\pi_i$ depends on both firm its own price vector and its competitors’ prices:

$$\pi_i = P_i^T Q_i(P) - C_i(Q_i(P)).$$

---

5We thus rule out changes in the non-price determining characteristics of products as considered in the literature on product repositioning (Mazzeo, 2002; Gandhi et al., 2008). See Section 2.5 for a discussion of how merger effects on non-price characteristics can be incorporated into our framework. See Telser (1972) for more details of the range of models that are special cases of this framework.

6Throughout we use the notation $\frac{\partial}{\partial}$ to refer to the Jacobian $\frac{\partial A}{\partial B} \equiv \begin{pmatrix} \frac{\partial A_1}{\partial B_1} & \cdots & \frac{\partial A_1}{\partial B_m} \\ \vdots & \ddots & \vdots \\ \frac{\partial A_n}{\partial B_1} & \cdots & \frac{\partial A_n}{\partial B_m} \end{pmatrix}$. 

---

47
where $C$ and $Q$ are the cost and demand functions. For brevity we write $Q_i$ for $Q_i(P)$ and $\mathbf{mc}_i$ for the vector of marginal costs. The firm’s vector of first-order conditions pre-merger can be written as:

$$0 = -\left(\frac{dQ_i}{dP_i}\right)^{-1} \mathbf{Q}_i - (P_i - \mathbf{mc}_i) \equiv f_i(P).$$

This formula is a natural extension of the standard, single-product oligopoly first-order condition: the mark-up on each product is equated to the matrix analog of the partial inverse hazard rate or Cournot distortion. Matrix limited to the firm’s products multiplied by that firm’s quantities, $\left(\frac{\partial Q_i}{\partial P_i}\right)^{-1} \mathbf{Q}_i$.

### 2.2.2 Incentives created by a merger

In studying the impact of a merger on firms’ incentives, it is useful to define a couple of terms.

**Definition.** If firms $i$ and $j$ merge, the post-merger diversion ratio matrix is

$$\tilde{D}_{ij} \equiv -\left(\frac{d^M Q_i}{dP_i}\right)^{\top} \left(\frac{d^M Q_j}{dP_i}\right)^{\top},$$

where $\frac{d^M Q_k}{dP_i} = \frac{\partial Q_k}{\partial P_i} + \frac{\partial Q_k}{\partial P_{-ij}} \frac{\partial P_{-ij}}{\partial P_i}$, which holds fixed the merging partner’s prices.\footnote{If more than 2 firms merge, then $Q_j$ is replaced by $Q_{j,k,...}$.}

The relevant matrix of diversion ratios is the matrix ratio of the quantity (anticipated to be) gained by the former rival’s products to that (anticipated to be) lost by one’s own products as a result of an increase in own price, holding fixed the price of the merger partner and allowing all other prices to adjust as they are expected to in equilibrium.

**Definition.** Let the pre-merger and post-merger first-order conditions, normalized to be quasi-linear in marginal cost be denoted $f(P)$ and $h(P)$, respectively; the pre-merger equilibrium is
defined by \( f(P^0) = 0 \); the post-merger equilibrium is defined by \( h(P^M) = 0 \). Then we define the Generalized Pricing Pressure (GePP) created by the merger to be

\[
g(P^0) \equiv h(P^0) - f(P^0).
\]

Thus GePP is the change in the first-order condition at the pre-merger prices. It captures only the unilateral effects of a merger, *holding fixed the strategies firms use (price v. quantity, etc.) and conjectures about other firms’ reactions*. The value of GePP is given in the following proposition.

**Proposition 8.** The GePP on firm \( i \) generated by a merger between firms \( i \) and \( j \) is \( g_i(P^0) \) where \( P^0 \) is the pre-merger equilibrium price vector and

\[
g_i(P) \equiv \tilde{D}_{ij} (P_j - mc_j) - \Delta \left( \left( \frac{dQ_i}{dP_i} \right)^{-1} \right)^T Q_i. \tag{2.3}
\]

Here \( \Delta(\cdot) \) denotes the change from pre- to post-merger value of its argument; the change is due to the merger partner’s strategy no longer reacting.\(^8\)

The first term of Equation (2.3) is the change in firm \( j \)'s profits when firm \( i \) increases its price enough to lose one marginal sale: the fraction of a unit gained by firm \( j \) for each unit lost by firm \( i \) times the value of a unit. The second term is the change in firm \( i \)'s own marginal profit due to the end of accommodating reactions: once the firms have merged, firm \( i \) no longer anticipates an accommodating reaction from its merger partner, so its demand becomes more elastic.

### 2.2.3 Marginal cost efficiencies

The GePP formula derived above assumes no cost efficiencies of the merger and as such can be seen as the baseline case. However, if estimates of expected efficiencies are available, they

\(^8\)Note that in the single-product firm case this is exactly Equation (2.2) from the introduction. If more than 2 firms merge, than the quantity and price vectors of firm \( j \) are replaced by vectors containing the quantities and prices of all merger partners other than \( i \).
can easily be incorporated. If post-merger firm $i$’s marginal costs are expected to be $\tilde{mc}_i$, then the GePP for firm $i$ after a merger of firms $i$ and $j$ is

$$\tilde{g}_i(P) = \tilde{D}_{ij}(P_j - \tilde{mc}_j) - \Delta \left( \left( \frac{dQ_i}{dP_i} \right)^{-1} \right)^T Q_i - (mc_i - \tilde{mc}_i).$$

This adjusted GePP can be used in the calculation of price changes and welfare effects or to calculate “compensating marginal cost reductions” (Werden, 1996). For the marginal cost reductions to counterbalance the other incentive effects and lead to no price change, it must be that

$$\begin{pmatrix} \tilde{g}_i(P) \\ \tilde{g}_j(P) \end{pmatrix} = 0,$$

which yields compensating cost reductions of

$$\begin{pmatrix} e^*_i \\ e^*_j \end{pmatrix} \equiv \begin{pmatrix} mc_i \\ mc_j \end{pmatrix} - \begin{pmatrix} \tilde{mc}_i \\ \tilde{mc}_j \end{pmatrix} = \begin{pmatrix} I & -\tilde{D}_{ij} \\ -\tilde{D}_{ji} & I \end{pmatrix}^{-1} \begin{pmatrix} g_i(P) \\ g_j(P) \end{pmatrix},$$

which is equivalent to Werden’s formula in the case of single-product Bertrand. Alternatively, if one wishes to apply the more permissive standard of Farrell and Shapiro (2010a), the off-diagonal terms are ignored and the GePP itself is contrasted to efficiencies.

### 2.2.4 Specific Contexts and Examples

This subsection illustrates the model under a few common equilibrium concepts. The formulae for Bertrand and Cournot are below, followed by a discussion of the effect of accommodating reactions on GePP. For the continuation of the main theory, see Section 2.3.
Bertrand

In the case of Bertrand, the expected accommodating reactions are zero, so GePP equals the (multi-product) UPP formula:

$$g_i(P) = -\left(\frac{\partial Q_i}{\partial P_i}\right)^{-1} \left(\frac{\partial Q_j}{\partial P_i}\right)^T (P_j - mc_j).$$

To help clarify, we now consider the explicit computation of GePP with two symmetric, multiproduct firms with constant marginal cost, playing Bertrand in a market with linear demand who then merge to (residual) monopoly.  

**Example 3.** Suppose that two symmetric, multiproduct firms 1 and 2 with symmetric products, all of which are substitutes for one another, and constant marginal cost vector $c$ face the linear demand system

$$\begin{pmatrix} Q_1 \\ Q_2 \end{pmatrix} = \begin{pmatrix} A \\ A \end{pmatrix} - \begin{pmatrix} B_o & B_x \\ B_x & B_o \end{pmatrix} \begin{pmatrix} P_1 \\ P_2 \end{pmatrix}.$$  

Note that $B_0$ is positive definite, $B_x$ has all entries negative, and symmetry implies $B_0 = B_0^T$ and $B_x = B_x^T$. Profits prior to the merger for either firm $i$ are $(P_i - c_i)^T Q_i = (P_i - c_i)^T (A - B_o P_i - B_x P_{-i})$ and thus, by the matrix product rule, the first-order condition is

$$0 = A - B_o P_i - B_x P_{-i} - B_o (P_i - c) \iff 2B_o P_i + B_x P_{-i} = A + B_o c.$$

If we solve for a symmetric equilibrium, $P_i = P_{-i} = P^0$, the equation becomes

$$(2B_o + B_x) P^0 = A + B_o c_i \iff P^0 = (2B_o + B_x)^{-1} (A + B_o c).$$

Thus the pre-merger mark-up is

$$\left(2B_o + B_x\right)^{-1} (A + B_o c) - c = \left(2B_o + B_x\right)^{-1} [A - (B_o + B_x) c].$$

\[9\text{Note that we have criticized the plausibility of this demand system in other work (Jaffe and Weyl, 2010). We rely on it here not for realism but rather because it nicely illustrates how one may compute our formula in a specific example.}\]
From the structure of demand, the diversion ratio is \(-B_o^{-1}B_x\). Thus the GePP on both firms is
\[-B_o^{-1}B_x \left(2B_o + B_x\right)^{-1} \left[A - (B_o + B_x) \mathbf{c}\right].\]

**Cournot**

The general formula for Cournot appears in the online appendix. Here we focus on a simple example of differentiated Cournot competition which illustrates how the Cournot model fits as a special case of our analysis.

**Example 4.** Consider two symmetric, single-product firms at a symmetric, Cournot equilibrium. The Slutsky matrix is both symmetric about both diagonals and thus without loss of generality may be specified as
\[
\frac{\partial Q}{\partial P} = \begin{pmatrix}
-(s + \sigma) & \sigma \\
\sigma & -(s + \sigma)
\end{pmatrix},
\]
where \(s, \sigma > 0\) as the goods are substitutes and demand is downward sloping when both prices rise by the same amount. To calculate \(\frac{dP_{-i}}{dP_i}\) we use the chain rule and the fact that under Cournot
\[
0 = \frac{dQ_{-i}}{dP_i} = \frac{\partial Q_{-i}}{\partial P_i} + \frac{\partial Q_{-i}}{\partial P_{-i}} \frac{dP_{-i}}{dP_i} = \sigma - (s + \sigma) \frac{dP_{-i}}{dP_i} \iff \frac{dP_{-i}}{dP_i} = \frac{\sigma}{s + \sigma}.
\]
This gives us \(\frac{dQ_i}{dP_i} + \frac{\partial Q_i}{\partial P_{-i}} \frac{dP_{-i}}{dP_i} = -(s + \sigma) + \sigma \frac{\sigma}{s+\sigma} = \frac{-s(s+2\sigma)}{s+\sigma}\). Prior to the merger, firms price according to the logic of Subsection 2.2.2.2.1, so
\[
P_i - mc_i = \frac{-Q_i}{\frac{\partial Q_i}{\partial P_i} + \frac{\partial Q_i}{\partial P_{-i}} \frac{dP_{-i}}{P_i}} = \frac{Q_i(s + \sigma)}{(s+2\sigma)s}.
\]
After the merger, there are no reactions, which means \(\frac{dM_{Q_i}}{P_i} = \frac{dQ_i}{dP_i} = -(s + \sigma)\), so the end of accommodating reactions term is
\[
Q \left(\frac{-1}{s+\sigma} - \frac{-(s + \sigma)}{(s+2\sigma)s}\right) = \frac{Q(s + \sigma)^2 - s(s + 2\sigma)}{s(s + 2\sigma)(s + \sigma)} = \frac{Q\sigma^2}{s(s+2\sigma)(s+\sigma)}.
\]
The appropriate diversion ratio is the one holding fixed the price of the merger partner; since this is a merger to monopoly, that is just the Bertrand diversion ratio $\frac{\sigma}{s+\sigma}$.

If there are marginal cost efficiencies with a change in marginal costs of $\Delta(mc) < 0$, then the GePP on each product is

$$g_i(p) = D(P - mc) - D\Delta(mc) - \Delta \left( \frac{1}{\frac{dQ_i}{dP_i}} \right) Q_i + \Delta(mc)$$

$$= \frac{\sigma}{s + \sigma} \frac{Q(s + \sigma)}{(s + 2\sigma)s} - \frac{Q\sigma^2}{s(s + 2\sigma)(s + \sigma)} + (1 - D)\Delta(mc)$$

$$= \frac{Q\sigma}{(s + 2\sigma)(s + \sigma)} + \frac{s}{s + \sigma} \Delta(mc).$$

Note that, without marginal cost efficiencies, as the products become undifferentiated ($\sigma \to \infty$) this formula converges to 0. This seems to indicate that a merger-to-monopoly in undifferentiated Cournot causes no increase in prices, which is clearly absurd. The problem, as we show in Appendix B.2, is that as the products become undifferentiated, the pass-through by which the GePP must be multiplied to obtain the price change explodes. This case shows both how our formula works out in a non-Bertrand, but canonical model and illustrates why considering pass-through, and not just pricing pressure, is often crucial.

**Accommodation**

A natural concern is that, especially in differentiated product industries, it may be difficult to determine empirically or even grasp intuitively what model of conduct is appropriate (Nevo, 1998). While for many questions this is a serious worry, it may not be a severe problem for merger analysis since anticipated accommodating reactions (arising either from the strategic variable not being price or from non-Nash behavior) have two off-setting effects. If firms 1 and 2 merge, accommodating reactions from the non-merging firms increase the GePP by increasing the (conjectured) diversion ratio: if other firms increase their prices when firm 1 does, that would both reduce the quantity of sales lost by firm 1 when it raises its price and increase the quantity gained by firm 2 (whose price is held fixed). On the other hand, the
greater the pre-merger accommodating reactions between the firms 1 and 2, the greater is the increase in firm 1’s elasticity of demand due to firm 2 no longer accommodating post-merger; therefore accommodation increases the (subtracted) term for the end of accommodating reactions, thereby decreasing the GePP. Which of these effects dominates will depend on how large the anticipated accommodation between the merging firms and other firms in the industry (first effect) is relative to the accommodation between the merging partners (second effect).

Because of these offsetting effects, the size of GePP may not differ as much across alternative conduct assumptions as it might at first appear. The following example demonstrates this phenomenon and provides another illustration of how GePP can be computed in specific models.

**Example 5.** Consider a pre-merger symmetric industry with \( n \) single-product firms; pre-merger they are in a symmetric equilibrium, earning mark-up \( m \), each selling quantity \( q \), with an aggregate (Bertrand) diversion ratio \( D \) to the \( n - 1 \) other firms in the industry. On the margin, each firm anticipates an increase in price of \( \lambda \) by all other firms in response to a one unit increase in their own price.

Prior to the merger, the first-order condition for a single firm requires that

\[
m = -\frac{q}{\frac{\partial Q_i}{\partial P_i}}.
\]

Symmetry implies that \( \frac{\partial Q_i}{\partial P_j} = \frac{\partial Q_j}{\partial P_i} = -\frac{\partial Q_i}{\partial P_i} \cdot \frac{D}{n-1} \). This allows us to calculate

\[
\frac{dQ_i}{dP_i} = \frac{\partial Q_i}{\partial P_i} + \sum_{j \neq i} \frac{\partial Q_i}{\partial P_j} \frac{\partial P_j}{\partial P_i} = \frac{\partial Q_i}{\partial P_i} + \frac{\partial Q_i}{\partial P_j} (n-1) \lambda = \frac{\partial Q_i}{\partial P_i} (1 - D \lambda),
\]

which gives

\[
\frac{\partial Q_i}{\partial P_i} = -\frac{q}{m (1 - D \lambda)}.
\]

Post-merger the price of the merger partner is held fixed, rather than increasing by \( \lambda \) in response to an increase in the firm’s price. To see how this difference affects the anticipated
accommodation, note that the total pre-merger accommodation firm \( i \) anticipates from each
other firm is the direct effect while holding fixed its merger partner’s price plus the indirect
effect via the impact on its merger partner. Therefore, post-merger \( \tilde{\lambda} \), the symmetric increase
in the \( n-2 \) remaining firms’ prices in response to an increase in one of the partners’ prices
must satisfy
\[
\lambda_{\text{pre-merger response}} \quad \text{holding fixed partner} \quad + \quad \lambda_{\text{equilibrium partner response}} \quad \cdot \quad \tilde{\lambda}_{\text{partner’s equilibrium effect holding } i \text{ fixed}},
\]
thus \( \tilde{\lambda} = \frac{\lambda}{1 + \lambda} \).

These quantities allow us to calculate the relevant post-merger derivatives. Call the
merging firms, firm 1 and firm 2 and consider \( \frac{d^M Q_1}{dP_1} \). This is composed of the direct effect
and the indirect effect from the change in the \( n-2 \) non-merging firm prices:
\[
\frac{d^M Q_1}{dP_1} = \frac{\partial Q_i}{\partial P_i} \cdot \frac{\partial Q_i}{\partial P_j} (n-2) \tilde{\lambda} = \frac{\partial Q_i}{\partial P_i} \left( 1 - \tilde{\lambda} \frac{n-2}{n-1} D \right).
\]
For the merger partner, firm 2, the sales gained are the direct diversion plus the indirect
diversion from the increase in the \( n-2 \) non-merging firms’ prices:
\[
\frac{d^M Q_2}{dP_1} = -\frac{\partial Q_i}{\partial P_i} \cdot \frac{D}{n-1} \left( 1 + \tilde{\lambda} [n-2] \right).
\]
Thus,
\[
\tilde{D}_{12} \cdot m = \frac{D}{n-1} \left( 1 + \tilde{\lambda} [n-2] \right) m = \frac{D \left( 1 + \tilde{\lambda} (n-2) \right)}{(n-1) \left( 1 - \tilde{\lambda} D \right) + \tilde{\lambda} D},
\]
and the change in accommodating reaction term is
\[
Q_1 \left( \frac{1}{\frac{d^M Q_1}{dP_1}} - \frac{1}{\frac{d^M Q_1}{dP_1}} \right) = \frac{q}{m} \left( 1 - \frac{1 - D\lambda}{1 - \tilde{\lambda} \frac{n-2}{n-1} D} \right) = Dm \frac{\left( 1 - \tilde{\lambda} \right) (n-1) + \tilde{\lambda}}{(n-1) \left( 1 - \tilde{\lambda} D \right) + \tilde{\lambda} D},
\]
Subtracting these two terms and using \( \lambda - \tilde{\lambda} = \frac{\tilde{\lambda}}{1 - \lambda} - \tilde{\lambda} = \frac{\tilde{\lambda}^2}{1 - \lambda} \) yields the GePP,
\[
D \cdot m \frac{1 + \tilde{\lambda} (n-3) - D(n-1) \frac{\tilde{\lambda}^2}{1 - \lambda}}{(1 - D\tilde{\lambda}) (n-1) + D\tilde{\lambda}}.
\]
If we focus on the case where $\lambda$ (and thus $\tilde{\lambda}$) is small so that we can ignore second-order terms, then (2.4) simplifies to

$$D \cdot m \frac{1 + \tilde{\lambda}(n-3)}{(1 - D\tilde{\lambda})(n-1) + D\tilde{\lambda}}.$$ \hspace{1cm} (2.4b)

Note that $\tilde{\lambda}$ is strictly increasing in $\lambda$. When $n = 2$, we are considering a merger to monopoly, Equation (2.4b) is proportional to $1 - \tilde{\lambda}$, so it is decreasing in $\lambda$. If accommodation by the merger partner is the only issue, GePP declines with the degree of accommodation as Farrell and Shapiro (2010a) conjecture. However, when $n = 3$ equation (2.4b) is proportional to $\frac{1}{2 - D\tilde{\lambda}}$, which is increasing in $\lambda$. This effect gets stronger as $n \to \infty$; in the limit the expression is proportional to $\frac{\tilde{\lambda}}{1 - D\tilde{\lambda}}$, which increases even more quickly in $\lambda$. Thus, in this basic example, “somewhere between” a merger to a monopoly and a merger by two firms within a triopoly, the effect of accommodation on GePP switches from negative to positive. The exact formula indicates that there is a larger space of parameters for which GePP is decreasing in $\lambda$; in highly collusive industries it may be that even in the case of 3 or 4 firms, more accommodation will lower GePP.

The strength of accommodation between each merging firm and the non-merging firms relative to the accommodation between merging firms matters for the effect of accommodation on GePP (See Appendix B.3 for an example in a non-symmetric industry.) Both of these analyses suggest that while accommodation and the type of oligopoly competition clearly impacts the level of prices, its effect on changes in prices caused by mergers may be smaller and certainly subtler.

**2.3 Price Changes**

GePP measures how much firm incentives shift when firms merge. However, policymakers are typically interested in such shifts in incentives only insofar as they predict changes in prices. We extend the work of Chetty (2009b) to show how a comparative static approach
without a fully-estimated structural model can be used to analyze structural changes such as mergers. If the change in incentives is small, the effect of a merger can be approximated the same way the effect of a tax would be, despite the fact that unlike a tax we cannot imagine a merger “going to zero” to make our formula exact.

Our approach is to apply the appropriate envelope theorem, viewing the change in incentives created by the merger - the pricing pressure, \( g \), - as a vector of local changes in the equilibrium conditions; we then apply Taylor’s Theorem for inverse functions to approximate the post-merger conditions around the pre-merger equilibrium. This allows us to derive an approximation of the effect of the merger based on local properties of demand and to get a bound on the error of the approximation based on the curvature of the first-order conditions and the size of the incentive change.

![Figure 2.1: The effect of pricing pressure for low and high pass-through.](image)

Note: The graph shows profits as a function of price for two different firms. They are chosen such that they have the same pre-merger optimal price and the same generalized pricing pressure, \( g \). The dotted (solid) lines shows pre- and post-merger profits for a firm with low (high) pass-through. The high pass-through firm has a much higher post-merger optimal price. (It’s off the graph to the right.)
For a graphical intuition of why the curvature of profits is important for price changes and GePP is an insufficient indicator, see Figure 2.1. It shows profits as a function of price for two demand systems, both pre-merger (thick lines) and post-merger (thin, blue lines). The profits are based on a constant pass-through demand system; the dotted lines are low pass-through (.1) and the solid lines are high pass-through (.4). Costs are such that pre-merger both firms’ profits are maximized at the same optimal price ($P^0 \approx 1.406$). They also have the same GePP, as evidenced by the tangency of the two post-merger (thin, blue) profit lines at $P^0$. However the curvature of the profit functions is very different and the post-merger profit-maximizing prices are very different. With low pass-through (dotted line), the post-merger price is approximately 1.445; with high pass-through (solid line), the post-merger price is off the graph to the right. This illustrates that demand systems with the same pre-merger prices and the same pricing pressures can have very different post-merger prices when they have different pass-through rates.

Theorem 3 provides our main result, which formalizes this intuition.

**Theorem 3.** Let $P^0$ be the pre-merger equilibrium price vector. If $\mathbf{f}$ is the vector of pre-merger first-order conditions and $\mathbf{g}$ is the GePP vector (so $h(P) = \mathbf{f}(P) + \mathbf{g}(P)$ is the the post-merger first-order condition) and $(\mathbf{f} + \mathbf{g})$ is invertible, then, to a first-order approximation, the price change induced by the merger is

$$\Delta P = - \left( \frac{\partial \mathbf{f}}{\partial P}(P^0) + \frac{\partial \mathbf{g}}{\partial P}(P^0) \right)^{-1} \cdot \mathbf{g}(P^0).$$

**Proof.** Let $h(P) = \mathbf{f}(P) + \mathbf{g}(P)$. Since $f(P^0) = 0$, we have $h(P^0) = g(P^0) \equiv r$. We want to find $P^M$ (the post-merger price), such that $h(P^M) = 0$. If $h$ is invertible, then

$$P^M - P^0 = h^{-1}(0) - h^{-1}(r) = \left( \frac{\partial h^{-1}}{\partial h}(r) \right)(0 - r) + O(\|r\|^2)$$

$$\approx - \left( \frac{\partial \mathbf{f}}{\partial P}(P^0) + \frac{\partial \mathbf{g}}{\partial P}(P^0) \right)^{-1} \cdot \mathbf{g}(P^0),$$

which completes the proof. \qed
As we show in Appendix B.4, the $i$-th entry of the error vector in Equation (2.5) takes the form

$$E_i = -\frac{1}{2} \sum_j \left[ \left( \frac{\partial h}{\partial P} \right)^{-1} \right]_{ij} \mathbf{g}^T (P^0) \left( \frac{\partial h^T}{\partial P} \right)^{-1} (D^2_{P}h_j) \left( \frac{\partial h}{\partial P} \right)^{-1} \mathbf{g} (P^0),$$

where $[A]_{ij}$ indicates the $ij$ element of matrix $A$, $D^2_{P}h_j$ indicates the Hessian of $h_j$, and the derivatives and Hessian are evaluated at some $P \in [P^0, P^M]$. This error is small whenever $\mathbf{g}$ is small and the first-order conditions are not highly curved in the relevant range.  

Willig (1976) famously argued that theoretical approximations of this sort are useful to the extent that the associated errors are small relative to other sources of error such as statistical sampling or mis-specification error involved in accounting for the factors the approximation ignores. Two recent studies have investigated the relative size of the error of our approximation compared to these other sources empirically and numerically.

Cheung (2011) does an empirical analysis of the US Airways-America West merger and found the approximation error of our formula to be small compared to the statistical estimation error, even when the wrong (post-merger) pass-through rates were used. Miller et al. (2012) found larger errors, of up to 30-40%, based on numerical simulations of demand systems commonly used in merger simulation for mergers with true price impact in the 5% range (a cutoff frequently used for merger approval). However these errors were robust across a range of commonly used demand systems and were an order of magnitude smaller than those arising from mis-specifying the curvature of the demand system within this class (e.g. using standard discrete choice demand systems when the real data came from linear demand).  

10Our approximation is equivalent, in the case of Bertrand conduct with constant marginal cost, to the first step of the Newton’s method approach to merger simulation proposed by Froeb et al. (2005), though the justification is different. For example, the second step of their approach does not correspond to the second-order term that would be derived from our expansion, as theirs relies on non-local but first-order information while ours uses local, higher-order derivatives.

11Of course, our formula adds little in the case that any one of these particular demand systems applies. However, given that there are many such smooth demand systems, and many others that are not commonly but are equally smooth and have very different demand curvatures and thus pass-through rates, our approach
This suggests that, if demand in potential mergers is in fact generated from the range of demand systems typically used for simulation, the error in our approximation is likely to be smaller than other common sources of error. Two features of our approximation contribute to this relative accuracy:

1. First, it is more accurate when incentive effects are smaller and so seems to do comparatively well for the price changes that are typically the concern of merger analysis—those in the 5%-10% range.

2. Second, almost all functional forms used in demand estimation are very smooth (have sharply bounded curvature of equilibrium conditions) and are therefore well-approximated by our formula. Since any method that tries to estimate merger effects from exclusively pre-merger data will struggle if the true demand system is not smooth, our method may provide robustness over the class of plausibly empirically applicable models, though not outside it.

Miller et al. (2012) find that our approximation is precise in all cases with linear demand. To illustrate theoretically why this is true, we now return to Example 3.

**Example 1 Continued.** The post-merger equilibrium can be calculated directly by maximizing \(2(P - c)(A - (B_o + B_x)P)\), which gives a first-order condition

\[
0 = -B_o(P - c) + A - B_o P - B_x P - B_x (P - c)
\]

which gives a post-merger price:

\[
P^M = \frac{1}{2} (B_o + B_x)^{-1} [A + (B_x + B_o) c].
\]

Thus, repeatedly using the Slutsky symmetry of the \(B\) matrices and symmetry of products provides a treatment that is robust across this class rather than just being valid for a particular assumed system, as in merger simulation. See Fabinger and Weyl (2012) for a detailed discussion of the restrictions beyond smoothness placed on demand curvature by typical demand systems.
to commute matrix multiplication,

$$
\Delta P = \frac{1}{2} (B_o + B_x)^{-1} [A + (B_x + B_o) c] - (2B_o + B_x)^{-1} (A + B_o c)
$$

$$
= \frac{1}{2} \left[(B_o + B_x) (2B_o + B_x)^{-1} - \left((2B_o + B_x) (A + (B_x + B_o) c) - 2 (B_o + B_x) (A + B_o c)\right) \right]
$$

$$
= \frac{1}{2} \left[(B_o + B_x) (2B_o + B_x)^{-1} B_x [-A + (B_o + B_x) c] \right].
$$

On the other hand, we can compute our approximation. The post-merger first-order condition for $P_i$ is

$$
A - B_o P_i - B_x P_{-i} - B_o P_i + B_o c_i - B_x (P_{-i} - c_{-i}) = 0
$$

$$
A - 2P_i - B_o^{-1} (B_x (2P_{-i} - c_{-i}) + c_i = 0
$$

(2.7)

Taking the derivative of (2.7), which is linear in own cost, with respect to $P_i$ yields $-2I$, where $I$ is the identity matrix of appropriate size. The derivative with respect to $P_{-i}$ is $-2B_o^{-1}B_x$, so that the merger pass-through is

$$
\rho = \frac{1}{2} \begin{bmatrix}
I & B_o^{-1}B_x \\
B_o^{-1}B_x & I
\end{bmatrix}^{-1}
$$

$$
= \frac{1}{2} \begin{bmatrix}
\left(I - (B_o^{-1}B_x)^2\right)^{-1} & \left(I - (B_o^{-1}B_x)^2\right)^{-1} B_o^{-1}B_x \\
B_o^{-1}B_x & \left(I - (B_o^{-1}B_x)^2\right)^{-1}
\end{bmatrix}
$$

Plugging in the GePP ($g(P^0) = -B_o^{-1}B_x (2B_o + B_x)^{-1} [A - (B_o + B_x) c]$) (and again heavily using symmetry) gives

$$
\Delta P \approx -\frac{1}{2} \left(I - (B_o^{-1}B_x)^2\right)^{-1} (I - B_o^{-1}B_x) \cdot B_o^{-1}B_x
$$

$$
= -\frac{1}{2} \left(I + B_o^{-1}B_x\right)^{-1} B_o^{-1} (2B_o + B_x)^{-1} B_x [A - (B_o + B_x) c]
$$

$$
= \frac{1}{2} \left[(B_o + B_x) (2B_o + B_x)^{-1} B_x [-A + (B_o + B_x) c] \right].
$$
This is identical to the expression in (2.6) and thus, in this simple case, our approximation is exact.

In order to simplify notation and shorten calculations, this example focused on the case of two symmetric, pre-merger-Bertrand firms with symmetric products merging to monopoly while facing Slutsky symmetric demand. However, our approximation formula remains exact so long as demand is linear and linear conjectures are maintained: Slutsky symmetry, merger to monopoly and symmetry across and within firms are not required.\textsuperscript{12}

2.4 Role of Pass-Through

Many papers have considered the role of pass-through in determining the price effects of a merger. Shapiro (1996) and Crooke et al. (1999) showed that demand forms with differing curvature but the same elasticities might lead to simulated merger effects differing by an order of magnitude. Froeb et al. (2005) argued that the demand systems that predict large pass-through of efficiencies (which are passed through at post-merger rates) also predict large anticompetitive merger effects. Weyl and Fabinger (2009) and Farrell and Shapiro (2010a) argued informally that because UPP is essentially the opportunity cost of sales created by the merger, multiplying it by the pre-merger pass-through rates should approximate merger effects. Farrell and Shapiro (2010b) show that in some cases bounds on pre-merger pass-through and UPP can bound merger effects.\textsuperscript{13}

We reconcile these views by showing how merger pass-through combines important aspects of both pre-merger and post-merger pass-through. Furthermore, we show that it is pass-through, which is affected by conduct and cost curvature, not demand curvature, that converts GePP into approximate changes in prices.

\textsuperscript{12}A proof is available on request.

\textsuperscript{13}Since they use a constant marginal cost framework under which pass-through and demand curvature are equivalent, it is not clear which is the relevant quantity.
Marginal costs enter quasi-linearly into $f_i$, the pre-merger first-order condition for firm $i$. Thus, if we were to impose a vector of quantity taxes $t$, the post-tax (but pre-merger) equilibrium would be characterized by

$$f(P) + t = 0,$$

so that by the implicit function theorem

$$\frac{\partial P}{\partial t} \frac{\partial f}{\partial P} = -I.$$

The pre-merger pass-through matrix is

$$\rho_{-} \equiv \frac{\partial P}{\partial t} = -\left( \frac{\partial f}{\partial P} \right)^{-1}. \quad (2.8)$$

After the merger between firm $i$ and firm $j$ takes place, the marginal cost of producing firm $i$’s goods enters into $h_i$ (its post-merger first-order conditions) quasi-linearly with a coefficient of 1, but also enters into $h_j$ quasi-linearly with a coefficient of $-\tilde{D}_{ji}$. This follows directly from the fact that the GePP for $j$ includes the mark-up on good $i$ which depends (negatively) on the tax. Thus if we let

$$K = \begin{pmatrix} I & -\tilde{D}_{ij} \\ -\tilde{D}_{ji} & I \end{pmatrix},$$

then the post-merger and post-tax equilibrium is characterized by

$$h(P) = -Kt,$$

which implies that the post-merger pass-through matrix is\(^{14}\)

$$\rho_{\rightarrow} \equiv \frac{\partial P}{\partial t} = -\left( \frac{\partial h}{\partial P} \right)^{-1} K. \quad (2.9)$$

\(^{14}\)The term with $\frac{\partial K}{\partial P} t^0$ drops out because the tax is zero to begin with.
Our result from the previous section is that
\[
P^M - P^0 \approx - \left( \frac{\partial h}{\partial P}(P^0) \right)^{-1} \cdot g(P^0).
\]
Thus, **merger pass-through**, \(- \left( \frac{\partial h}{\partial P} \right)^{-1}\), is not equal to pre-merger pass-through, \(- \left( \frac{\partial f}{\partial P} \right)^{-1}\), or post-merger pass-through, \(- \left( \frac{\partial h}{\partial P} \right)^{-1} K\); rather it depends on the curvature of the latter and the cost structure of the former. This makes sense since the post-merger first-order conditions are relevant, but the opportunity costs are direct shifts in the first-order conditions, not real (physical) costs that would enter wherever the marginal costs enter the post-merger first-order condition.

### 2.4.2 Calculation and approximation of merger pass-through

**Identification**

When can we identify the merger pass-through from the pre-merger pass-through? Since \(\frac{\partial f(P)}{\partial P}\) is equal to the negative inverse of the pass-through matrix, it is clearly calculable. In the case of two single-product firms merging under Bertrand equilibrium, the pass-through matrix, along with the first derivatives of demand, can be used to calculate

\[
\frac{\partial^2 Q_i}{\partial P_i^2}, \frac{\partial^2 Q_i}{\partial P_i \partial P_j}, \frac{\partial^2 Q_j}{\partial P_j^2}, \text{ and } \frac{\partial^2 Q_j}{\partial P_i \partial P_j},
\]

If one assumes Slutsky symmetry \(\left( \frac{\partial Q_i}{\partial P_j} = \frac{\partial Q_j}{\partial P_i} \right)\), then the other second derivatives are

\[
\frac{\partial^2 Q_j}{\partial P_i^2} = \frac{\partial}{\partial P_i} \frac{\partial Q_j}{\partial P_i} = \frac{\partial}{\partial P_j} \frac{\partial Q_i}{\partial P_j} = \frac{\partial}{\partial P_i} \frac{\partial Q_i}{\partial P_j},
\]

which are all that is needed to calculate \(\frac{\partial g(P)}{\partial P}\). Since there is little intuition to be gained from the form of \(\frac{\partial g(P)}{\partial P}\), we leave it to Appendix B.5. A similar procedure may be applied under Cournot competition.

In the case of more than two merging firms, derivatives of the form \(\frac{\partial^2 Q_i}{\partial P_j \partial P_k}\) are needed and
cannot be calculated from observed pass-through rates and first derivatives unless one places restriction on the form of demand. A slightly more restrictive version of the horizontality assumption of Weyl and Fabinger (2009),

\[ Q_i(P) = H \left( P_i + \sum_{j \neq i} F_j(P_j) \right), \]

is sufficient to calculate the necessary second partials, but it is a rather strong assumption.

**Approximation**

The difference between pre-merger and merger pass-through (and post-merger pass-through) may in fact be small. For our approximation to be valid, \( g(P^0) \) and the curvature of the equilibrium conditions need to be jointly sufficiently “small.” If \( g(P^0) \) is small, then it seems likely that \( \frac{\partial g}{\partial P} \) would also be small at \( P^0 \) and thus \( \left( \frac{\partial h}{\partial P} \right)^{-1} \approx \left( \frac{\partial f}{\partial P} \right)^{-1} \) at \( P^0 \). If this were not the case, then while \( g(P^0) \) is small, if \( g(P) \) were evaluated at a relatively close price in the direction of maximal gradient rather than at \( P^0 \) it would then no longer be small. To the extent that the smallness of \( g \) is “fragile” in this sense, it is unlikely to form a solid basis for using first-order approximations.

Thus, in many cases when the first-order approximation would be valid, the merger pass-through is approximately equal to pre-merger pass-through. Furthermore, if small diversion ratios, rather than other factors, cause \( g(P^0) \) to be small, then post-merger pass-through will also be close to merger pass-through since \( K \) will be close to the identity matrix. If a merger is likely to have a small impact on prices, then it is likely to have a small impact on pass-through rates and thus both pre- and post-merger pass-through rates will approximate merger pass-through. Miller et al. (2012) have an even more surprising finding in the numerical simulations they consider: results are typically more accurate when pre-merger pass-through is substituted for the theoretically-desired merger pass-through. While these results of course depend on the particular functional forms used in the simulations, they suggest that substituting pre-merger pass-through for merger pass-through may not cause
large errors. Similarly, Cheung (2011) shows that using post-merger pass-through in place of merger pass-through leads to total approximation error smaller than statistical estimation error in her application.

This interpretation, which views pre-, post- and merger pass-through as close to one another, has a number of benefits. First, it is consistent with the apparent coincidence (Froeb et al., 2005) that demand forms with high pre-merger pass-through rates have been found to generate high pass-through of merger efficiencies (which are driven by post-merger pass-through) and large anti-competitive effects (which are proportional to merger pass-through). Second, it shows that the logic of Froeb et al. (2005) and that of Shapiro and his co-authors are on some level consistent: when pre-merger and post-merger pass-through are good estimates of merger pass-through, they are also similar to each other. Finally, it shows that using intuitions about pre-merger pass-through rates to approximate the rate at which GePP is passed through to prices may be reasonable.

2.5 Welfare Changes

In this section we show how changes in prices calculated in Section 2.3 can be converted into estimates of changes in consumer or social surplus in a market (ignoring externalities and potential cross-market effects of the price changes). Dividing by the value of the market, $P^T Q$, converts these into unit-free indices.

Consumer Surplus

To a first-order, the change in consumer surplus in the evaluated market is just the sum across goods of the change in price times the quantity:

$$\Delta CS \approx -\Delta P^T Q.\textsuperscript{15}$$

\textsuperscript{15}Since we have calculated the first and second derivatives of Q, we could add higher order terms to this approximation, but, since $\Delta P$ itself is an approximation, that would be adding some second order terms and
Social Surplus

The predicted change in social surplus, again ignoring externalities and out-of-market effects, is the sum of the change in quantities, approximated by $\Delta Q \approx \frac{\partial Q}{\partial P} \Delta P$, multiplied by the absolute mark-ups:

$$\Delta SS \approx \Delta Q^T (P - mc) \approx \left(\frac{\partial Q}{\partial P} \Delta P\right)^T (P - mc).$$

The mark-ups can be pre-merger, post-merger or some combination of the two. It would also be natural to include (as an additional term) an expected change in fixed (or infra-marginal) costs due to the merger as in Williamson (1968).

Profits

While changes to profits are not typically an object of regulatory concern, an assumption that these must be positive (by the firms’ revealed preference for merging) may provide some information. If $\Delta F_i$ is the (presumably negative) change in firm $i$’s fixed costs and $\Delta mc_i$ is the (uniform) change in inframarginal costs then

$$\Delta \pi_i \approx (\Delta P_i - \Delta mc_i)^T Q_i + (P - mc_i)^T \left(\frac{\partial Q}{\partial P} \Delta P\right) - \Delta F_i.$$

The incentive for firms $i$ and $j$ to merge is just $\Delta \pi_i + \Delta \pi_j$.

Advantages of Normative Analysis

Estimating a unified, normatively significant quantity, such as the impact on consumer welfare, offers several potential benefits over estimating a group of price effects. First, while in some not others. The formula may be evaluated at pre-merger (in the spirit of Laspeyres) or post-merger (Paasche) quantities or an arithmetic (Marshall-Edgeworth) or geometric (Fisher) average of the two.

16 Using the tax inclusive price would include tax revenue in social surplus in the spirit of Kaplow (2004).

17 See Section 2.2.2.2.3 above for a discussion of changes in marginal cost.

18 A natural direction for future research would be to use such a formula in extending Deneckere and Davidson (1985)’s analysis of the incentives for a merger.
cases it is possible to find remedies addressing particular areas of concern without impacting others, often a package of impacts are inherently tied to one another and must be evaluated as a whole. It may frequently be the case that some of the prices of a firm’s products in a market are predicted to rise (or rise by a large amount) and others to fall (or rise only slightly) after a merger. When making a decision in such a case, it is necessary to aggregate the relevant information. Such an aggregation requires some implicit or explicit normative standard; welfare criteria are the natural choice, intuitively putting the greatest weight on the products with the largest markets.

Additionally, many of the potential benefits and harms of a merger may arise through channels different from or only indirectly related to a change in price. One example is consumption externalities (e.g. network or platform effects): in an industry with advertising-funded media, a primary harm from elevated prices to readers may be the reduction in the readership accessible to advertisers. A welfare standard facilitates accounting for such harms by making them comparable to price harms, as illustrated by White and Weyl (2012), who provide an extension of our formula to allow for benefits and harms from network externalities. These effects and others – such as from innovation or quality adjustment – are typically considered separately from price effects; a social welfare framework can include such effects whenever estimates or guesses as to their welfare effects are available.

### 2.6 Applying the formula

In this section we discuss implications of our results for applied merger analysis. The advantages of our approach, relative to UPP, come at a computational cost: direct use of the formulae we derive requires many more inputs than the calculation of UPP. In this section, we illustrate how one might go about applying our approach in practice.
2.6.1 Simplifying the formula

While it seems that UPP is, in some sense, a simpler calculation than those we suggest, this is because a UPP-based calculation imposes simplifying assumptions. For example, if we were to assume that all firms produced a single product, that conduct were Bertrand, that all cross-product pass-through rates were zero, and that own pass-through rates were symmetric ($\rho$), then our formula would simplify to $\rho \sum_i Q_i UPP_i$.

This is a somewhat extreme example, but it illustrates that beginning with our formula there are numerous simplifying assumptions one might make to reduce the complexity of the analysis. A few categories of assumptions one might consider are:

1. Pass-through: assuming zero cross-pass-through, either across firms or across products in a firm, would simplify the calculations. Alternatively, one could assume symmetry of pass-through rates or a demand structure, such as horizontality discussed in Subsection 2.4.2.4.2, that assumes a relationship between elasticities and relative pass-through rates.

2. Heterogeneity: assuming some form of symmetry across all firms or for non-merging firms can be a reasonable simplification. Grouping all non-merging firms into one can sometimes be an easy way to capture the important relationships.\footnote{See Appendix B.3 for an example.} Slutsky symmetry also reduces the number of parameters one needs to estimate.

3. Conduct: different conduct assumptions may be easier to work with depending on what data is available. For example, when information on diversion ratios comes from survey data on “next favorite alternative,” assuming Nash conduct may be easiest because that data directly reflects demand patterns. If internal firm documents are used, then a conjectures model that fits how firms discuss the reactions of competitors may be appropriate. Adjustments for biases introduced by using an incorrect conduct model could then be made using heuristics as in Subsection 2.2.2.2.4.
Given time limitations and other constraints of the judicial process, some potentially unattractive assumptions will inevitably be imposed. The full force of our formula is likely to be used in only rare cases, though Miller et al. (2012) show how it can be applied empirically with scanner data with identifying assumptions, data requirements, and time demands that do not greatly exceed those of standard empirical methods. When the data and time demands are excessive, our formulation allows easy selection and application of any combination of assumptions – it does not force all industries into one mold. Furthermore, it is easy to conduct GePP analysis under several combinations of assumptions, facilitating the comparison of the resulting conclusions and thus clarifying the exact role each of these assumptions plays. For example, simplifications could be made more extreme on mark-ups, which are often nearly as difficult to estimate as pass-through or conduct, while allowing greater flexibility or robustness on diversion ratios and conduct. Such variations can clarify the robustness (or weakness) of results in any specific case.

2.6.2 Approximateness

While our approach only approximates the effects of mergers, we want to emphasize that this is a direct result of the sparsity of assumptions we make. If one were to assume a functional form for demand, that assumption would define all the higher order terms for the Taylor expansion and yield a precise result. However, in practice, such assumptions typically restrict important quantities, such as pass-through rates and elasticities (Crooke et al., 1999; Weyl and Fabinger, 2009). We understand that it is typically difficult to estimate pass-through rates precisely, but we believe it is preferable to use any information available on these, even if it is informal, or be explicit about assuming their values rather than indirectly constraining them via functional form assumptions. The findings of Miller et al. (2012) seem to confirm this, as they show that the error created from mis-specifying pass-through rates is an order of magnitude greater than that arising from our approximation. Another advantage of our approach is that the robustness of results to differing functional form, cost-side, or conduct
assumptions can be tested by adjusting some of the numbers in the relevant matrices, without building additional computational models.

2.6.3 At which stage should our tools apply?

Merger review typically proceeds in stages, beginning with an initial screen, proceeding through a more thorough investigation if the screen indicate the the possibility of large anti-competitive effects and, if no settlement can be reached, proceeding to a full court case. As Werden and Froeb (2011) emphasize, the first-order approach is usually promoted as appropriate for an initial screen, with some value during an investigation, but inadequate for a thorough investigation or in-court proceedings where a detailed merger simulation will typically be more compelling.

An advantage of our approach is that it avoids a sharp distinction between these different phases. A version of the formula with many assumptions imposed may be used initially to accommodate limited time and data. As more time and data become available these assumptions can gradually be relaxed and replaced with estimates from data or detailed intuitions. If network effects, product repositioning or other factors are thought to be important, they may be incorporated into the analysis (using extensions of our formula as described in Section 2.5), initially in a highly restricted way and, again, more comprehensively as the analysis progresses. Thus our approach aims to provide a framework that can be used at multiple stages of analysis.

2.6.4 Other applications

While our focus has been almost exclusively on merger analysis, our approach and some of our results may apply to problems beyond this narrow context. Our approach to first-order approximation illustrates how local approximations may be used even in analyzing interventions that are in some sense discrete or discontinuous. Of course, this is valid only when the intervention is in some relevant sense small. However, there are many cases of
interest, at least in industrial economics, when an intervention (such as the introduction of a new product or the entry of a new firm) may have only a small impact on the prices of other products and on consumer welfare even though it may constitute a discrete change. In these cases, our technique allows the sufficient statistics or first-order identification approach advocated by Chetty (2009b) and Weyl (2009) to be applied more broadly than was originally envisioned.

2.7 Conclusion

We extend the modeling framework of Telser (1972) to incorporate price as firms’ choice variable and to allow for multi-product firms. We then quantify, in this general setting, the change in pricing incentives created by a merger. Next, we illustrate how first-order approximations may be applied to a discontinuous event such as a merger: using pass-through rates we derive formulae to approximate quantitative effects of mergers on prices and welfare.

In addition to proposing tools for applied merger analysis, we also hope to stimulate further work in this area. Perhaps the most natural extension of our analysis is the work started by Miller et al. (2012) to analyze the accuracy of the first-order approximation for various demand and cost systems. Another step would be to consider an actual second-order approximation to the merger effect, with a focus on what variation would be needed to identify such an approximation and its intuitive interpretation. In a similar spirit, it would be natural to add coordinated effects – changes in the strategy space and conjectures – using a more explicit model of dynamic coordination. The incorporation into our model of non-Jevons effects on consumer welfare, such as those arising when prices affect network size, is an active area of research being pursued by White and Weyl (2012).

Empirical work oriented towards measuring pass-through rates and how they vary across markets will be crucial in helping to calibrate policymakers’ intuitions about these important, but often difficult-to-measure parameters. Similarly, work on understanding the empirical
relationship between pre-merger, post-merger and merger pass-through rates will be important. Such work will help policymakers determine reasonable simplifying assumptions that can safely be made without sacrificing too much accuracy. The formulation of such simplifications is central to making the work here directly relevant to the often severely time-constrained analysis of particular mergers.
Chapter 3: Understanding Behavior in Strategic Settings: Evidence from a Million Rock-Paper-Scissors Games\textsuperscript{1}

Over the last several decades it would be difficult to find an idea that altered the social science landscape more profoundly than game theory. Across economics and its sister sciences, elements of Nash equilibrium are included in nearly every analysis of behavior in strategic settings. For their part, economists have developed deep theoretical insights into how people should behave in a variety of important strategic environments – from optimal actions during wartime to more mundane tasks such as how to choose a parking spot at the mall. Understanding whether people actually behave in accord with theoretical predictions, however, has considerably lagged behind. Although there are important tests of game theory in lab experiments (see, e.g., Dufwenberg and Gneezy (2000); Dufwenberg et al. (2010); Lergetporer et al. (2014); Sutter et al. (2013)), credibly testing whether behavior conforms to theory in the field has been difficult (yet, see Chiappori et al. (2002); Güth et al. (2003); Goette et al. (2012) and the cites therein).

We take a fresh approach to studying strategic behavior in the field, exploiting a unique dataset that allows us to observe play while the information shown to the player changes. In particular, we use data from over one million matches of rock-paper-scissors\textsuperscript{2} played on a

\textsuperscript{1}Co-authored with Dimitris Batzilis, Steven Levitt, John List and Jeffrey Picel

\textsuperscript{2}Two players each play rock, paper, or scissors. Rock beats scissors; scissors beats paper; paper beats rock. If they both play the same, it is a tie. The payoff matrix is in Figure 3.1.
historically popular Facebook application. Before each match (made up of multiple throws), players are shown a wealth of data about their opponent’s past history: the percent of past first throws in a match that were rock, paper, or scissors, the percent of all throws that were rock, paper, or scissors, and all the throws from the opponents’ most recent five games. These data thus allow us to investigate whether, and to what extent, players use information.

<table>
<thead>
<tr>
<th></th>
<th>Rock</th>
<th>Paper</th>
<th>Scissors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Player 1:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rock</td>
<td>(0,0)</td>
<td>(-1,1)</td>
<td>(1,-1)</td>
</tr>
<tr>
<td>Paper</td>
<td>(1,-1)</td>
<td>(0,0)</td>
<td>(-1,1)</td>
</tr>
<tr>
<td>Scissors</td>
<td>(-1,1)</td>
<td>(1,-1)</td>
<td>(0,0)</td>
</tr>
</tbody>
</table>

**Figure 3.1:** Payoffs for a single throw of rock-paper-scissors.

The informational variation makes the strategy space for the game potentially much larger than a basic rock-paper-scissors game. We show, however, that in Nash Equilibrium, players must expect their opponents to mix equally across rock-paper-scissors – same as in the one-shot game. Therefore, a player has no use for information on the opponent’s history when the opponent is playing Nash.

To the extent that an opponent systematically deviates from Nash, however, knowledge of that opponent’s past history can potentially be exploited.\(^3\) Yet, despite the simplicity of the game and the transparency of the information, it is not clear how one should utilize the information provided. Players can use the information to determine whether an opponent’s past play conforms to Nash, but they do not observe the past histories of the opponent’s previous opponents; without seeing what information the opponent was reacting to, it is hard to guess what non-Nash strategy the opponent may be using. Additionally, players are not shown information about their own past play, so if a player wants to exploit an opponent’s reaction, he has to keep track of his own history of play.

\(^3\)If the opponent is not playing Nash, then Nash is no longer a best response. In symmetric zero-sum games like RPS, deviating from Nash is costless if the opponent is playing Nash (since all strategies have an expected payoff of zero), but there is a profitable deviation from Nash if a player thinks he knows what non-Nash strategy his opponent is using.
Because of the myriad of possible responses, we start with a reduced-form analysis of the first throw in each match to describe how players respond to the provided information. We find that players use information: for example, they are more likely to play rock when their opponent has played less paper (which beats rock) or more scissors (which rock beats) on previous first throws. Players have a weak negative correlation across their own first throws. Overall, we find that most players, at some point in their histories, employ strategies consistent with Nash equilibrium. Even so, we do find considerable evidence of disequilibrium play; for example, 53% of players are reacting to information about their opponents’ history.

This finding motivated us to adopt a structural approach to evaluate the performance of two well-known disequilibrium models: level-$k$ and quantal response. The level-$k$ model posits that players are of different types according to the depth of their reasoning about the strategic behavior of their opponents (Stahl, 1993; Stahl and Wilson, 1994, 1995; Nagel, 1995). Players who are $k_0$ do not take into account their opponents’ strategies or incentives. This can either mean that they play randomly (e.g. Costa-Gomes and Crawford (2006)) or that they play some focal or salient strategy (e.g. Crawford and Iriberri (2007); Arad and Rubinstein (2012)). Players who are $k_1$ respond optimally to a $k_0$ player, which in our context means responding to the focal strategy of the opponent’s (possibly skewed) historical distribution of throws; $k_2$ players respond optimally to $k_1$, etc.\(^4\)

Level-$k$ theory acknowledges the difficulty of calculating equilibria and of forming equilibrium beliefs, especially in one shot games. It has been applied to a variety of laboratory games (e.g. Costa-Gomes et al. (2001); Costa-Gomes and Crawford (2006); Hedden and Zhang (2002); Crawford and Iriberri (2007); Ho et al. (1998)), but this is one of the first applications of level-$k$ theory to a naturally-occurring environment (e.g. Bosch-Domenech et al. (2002); Ostling et al. (2011); Gillen (2009); Goldfarb and Xiao (2011); Brown et al. (2012)). We also have substantially more data than most other level-$k$ studies, both in number of observations.

\(^4\)Since the focal $k_0$ strategies can be skewed, our $k_1$ and $k_2$ strategies usually designate a unique throw, which would not be true if $k_0$ were constrained to be a uniform distribution.
and in the richness of the information structure.

We adapt level-$k$ theory to our repeated game context. Empirically, we use maximum likelihood to estimate how often each player plays $k_0$, $k_1$, and $k_2$, assuming that they are restricted to those three strategies. We find that the majority of play is best described as $k_0$ (about 74%). On average, $k_1$ is used in 19% of throws. The average $k_2$ estimate is 7.7%, but for only 12% of players do we reject at the 95% level that they never play $k_2$. Most players use a mixture of strategies, mainly $k_0$ and $k_1$. We also find that 20% of players deviate significantly from $\frac{1}{3}, \frac{1}{3}, \frac{1}{3}$ when playing $k_0$.

An interesting result is that play is more likely to be consistent with $k_1$ when the expected return to $k_1$ is higher. This effect is larger when the opponent has a longer history – that is, when the skewness in history is less likely to be noise. The fact that players respond to the level of the perceived expected $(k_1)$ payoff, not just whether it is the highest, is related to the idea of quantal response: that players’ probability of using a pure strategy is increasing in the relative perceived expected payoff of that strategy. This can be thought of as a more continuous version of a $k_1$ strategy. Rather than always playing the strategy with the highest expected payoff as under $k_1$, the probability of playing a strategy increases with the expected payoff. As the random error in quantal response approaches zero (or the responsiveness of play to the expected payoff goes to infinity) this converges to the $k_1$ strategy.

On average, we find that increasing the expected payoff to a throw by one standard deviation increases the probability it is played by 5.2 percentage points (more than one standard deviation). The coefficient is positive and statistically significant for 60% of players. To interpret these results, one must consider that if players were using the $k_1$ strategy, then we

\[ \begin{align*}
5 \text{As we discuss in Section 3.4 there are several reasons that may explain why we find lower estimates for } k_1 \\
\text{and } k_2 \text{ play than in previous work. Many players may not remember their own history, which is necessary for playing } k_2. \text{ Also, given that } k_0 \text{ is what players would most likely play if they were not shown the information (i.e. when they play RPS outside the application), it may be more salient than in other contexts.}
\end{align*} \]

\[ \begin{align*}
6 \text{Because we think players differ in the extent to which they respond to information and consider expected payoffs, we do not impose the restriction from Quantal Response Equilibrium theory (McKelvey and Palfrey, 1995) that the perceived expected payoffs are correct.}
\end{align*} \]
would also find that expected payoffs have a positive effect on probability of play. Similarly, if players used quantal response, many of their throws would be consistent with \( k_1 \) and our maximum likelihood analysis would indicate some \( k_1 \) play.

Therefore, this evidence does not allow us to formally state which model is a better fit for the data. To preform that task, we compare the model likelihoods to test whether \( k_1 \) or quantal response better explains play. The quantal response model is significantly better than the maximum likelihood for 17.8 percent of players, yet the \( k_1 \) model is significantly better for 18.4 percent of players. We interpret this result as suggesting that there are some players whose strategies are close to \( k_1 \) and a distinct set of players who use strategies resembling quantal response. In sum, our data paint the picture that there is a fair amount of equilibrium play, and when we observe disequilibrium play, extant models have power to explain the data patterns.

The remainder of the chapter is structured as follows. Section 3.1 describes the Facebook application in which the game is played and presents summary statistics of the data. Section 3.2 describes the theoretical model underlying the game, and the concept (and implications) of Nash equilibrium in this setting. Section 3.3 explores how players respond to the information about their opponents’ histories. Section 3.4 explains how we adapt level-\( k \) theory to this context and provides parameter estimates. Section 3.5 adapts the quantal response model to our setting and Section 3.6 uses maximum likelihood to compare the level-\( k \) and quantal response models. Section 3.7 concludes.

3.1 An Introduction to Roshambull

Rock-Paper-Scissors, also known as Rochambeau and jan-ken-pon, is said to have originated in the Chinese Han dynasty, making its way to Europe in the 18th century. To this day, it continues to be played actively around the world. There is even a world Rock-Paper-Scissors
championship sponsored by Yahoo.\textsuperscript{7}

The source of our data is an early Facebook ‘app’ called Roshambull,\textsuperscript{8} which allowed users to play rock-paper-scissors against other Facebook users – either by challenging a specific person to a game or by having the software choose an opponent. It was a very popular app for its era with 340,213 users ($\approx 1.7\%$ of Facebook users) playing at least one match in the first three months of the game’s existence. Users played best-two-out-of-three matches for prestige points known as ‘creds.’ They could share their records on their Facebook page and there was a leader board with the top players’ records.

To make things more interesting for players, before each match the app showed them a “scouting sheet” with information on the opponent’s history of play.\textsuperscript{9} In particular, the app showed each player the opponent’s distribution of throws on previous first throws of a match (and the number of matches) and on all previous throws (and the number of throws), as well as the opponent’s lifetime win record and a play-by-play breakdown of the opponent’s previous five matches. It also shows the opponent’s win-loss records and the number of creds wagered. Figure 3.2 shows a sample screenshot from the game.

Our dataset contains 2,636,417 matches, all the matches played between May 23rd, 2007 (when the program first became available to users) and August 14th, 2007. For each throw, the dataset contains a player ID, match number, throw number, throw type, and the time and date at which the throw was made.\textsuperscript{10} This allows us to create complete player histories

\textsuperscript{7}Rock-paper-scissors is usually played for low stakes, but sometimes the result carries with it more serious ramifications. During the World Series of Poker, an annual $500 per person rock-paper-scissors tournament is held, with the winner taking home $25,000. Rock-paper-scissors was also once used to determine which auction house would have the right to sell a $12 million Cezanne painting. Christie’s went to the 11-year-old twin daughters of an employee, who suggested “scissors” because “Everybody expects you to choose ‘rock’.” Sotheby’s said that they treated it as a game of chance and had no particular strategy for the game, but went with “paper” (Vogel, 2005).

\textsuperscript{8}The name is a combination of a bastardized spelling of Rochambeau and the name of the firm sponsoring the app, Red Bull.

\textsuperscript{9}Bart Johnston, one of the developers said, “We’ve added this intriguing statistical aspect to the game... You’re constantly trying to out-strategize your opponent” (Facebook, 2010).

\textsuperscript{10}Unfortunately we only have a player id for each player; there is no demographic information or information
Figure 3.2: Screenshot of the Roshambull App.

Note: This is a sample screenshot from the start of a game of Roshambull.
at each point in time. Most players play relatively few matches in our three month window: the median number of matches is 5 and the mean number is 15.34. Figure 3.3 shows the distribution of the number of matches a player played.

![Figure 3.3: Distribution of the number of matches played by Roshambull users.](image)

Note: The figure shows the number of clean matches played by the 334,661 players who had at least one clean, completed match. The data has a very long right tail, so all the players with over 200 matches are grouped together in the right-most bar.

Some of our inference depends upon having a large number of observations per player; for those sections, our analysis is limited to the 7751 “experienced” players for whom we observe at least 100 clean matches. They play an average of 195.6 matches; the median is about their out-of-game connections to other players.

We exclude the small fraction of player-pairs for which one player won an implausibly high share of the matches (suggesting collusion). We of course include those matches when forming the histories.
151 and the standard deviation is 141.8. Because these are the most experienced players, their strategies may not be representative; one might expect more sophisticated strategies in this group relative to the Roshambull population as a whole.

For all of the empirical analysis we focus on the first throw in each match. Modeling non-equilibrium behavior on subsequent throws is more complicated because in addition to their opponent’s history, a player may also respond to the prior throws in the match. Table 3.1 summarizes the play and opponents’ histories shown in the first throw of each match, for both the entire sample and the experienced players.

Table 3.1: Summary Statistics of First Throws

<table>
<thead>
<tr>
<th>Variable</th>
<th>Full Sample</th>
<th></th>
<th>Restricted Sample</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>(SD)</td>
<td>Mean</td>
<td>(SD)</td>
</tr>
<tr>
<td>Throw Rock (%)</td>
<td>33.39</td>
<td>(47.16)</td>
<td>32.57</td>
<td>(46.86)</td>
</tr>
<tr>
<td>Throw Paper (%)</td>
<td>34.78</td>
<td>(47.63)</td>
<td>34.57</td>
<td>(47.56)</td>
</tr>
<tr>
<td>Throw Scissors (%)</td>
<td>31.83</td>
<td>(46.58)</td>
<td>32.86</td>
<td>(46.97)</td>
</tr>
<tr>
<td>Opp’s Historical %Rock</td>
<td>34.15</td>
<td>(18.40)</td>
<td>33.59</td>
<td>(13.71)</td>
</tr>
<tr>
<td>Opp’s Historical %Paper</td>
<td>35.12</td>
<td>(18.12)</td>
<td>34.76</td>
<td>(13.45)</td>
</tr>
<tr>
<td>Opp’s Historical %Scissors</td>
<td>30.73</td>
<td>(17.03)</td>
<td>31.65</td>
<td>(12.83)</td>
</tr>
<tr>
<td>Opp’s Historical Skew</td>
<td>9.68</td>
<td>(17.29)</td>
<td>5.39</td>
<td>(12.43)</td>
</tr>
<tr>
<td>Opp’s Historical %Rock (all)</td>
<td>35.36</td>
<td>(11.49)</td>
<td>34.81</td>
<td>(8.58)</td>
</tr>
<tr>
<td>Opp’s Historical %Paper (all)</td>
<td>34.03</td>
<td>(11.25)</td>
<td>34.10</td>
<td>(8.35)</td>
</tr>
<tr>
<td>Opp’s Historical %Scissors (all)</td>
<td>30.61</td>
<td>(10.62)</td>
<td>31.09</td>
<td>(7.98)</td>
</tr>
<tr>
<td>Opp’s History Length</td>
<td>59.06</td>
<td>(125.12)</td>
<td>99.13</td>
<td>(162.14)</td>
</tr>
</tbody>
</table>

Note: The Restricted Sample uses data only from players who play at least 100 matches. The first 3 variables are dummies for when a throw is rock (R), paper (P), or scissors (S). The next 3 are the percentages opponent’s past first throws that were R, P or S. The “all throws” are the corresponding percentages for all of the opponents’ past throws. Skew measures the extent to which the opponent’s history of first throws deviates from random. Opp’s Historical Length is the number of previous matches the opponent played.

\(^{12}\text{For some analyses we only use players who have 100 clean games where the relevant strategies indicate a unique throw, so we use between 5732 and 7751 players.}\)
3.2 Model of the game

A standard game of rock-paper-scissors is a simple $3 \times 3$ zero-sum game. The payoffs are shown in Figure 3.1. The only Nash Equilibrium is for players to mix $\frac{1}{3}, \frac{1}{3}, \frac{1}{3}$ across rock, paper, and scissors. Because each match is won by the first player to win two throws, and players play multiple matches, the strategies in Roshambull are potentially substantially more complicated: players could condition their play on various aspects of their own or their opponents’ histories. A strategy would be a mapping from (1) the match history for the current match so far, (2) one’s own history of all matches played, and (3) the space of information one might be shown about one’s opponent’s history, onto a distribution of throws.

In addition, Roshambull has a matching process operating in the background, in which players from a large pool are matched into pairs to play a match and then are returned to the pool to be matched again. In the Appendix, we formalize Roshambull in a repeated game framework.

Despite this potential for complexity, however, the equilibrium strategies are still simple.

Proposition 9. In any Nash Equilibrium, for every throw of every match, each player correctly expects his opponent to mix $\frac{1}{3}, \frac{1}{3}, \frac{1}{3}$ over rock, paper, and scissors.\footnote{Players could use aspects of their history that are not observable to the opponent as a private randomization devices, but conditional on all information available to the opponent, they must be mixing $\frac{1}{3}, \frac{1}{3}, \frac{1}{3}$.}

Proof. See the Appendix.

The proof uses the fact that it is a symmetric, zero sum game to show that players continuation values at the end of every match must be zero. Therefore players are only concerned with winning the match, and not with the effect of their play on their resulting history. We then show that for each throw in the match, if player A correctly believes that player B is not randomizing $\frac{1}{3}, \frac{1}{3}, \frac{1}{3}$, then player A has a profitable deviation.
Nash Equilibrium implies that players randomize $\frac{1}{3}, \frac{1}{3}, \frac{1}{3}$ both unconditionally and conditionally on any information available to the opponent. Out of equilibrium, players may condition their throws on their or their opponents’ histories in a myriad of ways. The resulting play may or may not result in an unconditional distribution of play that differs substantially from $\frac{1}{3}, \frac{1}{3}, \frac{1}{3}$. In Section 3.3, we present evidence that 83% of players have throw distributions that do not differ from $\frac{1}{3}, \frac{1}{3}, \frac{1}{3}$. Yet, when throw distributions are exploitable, players respond to their opponents’ histories.\textsuperscript{14}

### 3.3 Players respond to information

Before examining the data for specific strategies players may be using, we present reduced-form evidence that players respond to the information available to them. To keep the presentation clear and simple, for each analysis we focus on rock, but the results are similar for paper and scissors.

We start by examining the dispersion across players in how often they play rock. Figure 3.4 shows the distribution across experienced players of the fraction of their last 100 throws that are rock. It also shows this distribution for simulated players who play rock, paper and scissors with probability one-third on each throw. The distribution from the actual data is substantially more disperse than the simulations, suggesting that the fraction of rock played deviates from one-third more than one would expect from pure randomness. Doing a chi-squared test on all of players’ throws we reject uniform random play for 17% of experienced players.

Given this dispersion in the frequency with which players play rock, we test whether players respond to the information they have about their opponents tendency to play rock – the opponents’ historical rock percentage. Table 3.2 bins opponents by their historical percent rock and shows the fraction of paper, rock, and scissors play. Note that the percent paper is

\textsuperscript{14}We also find serial correlation both across throws within a match and across matches, which is inconsistent with Nash Equilibrium.
Figure 3.4: Observed and Simulated Percent of Throws that are Rock

Note: For each of the 7751 players with at least 100 matches we calculate the percent of his or her last 100 throws that were rock (white distribution). We overlay this on the distribution of the percent of throws that are rock for 7751 simulated players who each play 100 throws and throw rock, paper and scissors each with a probability one-third (blue distribution).
Table 3.2: Response to Percent Rock

<table>
<thead>
<tr>
<th>Opp’s Historical %Rock</th>
<th>Paper (%)</th>
<th>Rock (%)</th>
<th>Scissors (%)</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>0% - 25%</td>
<td>24.33</td>
<td>34.12</td>
<td>41.55</td>
<td>216542</td>
</tr>
<tr>
<td>25% - 30%</td>
<td>25.75</td>
<td>34</td>
<td>40.25</td>
<td>259141</td>
</tr>
<tr>
<td>30% - 33(\frac{1}{3}) %</td>
<td>29.84</td>
<td>33.98</td>
<td>36.17</td>
<td>259427</td>
</tr>
<tr>
<td>33(\frac{1}{3}) % - 37%</td>
<td>35.41</td>
<td>33.48</td>
<td>31.1</td>
<td>294979</td>
</tr>
<tr>
<td>37% - 42%</td>
<td>42.74</td>
<td>31.38</td>
<td>25.87</td>
<td>185369</td>
</tr>
<tr>
<td>42% - 100%</td>
<td>51.11</td>
<td>27.61</td>
<td>21.28</td>
<td>219066</td>
</tr>
</tbody>
</table>

Note: Using only players with at least 100 matches, we bin matches by the opponent’s historical percent of rock throws prior to the match. For each range of opponent’s historical percent rock, we show the distribution of rock, paper, and scissors throws the players use.

increasing across the bins and percent scissors is decreasing. Paper goes from less than a third chance to more than a third chance (and scissors goes from more to less) right at the cutoff where rock goes from less often than random to more often than random. The percent rock a player throws does not vary nearly as much across the bins.

For a more quantitative analysis of how this and other information presented to players affects their play, Table 3.3 presents regression results. The dependent variable is binary, indicating whether a player throws rock. The coefficients all come from one regression. The first column is the effect for all players, the second column is the additional effect of the covariates for players in the restricted sample; the third column is the additional effect for those players after their first 99 games. For example, a standard deviation increase in the opponents historical fraction of scissors (.18) increases the probability than an inexperienced player plays rock by 4.5 percentage points \((100 \cdot .18 \cdot .2531)\) and for an experienced player who already played at least 100 games, the increase is 9 percentage points \((100 \cdot .18 \cdot (.2531 + .1409 + .1379))\).

As expected, the effects of the opponents percent of first throws that were paper is negative and gets stronger with experience, and the effect for scissors is positive and gets stronger with experience. The effect of the opponent’s distribution of all throws and lagged throws is
### Table 3.3: Probability of Playing Rock

<table>
<thead>
<tr>
<th></th>
<th>Overall Effect</th>
<th>Additional Effect on Experienced Players</th>
<th>Additional Effect when ≥100 Games</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
</tr>
<tr>
<td>Opp’s Frac Paper (first)</td>
<td>-0.0415***</td>
<td>-0.0695***</td>
<td>-0.0955***</td>
</tr>
<tr>
<td></td>
<td>(0.0022)</td>
<td>(0.0056)</td>
<td>(0.0087)</td>
</tr>
<tr>
<td>Opp’s Frac Scissors (first)</td>
<td>0.2531***</td>
<td>0.1409***</td>
<td>0.1379***</td>
</tr>
<tr>
<td></td>
<td>(0.0023)</td>
<td>(0.0061)</td>
<td>(0.0094)</td>
</tr>
<tr>
<td>Opp’s Frac Paper (all)</td>
<td>0.0018</td>
<td>0.0224*</td>
<td>0.0041</td>
</tr>
<tr>
<td></td>
<td>(0.0033)</td>
<td>(0.0088)</td>
<td>(0.0138)</td>
</tr>
<tr>
<td>Opp’s Frac Scissors (all)</td>
<td>0.0401***</td>
<td>0.0245**</td>
<td>-0.0209</td>
</tr>
<tr>
<td></td>
<td>(0.0035)</td>
<td>(0.0094)</td>
<td>(0.0146)</td>
</tr>
<tr>
<td>Opp’s Paper Lag</td>
<td>0.0042***</td>
<td>-0.0012</td>
<td>-0.0040*</td>
</tr>
<tr>
<td></td>
<td>(0.0007)</td>
<td>(0.0016)</td>
<td>(0.0019)</td>
</tr>
<tr>
<td>Opp’s Scissors Lag</td>
<td>0.0145***</td>
<td>0.0048**</td>
<td>-0.0014</td>
</tr>
<tr>
<td></td>
<td>(0.0008)</td>
<td>(0.0016)</td>
<td>(0.0020)</td>
</tr>
<tr>
<td>Own Paper Lag</td>
<td>0.0001</td>
<td>0.0073***</td>
<td>0.0052**</td>
</tr>
<tr>
<td></td>
<td>(0.0007)</td>
<td>(0.00015)</td>
<td>(0.0019)</td>
</tr>
<tr>
<td>Own Scissors Lag</td>
<td>0.0039***</td>
<td>0.0029</td>
<td>-0.0046*</td>
</tr>
<tr>
<td></td>
<td>(0.0007)</td>
<td>(0.00015)</td>
<td>(0.0019)</td>
</tr>
<tr>
<td>Constant</td>
<td>0.3385***</td>
<td>-0.0106***</td>
<td>-0.0015</td>
</tr>
<tr>
<td></td>
<td>(0.0007)</td>
<td>(0.0009)</td>
<td>(0.0018)</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.0142</td>
<td>0.0239</td>
<td>0.0258</td>
</tr>
<tr>
<td>N</td>
<td>4210005</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: All columns show OLS coefficients for the effect on whether a throw is rock; all the coefficients come from one regression. The first column is the effect for all players, the second column is the additional effect of the covariates for players in the restricted sample; the third column is the additional effect for those players after their first 99 games. Opp’s Fraction Paper refers to the fraction of the opponent’s previous first throws that were paper. Opp’s Fraction Paper (all) refers to the fraction of all the opponent’s previous throws that were paper. Opp’s Paper Lag is a dummy for whether the opponent’s most recent first throw in a match was paper. Own Paper Lag is a dummy for whether the player’s own most recent first throw in a match was paper (similarly for scissors). The regression also control for the opponent’s number of previous matches.
The consistent and strong reactions to the opponent’s distribution of first throws motivates our use of that variable in the structural models.

The fact that players respond to their opponents’ histories makes their play somewhat predictable and potentially exploitable. To measure this exploitability, we run the regression from Table 3.3 on half the restricted sample and use the coefficients to predict the probability of playing rock on each throw for the other half of the restricted sample. We do the same for paper and scissors. We then calculate how often a player who was optimally responding to this predicted play would win, draw, and lose a throw. We compare these rates to the players in our sample, keeping in mind that predicted play is based largely on the opponent’s history, so responding to it optimally would require that the opponent know his own history. Table 3.4 presents the results. If players bet $100 on each throw, the average experienced player would win $1.49 on the average throw. This is better than the $0.66 the average player wins, but someone responding optimally to their predictability would win $16.86 on average.

### Table 3.4: Win Percentages

<table>
<thead>
<tr>
<th>Data</th>
<th>Wins (%)</th>
<th>Draws (%)</th>
<th>Losses (%)</th>
<th>W - L (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full Sample</td>
<td>34.14</td>
<td>32.38</td>
<td>33.48</td>
<td>0.66</td>
</tr>
<tr>
<td>Experienced Sample</td>
<td>34.65</td>
<td>32.18</td>
<td>33.16</td>
<td>1.49</td>
</tr>
<tr>
<td>Best Response to Predicted</td>
<td>42.14</td>
<td>29.7</td>
<td>25.28</td>
<td>16.86</td>
</tr>
</tbody>
</table>

Note: Experienced refers to players who play at least 100 games. “Best Response” is how a player playing against the experienced sample would do if she always played the best response to how the player is predicted to play by Table 3.3. We used half the sample to calculate the coefficients, which were used to predict the play for the other half. W-L is the difference between the first and third column; it equals the average winnings per throw if players bet $100 on a throw.

Given these incentives to exploit players’ predictability, we want to check whether their opponents do. They do not appear to. Given the predicted probabilities of play for experienced players...
players, we calculate the expected payoff to an opponent of playing rock. Table 3.5 bins throws by the expected payoff to playing rock and shows the distribution of opponent throws. The probability of playing rock bounces around – opponents are not more likely to play rock when the actual expected payoff is high; the predictability of players’ throws is not effectively exploited.

Table 3.5: Opponents’ Response to Expected Payoff of Rock

<table>
<thead>
<tr>
<th>Opponent’s Expected Payoff of Rock</th>
<th>Opponent’s Throw (%)</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>[−1, −0.666]</td>
<td>29.5  42.57  27.93</td>
<td>3183</td>
</tr>
<tr>
<td>[−0.666, −0.333]</td>
<td>29.57 41.34  29.09</td>
<td>14723</td>
</tr>
<tr>
<td>[−0.333, 0]</td>
<td>32.65 33.81  33.54</td>
<td>354952</td>
</tr>
<tr>
<td>[0, 0.333]</td>
<td>34.53 32.46  33</td>
<td>329709</td>
</tr>
<tr>
<td>[0.333, 0.666]</td>
<td>35.42 33.85  30.73</td>
<td>11441</td>
</tr>
</tbody>
</table>

Note: Using the players’ predicted play from Table 3.3 (and similar for paper and scissors), we calculate the expected payoff to their opponents of playing rock. This table shows the distribution of opponents’ play for different ranges of that expected payoff.

Since players are mostly responding to their opponent’s history, exploiting those response requires that a player remember her own history of play (since the game does not show one’s own history). So it is perhaps not surprising that players’ predictability is not exploited and therefore unsurprising that they react in a predictable manner. If we do the analysis at the player level, 53% of players significantly respond to their opponents’ historical distributions of past throws.

Having described in broad terms how players react to the information presented, we turn to existing structural models to test whether play is consistent with these hypothesized non-equilibrium strategies.
3.4 Level-\(k\) behavior

While level-\(k\) theory was developed to analyze single shot games, it is a useful framework for exploring how players incorporate information about their opponent.\(^\text{17}\) The \(k_0\) strategy is to ignore the information about one’s opponent and play a (possibly random) strategy independent of the opponent’s history. While much of the existing literature assumes that \(k_0\) is uniform random, some studies assume that \(k_0\) players use a salient or focal strategy. In this spirit, we allow players to randomize non-uniformly (imperfectly) when playing \(k_0\) and assume that the \(k_1\) strategy best responds to a focal strategy for the opponent — \(k_1\) players best respond to the opponent’s past distribution of first throws.\(^\text{18}\) It seems natural that a \(k_1\) player who assumes his opponent is non-strategic would use this description of past play as a predictor of future play. When playing \(k_2\), players assume that their opponents are playing \(k_1\) and respond accordingly.

Given players’ assumptions about their opponents’ play, their strategies then depend on the value function they are maximizing. We assume that players are myopic and ignore the effect of their throw on their continuation value.\(^\text{19}\) This approach is consistent with the literature that analyzes some games as “iterated play of a one-shot game” instead of as an infinitely repeated game (Monderer and Shapley, 1996). More generally, we think it is a reasonable simplifying assumption. While not impossible, it is hard to imagine how one would manipulate one’s history to affect future payoffs with an effect large enough to outweigh the effect on this period’s payoff.\(^\text{20}\)

\(^{17}\)Though players play multiple games, they might struggle to form accurate beliefs about opponents’ strategies since players are playing against many different opponents each of whom may be using a complicated mixed strategy.

\(^{18}\)The reduced form results indicate that players react much more strongly to the distribution of first throws than to the other information provided.

\(^{19}\)In the proof of Proposition 9 we show that in Nash Equilibrium, histories do not affect continuation values, so in equilibrium it is a result, not an assumption, that players are myopic. However, out of Nash Equilibrium, it is possible that what players throw now can affect their probability of winning later rounds.

\(^{20}\)One statistic that we thought might affect continuation values is the skew of a player’s historical
Formal definitions of the different level-$k$ strategies in our context are as follows:

**Definition.** When a player uses the $k_0$ strategy in a match, his choice of throw is unaffected by his history or his opponent’s history.

We should note that using $k_0$ is not necessarily unsophisticated. It could be playing the Nash equilibrium strategy. However there are two reasons to think that $k_0$ might not represent sophisticated play. First, for some players the frequency distribution of their $k_0$ play differs significantly from $\frac{1}{3}, \frac{1}{3}, \frac{1}{3}$, suggesting that if they are trying to play Nash, they are performing poorly. Second, more subtly, it is not sophisticated to play the Nash equilibrium if your opponents are failing to play Nash. With most populations who play the beauty contest game, people who play Nash do not win (Nagel, 1995). In RPS, if there is a possibility that one’s opponent is playing something other than Nash, there is a strategy that has a positive expected return, whereas Nash always has a zero expected return. (If it turns out the opponent is playing Nash, then every strategy has a zero expected return and so there is little cost to trying something else.) Given that some players differ from $\frac{1}{3}, \frac{1}{3}, \frac{1}{3}$ when playing $k_0$ and most don’t always play $k_0$, Nash is frequently not a best response.

**Definition.** When a player uses the $k_1$ strategy in a match, he plays the throw that has the highest expected payoff on this throw if his opponent randomizes according to the opponent’s historical distribution of first throws.$^{21}$

We have not specified how a player using $k_0$ chooses a throw, but provided the process is not changing over time, his past throw history is a good predictor of play in the current distribution. As a player’s history departs further from random play, the more opportunity for opponent response and player manipulation of opponent response. We ran multinomial logits for each player on the effect of own history skewness on the probability of winning, losing or drawing. The coefficients were significant for less than 5% of players. The mean coefficient implied that if a player’s skewness is a standard deviation higher (relative to the population) her probability of winning is .37 percentage points higher. This provides some support to our assumption that continuation values are not a primary concern.

$^{21}$Sometimes opponents’ distributions are such that there are multiple throws that are tied for the highest expected payoff. For our baseline specification we ignore these throws. As a robustness check we define alternative $k_1$-strategies where one throw is randomly chosen to be the $k_1$ throw when payoffs are tied or where both throws are considered consistent with $k_1$ when payoffs are tied. The results to do not change substantially.
match. To calculate the $k_1$ strategy for each throw, we calculate the expected payoff to each of rock, paper, and scissors against a player who randomizes according to the distribution of the opponent’s history. The $k_1$ strategy is the throw that has the highest expected payoff.

**Definition.** When a player uses the $k_2$ strategy in a match, he plays the throw that is the best response if his opponent randomizes uniformly between the throws that maximize her expected payoff against the player’s own historical distribution.

The $k_2$ strategy is to play “the best response to the best response” to one’s own history. In this particular game $k_2$ is in some sense harder than $k_1$ because the software shows only one’s opponent’s history, but players could keep track of their own history.

Having defined the level-$k$ strategies in our context, we now turn to the data for evidence of level-$k$ play.

### 3.4.1 Reduced-form evidence for level-$k$ play

One proxy for $k_1$ and $k_2$ play is players choosing throws that are consistent with these strategies. Whenever a player plays $k_1$ (or $k_2$) her throw is consistent with that strategy. However, the converse is not true. Players playing the NE strategy of $\frac{1}{3}, \frac{1}{3}, \frac{1}{3}$ would be consistent with $k_1$ a third of the time (on average).

For each player we calculate the fraction of throws that are $k_1$-consistent; these fractions are upper bounds on the amount of $k_1$ play. The highest percentage of $k_1$-consistent behavior for an individual in our restricted sample is 84.6 percent, indicating that no player uses $k_1$ consistently. Figure 3.5a shows the distribution of the fraction of $k_1$-consistency across players. It suggests that at least some players use $k_1$ at least some of the time: the distribution is to the right of the vertical $\frac{1}{3}$-line and there is a right tail. To complement the graphical evidence, we formally test whether the observed frequency of $k_1$-consistent play is significantly greater than expected under random play. Using this test, we can reject the null of no $k_1$ play at a 95 percent confidence interval for 71.7 percent of players in the sample.
Figure 3.5: Level-$k$ consistency

Note: These graphs show the distribution across players of the fraction of throws that are $k_1$- and $k_2$-consistent. They include the 6674 players who have 100 games with well-defined $k_1$ and $k_2$ strategies. The vertical line indicates $\frac{1}{3}$, which we would expect to be the center of the distribution if throws were random.

Given that players seem to play $k_1$ some of the time, players could benefit from playing $k_2$. Figure 3.5b shows the distribution of the fraction of actual throws that are $k_2$-consistent. Perhaps unsurprisingly given that players are not shown the necessary information, we do not find much evidence of $k_2$ play. The observed frequency of $k_2$ play is significantly greater than expected with random play for only 7.5 percent of players, barely more than the 5% we would find if no one played $k_2$.

If we assume that players use either $k_0$, $k_1$, or $k_2$ we can use the percentage of throws that are consistent with neither $k_1$ nor $k_2$ to obtain a lower bound on how often each player is playing $k_0$. We calculate $k_0 = 1 - \bar{k}_1 - \bar{k}_2$. We do not expect this bound to be tight because, in expectation, a randomly chosen $k_0$ play will be consistent with either the $k_1$ or $k_2$ strategy relatively often. The mean lower bound across players is 37 percent. The minimum is 9.3 percent and the maximum is 74 percent.

Multinomial Logit

Before turning to the structural model, we can use a multinomial logit model to explore whether a throw being $k_1$-consistent increases the probability that a player chooses that throw.
Figure 3.6: Distribution across players of the coefficient in the multinomial logit.

Note: The graph shows distribution across 6670 players of the coefficient from the multinomial logit
\[ U_i = \alpha_i + \beta \cdot 1\{k_1 = i\} + \varepsilon_i, \]
where \( i = r, p, s \) and \( 1\{k_1 = i\} \) is an indicator for when \( i \) is the \( k_1 \)-consistent thing to do.

Figure 3.6 shows the distribution of \( \beta \)'s across players. The mean is .532.

For each player, we estimate a multinomial logit where the utilities are

\[ U_i = \alpha_i + \beta \cdot 1\{k_1 = i\} + \varepsilon_i, \]

where \( i = r, p, s \) and \( 1\{k_1 = i\} \) is an indicator for when the \( k_1 \)-consistent action is to throw \( i \).

The marginal effect varies slightly with the baseline probabilities, 
\[ \frac{\partial Pr\{i\}}{\partial x_i} = \beta \cdot Pr\{i\}(1 - Pr\{i\}), \]
but is approximately \( \frac{1}{3} (1 - \frac{1}{3}) = \frac{2}{9} \) times the coefficient. Hence, on average, a throw being \( k_1 \)-consistent means it is 12 percentage points more likely to be played. Given that the standard deviation across players in the percent of rock, paper, or scissors throws is about 5 percentage points, this is a large average effect. The individual level coefficient is positive and significant for 61 percent of players.
3.4.2 Maximum likelihood estimation of a structural model of level-$$k$$ thinking

The results presented in the previous sections provide some evidence as to what strategies are being employed by the players in our sample, but they do not allow us to identify with precision the frequency with which strategies are employed. To obtain point estimates of each player’s proportion of play by level-$$k$$, along with standard errors, we need additional assumptions.

**Assumption 1.** All players use only the $$k_0$$, $$k_1$$, or $$k_2$$ strategies in choosing their actions.

Assumption 1 restricts the strategy space, ruling out any approach other than level-$$k$$, and also restricting players not to use levels higher than $$k_2$$. We limit our modeling to levels $$k_2$$ and below, both for mathematical simplicity and because there is little reason to believe that higher levels of play are commonplace, both based on the low rates of $$k_2$$ play in our data, and rarity of $$k_3$$ and higher play in past experiments.$^{22}$

**Assumption 2.** Whether players chose to play $$k_0$$, $$k_1$$, or $$k_2$$ on a given throw is independent of which throw (rock, paper, or scissors) each of the strategies would have them play.

Assumption 2 implies, for example, that the likelihood that a player chooses to play $$k_2$$ will not depend on whether it turns out that the best $$k_2$$ action is rock or is paper. This independence is critical to the conclusions that follow. Note that Assumption 2 does not require that a player commit to having the same probabilities of using $$k_0$$, $$k_1$$, and $$k_2$$ strategies across different throws.

Given these assumptions we can calculate the likelihood of observing a given throw in terms of 4 parameters, given in Table 3.6. Given these parameters, the probability of

---

$^{22}$As an aside, in the case of rock-paper-scissors the level $$k + 6$$ strategy is identical to the level $$k$$ strategy for $$k \geq 1$$, so it is impossible to identify levels higher than 6. One might expect $$k$$ to be equivalent to $$k + 3$$, but levels 1, 3, and 5 depend on the opponent’s history, with one being rock, one being paper, and one being scissors, while levels 2, 4, and 6 depend on one’s own history. So with many games all $$k < 7$$ are separately identified. This also implies that the $$k_1$$ play we observe could in fact be $$k_7$$ play, but we view this as highly unlikely.
Table 3.6: Parameters of the Structural Model

<table>
<thead>
<tr>
<th>Variable</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{r}_0$</td>
<td>fraction of the time a player plays $k_0$ and chooses rock</td>
</tr>
<tr>
<td>$\hat{p}_0$</td>
<td>fraction of the time a player plays $k_0$ and chooses paper</td>
</tr>
<tr>
<td>$\hat{s}_0$</td>
<td>fraction of the time a player plays $k_0$ and chooses scissors</td>
</tr>
<tr>
<td>$\hat{k}_1$</td>
<td>fraction of the time a player plays $k_1$</td>
</tr>
<tr>
<td>$(\hat{k}_2)$</td>
<td>$1 - \hat{k}_1 - \hat{r}_0 - \hat{p}_0 - \hat{s}_0$ (not an independent parameter)</td>
</tr>
</tbody>
</table>

$\hat{r}_0$ is not equal to the fraction of $k_0$ throws that are rock; that conditional probability is given by $\frac{\hat{r}_0}{\hat{r}_0 + \hat{p}_0 + \hat{s}_0}$.

observing a given throw $i$ is

\[
\hat{k}_1 \cdot 1\{k_1 = i\} + \hat{k}_2 \cdot 1\{k_2 = i\} + (\hat{r}_0 \cdot 1\{i = r\} + \hat{p}_0 \cdot 1\{i = p\} + \hat{s}_0 \cdot 1\{i = s\}),
\]

where $1\{\cdot\}$ is an indicator function, equal to one when the statement in braces is true and zero otherwise. This reflects the fact that the throw will be $i$ if the player plays $k_1$ and the $k_1$ strategy says to play $i$ ($\hat{k}_1 \cdot 1\{k_1 = i\}$) or the player plays $k_2$ and the $k_2$ strategy says to play $i$ ($\hat{k}_2 \cdot 1\{k_2 = i\}$) or the player plays $k_0$ and chooses $i$ ($\hat{r}_0 \cdot 1\{i = r\} + \hat{p}_0 \cdot 1\{i = p\} + \hat{s}_0 \cdot 1\{i = s\}$).

For each player, the overall log-likelihood depends on twelve statistics from the data. For each throw type ($i = R, P, S$), let $n_{i12}$ be the number of throws that are of type $i$ and consistent with $k_1$ and $k_2$, $n_i^1$ the number of $i$ throws consistent with just $k_1$, $n_i^2$ the number of $i$ throws consistent with just $k_2$, and $n_i^0$ the number of $i$ throws consistent with neither $k_1$ nor $k_2$. Given these statistics, the log-likelihood function is

\[
L(\hat{k}_1, \hat{k}_2, \hat{r}_0, \hat{p}_0, \hat{s}_0) = \sum_{i = r,p,s} \left( n_{i12}^1 \ln(\hat{k}_1 + \hat{k}_2 + \hat{r}_0) + n_{i1}^1 \ln(\hat{k}_1 + \hat{r}_0) + n_{i2}^1 \ln(\hat{k}_2 + \hat{r}_0) + n_{i0}^1 \ln(\hat{r}_0) \right).
\]

For each player we use maximum likelihood to estimate $(\hat{k}_1, \hat{k}_2, \hat{r}_0, \hat{p}_0, \hat{s}_0)$. Given the estimates, standard errors are calculated analytically.\(^{23}\)

\(^{23}\)We derive the Hessian of the likelihood function, plug in the estimates, and take the inverse.
Table 3.7: Summary of $k_0$, $k_1$, and $k_2$ estimates.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>SD</th>
<th>Median</th>
<th>Min</th>
<th>Max</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k_0$</td>
<td>73.79</td>
<td>15.7</td>
<td>75.2</td>
<td>18.65</td>
<td>100</td>
<td>6635</td>
</tr>
<tr>
<td>$k_1$</td>
<td>18.52</td>
<td>14.48</td>
<td>16.12</td>
<td>0</td>
<td>76.66</td>
<td>6635</td>
</tr>
<tr>
<td>$k_2$</td>
<td>7.69</td>
<td>7.76</td>
<td>5.87</td>
<td>0</td>
<td>40.65</td>
<td>6635</td>
</tr>
</tbody>
</table>

Note: Based on the 6635 players who have 100 clean matches where the $k_1$ and $k_2$ strategies are well-defined.

Table 3.8: Percent of players we reject always or never playing a strategy.

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>95% CI does not include 0</th>
<th>95% CI does not include 1</th>
<th>95% CI does not include 0 or 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k_0$</td>
<td>6399</td>
<td>93.06</td>
<td>58.26</td>
<td>57.45</td>
</tr>
<tr>
<td>$k_1$</td>
<td>6399</td>
<td>62.78</td>
<td>99.97</td>
<td>62.78</td>
</tr>
<tr>
<td>$k_2$</td>
<td>6399</td>
<td>11.55</td>
<td>95.55</td>
<td>11.55</td>
</tr>
</tbody>
</table>

Note: All percentages refer to the 6399 players who have 100 matches where the $k_1$ and $k_2$ strategies are well-defined and for whom we can calculate standard errors on the estimates.

Table 3.7 summarizes the estimates of $k_0$, $k_1$, and $k_2$: the average player uses $k_0$ for 73.8 percent of throws, $k_1$ for 18.5 percent of throws and $k_2$ for 7.7 percent of throws. Weighting by the precision of the estimates or by the number of games does not change these results substantially. As the minimums and maximums suggest, these averages are not the result of some people always playing $k_1$ while others always play $k_2$ or $k_0$. Most players mix, using a combination of mainly $k_0$ and $k_1$.24

Table 3.8 reports the share of players for whom we can reject with 95 percent confidence their never playing a particular level-$k$ strategy. Almost all players (93 percent) appear to use $k_0$ at some point. About 63 percent of players use $k_1$ at some stage, but we can reject exclusive use of $k_1$ for all but two out of 6,399 players. Finally, only about 12 percent of players are we confident use $k_2$.25

24 Other work has found evidence of players mixing levels of sophistication of across different games, e.g. Georganas et al. (2010).

25 Always playing $k_1$ would allow a player to win 34.5% (and lose 32.9%) of the time; always playing $k_2$ leads to wins 42.1% (and losses 28.0%) of the time. It seems that memory or informational constraints prevent players from employing what would be a very effective strategy.
Table 3.9: Summary of $ch_0$, $ch_1$, and $ch_2$ estimates.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>SD</th>
<th>Median</th>
<th>Min</th>
<th>Max</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>$ch_0$</td>
<td>74.85</td>
<td>16.03</td>
<td>76.84</td>
<td>16.99</td>
<td>100</td>
<td>6853</td>
</tr>
<tr>
<td>$ch_1$</td>
<td>16.31</td>
<td>14.08</td>
<td>13.78</td>
<td>0</td>
<td>76.66</td>
<td>6853</td>
</tr>
<tr>
<td>$ch_2$</td>
<td>8.85</td>
<td>6.89</td>
<td>8.12</td>
<td>0</td>
<td>50.61</td>
<td>6853</td>
</tr>
</tbody>
</table>

Note: Based on the 6853 players who have 100 clean matches where the $ch_1$ and $ch_2$ strategies are well-defined.

For each player, we can also examine the estimated fraction of rock, paper, and scissors when they play $k_0$. The distribution differs significantly from random uniform for 1257 or 20% of players, similar to the fraction of players whose raw throw distributions differ significantly from uniform (17%).

### 3.4.3 Cognitive Hierarchy

The idea that players might use a distribution over the level-$k$ strategies naturally connects to the Cognitive Hierarchy model of Camerer et al. (2004). They also model players as having different levels of reasoning, but the higher types are more sophisticated than in level-$k$.

The strategies for $ch_0$ and $ch_1$ are the same as $k_0$ and $k_1$; $ch_2$ assumes that other players are playing either $ch_0$ or $ch_1$, in proportion to their actual use in the population, and best responds to that mixture. To test if this more sophisticated version of two levels of reasoning fits the data better, we do another maximum likelihood estimation. The definitions of $ch_0$ and $ch_1$ are the same as $k_0$ and $k_1$.

**Definition.** When a player uses the $ch_2$ strategy in a match, he plays the throw that is the best response to the opponent playing according to the opponent’s historical distribution 79.94% of the time and randomizing between the throws that maximize expected payoff against the player’s own historical distribution the other 20.05% of the time.

The percents come from observed frequencies in the level-$k$ estimation. When players play either $k_0$ or $k_1$, they play $k_0 \frac{79.79}{73.79+18.52} = 79.94\%$ of the time.\(^\text{26}\)

\(^{26}\)To fully calculate the equilibrium, we could repeat the analysis using the frequencies of $ch_0$ and $ch_1$.
Analogous to Assumptions 1 and 2 above, we assume that players use only $ch_0$, $ch_1$ and $ch_2$, and that which strategy they chose is independent of what throw the strategy dictates.

Table 3.9 summarizes the estimates: the average player uses $ch_0$ for 74.9 percent of throws, $ch_1$ for 16.31 percent of throws, and $ch_2$ for 8.85 percent of throws. Weighting by the precision of the estimates or by the number of games a player plays does not change these substantially. These results are similar to what we found for level-$k$ strategies, suggesting that the low rates of using two iterations of reasoning were not a result of restricting that strategy to ignoring $k_0$ play.

3.4.4 Naive level-$k$ strategies

Even if a player expects his opponent to play as she did in the past, he may not calculate the expected return to each strategy. Instead he may employ the simpler strategy of playing the throw that beats the opponent’s most common historical throw. Put another way, he may only consider maximizing his probability of winning instead of weighing it against the probability of losing as is done in an expected payoff calculation. We consider this play naive and define alternative versions of $k_1$ and $k_2$ accordingly.

**Definition.** When a player uses the naive $k_1$ strategy in a match, he plays the throw that will beat the throw that his opponent has played most frequently in the past.

**Definition.** When a player uses the naive $k_2$ strategy in a match and has played throw $i$ most frequently in the past, then he plays the throw that beats the throw that beats $i$.

Table 3.10 summarizes the estimates for naive play. The average player uses $k_0$ for 72.3 percent of throws, naive $k_1$ strategy for 21.1 percent of throws and naive $k_2$ strategy for 6.7 percent of throws. As before, weighting by the precision of the estimates or by the number of games a player plays does not change these results substantially. Most players use a mixed strategy, mixing primarily over $k_0$ and naive $k_1$ strategy.
Table 3.10: Summary of Naive $k_0$, $k_1$, and $k_2$ estimates.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>SD</th>
<th>Median</th>
<th>Min</th>
<th>Max</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k_0$</td>
<td>72.25</td>
<td>17.73</td>
<td>74.72</td>
<td>6.54</td>
<td>100</td>
<td>5732</td>
</tr>
<tr>
<td>$k_1$</td>
<td>21.06</td>
<td>16.71</td>
<td>18.04</td>
<td>0</td>
<td>93.46</td>
<td>5732</td>
</tr>
<tr>
<td>$k_2$</td>
<td>6.69</td>
<td>7.71</td>
<td>4.39</td>
<td>0</td>
<td>45.75</td>
<td>5732</td>
</tr>
</tbody>
</table>

Note: Based on the 5732 players who have 100 clean matches where the naive $k_1$ and $k_2$ strategies are well-defined.

Compared with our standard level-$k$ model, the results for our naive level-$k$ model show slightly more use of the *naive $k_1$ strategy*. The average player used it for 21.1 percent of throws compared to 18.6 percent for standard $k_1$ in the former model. In the naive level-$k$ model, players used $k_0$ and the *naive $k_2$ strategy* slightly less often compared to their standard level-$k$ counterparts.

The opposite naive strategy would be for players to minimize their probability of losing, playing the throw that is least likely to be beat. Running the same model for that strategy we find almost no evidence of $k_1$ or $k_2$ play, suggesting that players are more focused on the probability of winning.

3.4.5 Comparison to the literature

All three of these related models suggest that players of Roshambull use considerably fewer levels of iteration in their reasoning process compared to estimates from other games and other experiments. Bosch-Domenech et al. (2002) found that less than a fourth of the players who used the k-strategies discussed here were $k_0$ players. Whereas, depending on the model, we found between 72.25% and 74.85% of plays involved zero iterations of reasoning. Camerer et al. (2004) suggest that players iterate 1.5 steps on average in many games. In comparison, in our level-$k$ model we find that our average player uses $1 \times .185 + 2 \times .077 = .339$ levels of iterated rationality. Stahl and Wilson (1994) reported that an insignificant fraction of players were $k_0$, 24 percent were $k_1$ players, 49 percent were $k_2$ players, and the remaining 27 percent think this computationally intense exercise is necessary.

100
were “Nash types.” In contrast, we found that the majority of plays were $k_0$ (ch0) and that $k_1$ (ch1) outnumbered $k_2$ (ch2), though in this game $k_0$ is closer to Nash than either $k_1$ or $k_2$. The dearth of $k_2$ play is especially striking in our context given the high returns to playing $k_2$.

One explanation for the differences between our results and the past literature is that most of the players do not deviate substantially from equilibrium play, making the expected payoffs to $k_1$ relatively small. Also, the set-up of rock-paper-scissors does not suggest a level-$k$ thinking mindset as strongly as the p-beauty contest games or other games specifically designed to measure level-$k$ behavior. Our more flexible definition of $k_0$ play may also explain its higher estimate. The low level of $k_2$ play is likely a result of the fact that the Facebook application did not show players their histories so players had to keep track of that on their own in order to effectively play $k_2$.

Another explanation is that we restrict the strategy space to exclude both Nash Equilibrium and different ways in which the players can react rationally to their opponent’s information. It seems players respond more to the first throw history than other information, but there may be many other strategies which are rationalizable in ways which we do not model. Bosch-Domenech et al. (2002), for example, considered equilibrium, fixed point, degenerate and non-degenerate variants of iterated best response, iterated dominance, and even experimenter strategies. Not all of these translate into the RPS set-up, but any strategies that our model left out might look like $k_0$ play when the strategy space is restricted.

### 3.4.6 When are players’ throws consistent with $k_1$?

Though we find relatively low levels of $k_1$ play, we do find some and the result that many of the players seem to be mixing strategies raises the question of when they chose to play $k_0$, $k_1$, and $k_2$. Our structural model assumes that which strategy players chose is independent of the throw dictated by each of the strategies. It does not require that which strategy they chose be independent of the expected payoffs, but the MLE model cannot give us insight into
how expected payoffs may affect play. This is partially because the MLE model does not allow us to categorize individual throws as being a given strategy.

Therefore, to try to get at when players use $k_1$, we return to using $k_1$-consistency as a proxy for possible $k_1$ play. We test two hypotheses. First, the higher the expected payoff to playing $k_1$, the more likely a player is to play $k_1$. For example, if an opponent is $k_0$, the expected returns to playing $k_1$, relative to playing randomly, are much higher if the opponent’s history (or expected distribution) is 40 percent rock, 40 percent paper, 20 percent scissors than if it is 34 percent rock, 34 percent paper, 32 percent scissors.

The second hypothesis is that a player will react more to a higher $k_1$ payoff when his opponent has played more games. A 40 percent rock, 40 percent paper, 20 percent scissors history is more informative if it is based 100 past throws than if it is based on only 10 throws.\(^{27}\)

We also analyze whether these effects vary by player experience; we interact all the covariates with whether a player is in the restricted sample (they eventually play $\geq 100$ matches) and whether they have played 100 matches before the current match.

Table 3.11 presents empirical results from testing these hypotheses. The $k_1$ payoff is the expected payoff to playing $k_1$ assuming the opponent randomizes according to his history. Its standard deviation is .25, so for inexperienced players a one standard deviation increase in payoff to the $k_1$ strategy, increases the probability the throw is $k_1$-consistent by 1.8 percentage points when opponents have a short history, 10 percentage points when opponents have a medium history and 15 percentage points when opponents have played over 94 games. Given that 45% of all throws are $k_1$-consistent, these latter two effects are substantial. Experienced players react slightly less to the $k_1$-payoff when opponents have short histories, but their reactions to opponents with medium or long histories are somewhat larger.

\(^{27}\)Similar predictions could be made about $k_2$ play; however, since we find that $k_2$ is used so little, we do not model $k_2$ play in this section.
Table 3.11: Effect of Expected \( k_1 \) Payoff on \( k_1 \)-Consistency (OLS)

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>K1 Payoff</td>
<td>0.076***</td>
<td>0.107***</td>
<td>0.071***</td>
</tr>
<tr>
<td></td>
<td>(0.0012)</td>
<td>(0.0014)</td>
<td>(0.0017)</td>
</tr>
<tr>
<td>High Opp Exp</td>
<td></td>
<td>-0.076***</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0017)</td>
<td></td>
</tr>
<tr>
<td>Medium Opp Exp</td>
<td></td>
<td>-0.056***</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0015)</td>
<td></td>
</tr>
<tr>
<td>K1 Payoff X High Opp Exp</td>
<td>0.541***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.011)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>K1 Payoff X Medium Opp Exp</td>
<td>0.324***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0051)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experienced</td>
<td>0.014***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0016)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exp X K1 Payoff</td>
<td>0.061***</td>
<td>0.043***</td>
<td>-0.005</td>
</tr>
<tr>
<td></td>
<td>(0.0039)</td>
<td>(0.0039)</td>
<td>(0.0051)</td>
</tr>
<tr>
<td>Exp X High Opp Exp</td>
<td></td>
<td>-0.007**</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0038)</td>
<td></td>
</tr>
<tr>
<td>Exp X Medium Opp Exp</td>
<td></td>
<td>-0.012***</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0037)</td>
<td></td>
</tr>
<tr>
<td>Exp X K1 Payoff X High Opp Exp</td>
<td>0.072***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.022)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exp X K1 Payoff X Medium Opp Exp</td>
<td>0.075***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.012)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Own Games&gt;100</td>
<td>-0.002</td>
<td>0.002</td>
<td>0.031***</td>
</tr>
<tr>
<td></td>
<td>(0.0022)</td>
<td>(0.0017)</td>
<td>(0.0044)</td>
</tr>
<tr>
<td>Own Games&gt;100 X K1 Payoff</td>
<td>0.081***</td>
<td>0.066***</td>
<td>-0.026***</td>
</tr>
<tr>
<td></td>
<td>(0.0068)</td>
<td>(0.0059)</td>
<td>(0.0080)</td>
</tr>
<tr>
<td>Own Games&gt;100 X High Opp Exp</td>
<td>-0.017**</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0049)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Own Games&gt;100 X Medium Opp Exp</td>
<td>-0.010*</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0052)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Own Games&gt;100 X K1 Payoff X High Opp Exp</td>
<td>0.005</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.024)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Own Games&gt;100 X K1 Payoff X Medium Opp Exp</td>
<td>0.065***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.017)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>3921798</td>
<td>3921798</td>
<td>3921798</td>
</tr>
<tr>
<td>Player Fixed Effects</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Adjusted ( R^2 )</td>
<td>0.003</td>
<td>0.004</td>
<td>0.008</td>
</tr>
</tbody>
</table>

Notes: '\( k_1 \) Payoff' is the expected payoff to playing \( k_1 \) if the opponent randomizes according to his history. 'High opp exp' is a dummy for opponents with between 31 and 94 past games. 'Experienced' is a dummy for players who eventually play at \( \geq \) 100 games. 'Own Games >100' indicates the player has already played at least 100 games. The 'X' indicates the interaction between the dummies and other covariates. SEs are clustered by player.

\* \( p < .10, \quad ** p < .05, \quad *** p < .01 \)
length on $k_1$ consistency, the mean marginal effect across players is similar to the OLS coefficients. We only report the overall results because the individual analyses lack power — very few of the individual-level coefficients are statistically different from zero. While we expect the correlation between opponent’s history-length and playing $k_1$ to be negative – since longer histories are less likely to show substantial deviation from random – we do not have a good explanation for why the direct effect of opponent’s history length is negative, even when controlling for the $k_1$ payoff. Perhaps the players are more wary of trying to exploit a more experienced player.

### 3.5 Quantal Response

The above evidence that $k_1$-consistent play is more likely when the expected payoff is higher, naturally leads us to a model of play that is more continuous. In some sense level-$k$ strategies are all or nothing. If a throw has the highest expected payoff against the opponent’s historical distribution, then the $k_1$ strategy says to play it, even if expected payoff is very small. A related, but different strategy is for players to choose each throw with a probability that is increasing in its expected payoff against the opponent’s historical distribution of play. This is related to the idea behind Quantal Response Equilibrium (McKelvey and Palfrey, 1995), but without requiring that players be in equilibrium. Non-equilibrium quantal response has been used in a variety of economic contexts. See McFadden (1976) and cites therein.

In this context, players doing one iteration of reasoning would have probabilities of play

$$P_i \propto \exp(\alpha_i + \beta E[i \mid \text{opponent's distribution}]).$$

Their probability of playing a given throw is increasing in expected return to that throw, assuming the opponent plays according to his historical distribution. This smooths the threshold response of the $k_1$ strategy into a more continuous response.\(^{28}\)

---

\(^{28}\)A second level of reasoning would expect opponents to play according to the distribution induced by one’s
Figure 3.7: *Distribution across players of the coefficient in the quantal response model.*

Note: The distribution across 6670 players.

Figure 3.7 shows the distribution of coefficients across individuals. The mean coefficient is .01. The expected return is the percent chance of winning minus the percent chance of losing, so it ranges from -100 to 100. The standard deviation is 23.2, so, on average, a standard deviation increase in the expected return to an action, increases the percent chance it is played by

\[
23.2 \cdot .01 \cdot 100 \cdot \frac{2}{9} = 5.2 \text{ percentage points.}
\]

(We multiply by 100 to convert to percentages and by 2/9 to evaluate the margin at the means.) The standard deviation across players in the percent of the time they play a throw is 5%, so this is a significant, but not a huge effect.

However, given the low levels of \(k_2\) play we find and the econometric difficulties of including own history in the logit, we only analyze the first iteration of reasoning.

\[
29 \text{In the reduced form results (Table 3.5) we showed that players did not respond to the expected payoff calculated from \textit{predicted} opponent play, whereas this shows that players do respond to expected payoffs calculated from \textit{historical} opponent play.}
\]
The coefficient on expected return is significant for 60.7% of players. The mean of the effect size conditional on being significant is .025. Converting to margins, this corresponds to a standard deviation increase in expected return resulting in a 12.9 percentage point increase in the probability of playing a given throw, which is quite large.

### 3.6 Likelihood Comparison

Which is a better model of player behavior, the discrete “if it’s the $k_1$ throw, play it” or the more continuous “if its $k_1$ payoff is higher, play it with higher probability”? Since the strategies are similar, if players were using one there would still be some evidence for the other, so we use a likelihood test to see which model better fits players’ behavior. For each player we can calculate the likelihood of observing the set of throws he plays given the level-$k$ maximum likelihood model and given the quantal response model. To facilitate the comparison, we estimate a version of the maximum likelihood model with no $k_2$ so that each model has three independent parameters.

If we assume that one of two models, level-$k$ (LK) and QE, generated the data and have a flat prior over which it was, then the probability that it was QE is

$$P(\text{QE} \mid \text{data}) = \frac{P(\text{data} \mid \text{QE})}{P(\text{data} \mid \text{LK}) + P(\text{data} \mid \text{QE})}.$$  

Figure 3.8 plots the distribution of the probability of the quantal response model across players. The probability is less than .5 for 3,711 players, so for 55.6% of players the quantal response is a better model. More interestingly, there are substantial numbers of players both to the left of .05 and to the right of .95. For 1,228 players (18.4%) the MLE model is a statistically better fit and for 1,186 players (17.8%) the quantal response is a statistically better fit. This suggests some players’ strategies focus more on whether the throw has the highest expected return ($k_1$) and others’ strategies respond more to the level of the expected return (quantal response).
3.7 Conclusion

The 20th century witnessed several break through discoveries in economics. Arguably the most important revolved around understanding behavior in strategic settings, which originated with John von Neumann’s (1928) minimax theorem. In zero-sum games with unique mixed-strategy equilibria, minimax logic dictates that strategies should be randomized to prevent exploitation by one’s opponent. The work of Nash enhanced our understanding of optimal play in games, and several theorists since have made seminal discoveries.

We take this research in a different direction by analyzing an enormous set of naturally generated data on rock-paper-scissors with information about opponents’ past play. In doing so, we are able to explore the models — both equilibrium and non-equilibrium — that best describe the data. While we find that most people employ strategies consistent with Nash, at least some of the time, there is considerable deviation from equilibrium play. Adapting level-$k$ thinking to our repeated game context, we use maximum likelihood to estimate the
frequency with which each player uses $k_0$, $k_1$ and $k_2$. We find that about three-quarters of all 
throws are best described as $k_0$. A little less than one-fifth of play is $k_1$, with $k_2$ level play 
accounting for less than one-tenth of play. Interestingly, we find that most players are mixing 
over at least two levels of reasoning. Since players mix across levels, we explore when they 
are most likely to play $k_1$. We find that consistency with $k_1$ is increased when the expected 
return to $k_1$ is higher.

We also explore the quantal response model. Our adapted version of quantal response 
has players paying attention to the expected return to each strategy. We find that a one 
standard deviation increase in expected return increases the probability of a throw by 5.2 
percentage points. In addition, for about a fifth of players the quantal response model fits 
significantly better than the level-$k$ model, but for another one-fifth the level-$k$ model fits 
significantly better. It seems that some players focus on the levels of the expected returns, 
while others focus on which throw has the highest expected return.

Beyond theory testing, we draw several methodological lessons. First, while our setting is 
very different from the single-shot games that level-$k$ theory was originally developed, the 
evidence that players mix across strategies raises questions for experiments that attempt to 
categorize players as a $k$-type, based on only a few instances of play. Second, with large data 
sets subtle differences in theoretical predictions can be tested with meaningful power. As the 
internet continues to provide unique opportunities for such large-scale data, we hope that 
our study can serve as a starting point for future explorations of behavior in both strategic 
and non-strategic settings.


Facebook (2010). Developer partner case study: Red bull application.


Chapter A: Appendix to Chapter 1

A.1 Proofs Omitted from the Text

Existence of Stable Arrangements

In this section, we use results from the literature on matching with contracts to show the existence of stable arrangements in our framework. For a given transfer vector \( t \), the demand of manager \( m \in M \), denoted \( D^m(t) \), is

\[
D^m(t) \equiv \arg \max_{Y \subseteq W} \left\{ \alpha_Y^m - t^m \to Y \right\}.
\]

Definition (Kelso and Crawford (1982)). The preferences of manager \( m \in M \) are substitutable if for any transfer vectors \( t \) and \( \tilde{t} \) with \( \tilde{t} \geq t \), there exists, for each \( Y \in D^m(t) \), some \( \tilde{Y} \in D^m(\tilde{t}) \) such that

\[
\tilde{Y} \supseteq \{ w \in Y : t^m \to w = \tilde{t}^m \to w \}.
\]

That is, the preferences of \( m \in M \) are substitutable if an increase in the “prices” of some workers cannot decrease demand for the workers whose prices remain unchanged.\(^1\)

Theorem 2 of Kelso and Crawford (1982) shows that under the assumption that all managers’ preferences are substitutable, there is an arrangement \([\mu; t]\) that is strict core, in the sense that: \(^2\)

- Each agent (weakly) prefers his assigned match partner(s) (with the corresponding transfer(s)) to being unmatched, that is,

\[
w^i([\mu; t]) \geq 0 \quad \forall i \in M \cup W.
\]

- There does not exist a manager \( m \in M \), a set of workers \( Y \subseteq W \), and a transfer vector

\(^1\)Theorem A.1 of Hatfield et al. (2013) shows that in our setting the Kelso and Crawford (1982) substitutability condition is equivalent to the choice-based substitutability condition of Hatfield and Milgrom (2005), that we describe in the main text: the availability of new workers cannot make a manager want to hire a worker he would otherwise reject.

\(^2\)Strictly speaking, Kelso and Crawford (1982) have one technical assumption not present in our framework: they assume that \( \alpha_w^m + \gamma_w^m \geq 0 \), in order to ensure that all workers are matched. However, examining the Kelso and Crawford (1982) arguments reveals that this extra assumption is not necessary to ensure that a strict core arrangement exists – the Kelso and Crawford (1982) salary adjustment processes can be started at some arbitrarily low (negative) salary offer and all of the steps and results of Kelso and Crawford (1982) remain valid, with the caveat that some workers may be unmatched at core outcomes.
Such that
\[
\alpha_Y^m - \tilde{t}^m \rightarrow Y \geq \alpha_{\mu(m)}^m - t^m \rightarrow \mu(m),
\]
and
\[
\gamma_w^m + \xi(t^m \rightarrow w) \geq \gamma_{\mu(w)}^w + \xi(t^{\mu(w)} \rightarrow w) \quad \forall w \in Y,
\]
with strict inequality for at least one \( i \in (\{m\} \cup Y) \).

The Kelso and Crawford (1982) (p. 1487) construction of competitive equilibria from strict core allocations then implies that there is some transfer vector \( \hat{t} \), having \( \hat{t}^{\mu(w) \rightarrow w} = t^{\mu(w) \rightarrow w} \) (for each \( w \in W \)), such that \([\mu; \hat{t}]\) is stable in our sense.

**Proof of Lemma 1**

We let \( \mathcal{B} \) be the set of managers who are matched at \( \mu \) and let \( \mathcal{B} \) be the set of workers who are matched at \( \mu \). This means that
\[
\mu(m) \subseteq \mathcal{B} \quad \forall m \in \mathcal{B},
\]
\[
\mu(w) \in \mathcal{B} \quad \forall w \in W.
\]

These observations, combined with the fact that \( t^m \rightarrow m = t^w \rightarrow w = 0 \), enable us to show that
\[
\sum_{m \in M} t^m \rightarrow \mu(m) = \sum_{m \in \mathcal{B}} t^m \rightarrow \mu(m) + \sum_{m \in M \setminus \mathcal{B}} t^m \rightarrow \mu(m),
\]
\[
= \sum_{m \in \mathcal{B}} t^m \rightarrow \mu(m),
\]
\[
= \sum_{m \in \mathcal{B}} \sum_{w \in \mu(m)} t^m \rightarrow w,
\]
\[
= \sum_{w \in W} \sum_{m \in \mathcal{B}, \mu(m) = w} t^m \rightarrow w + \sum_{w \in W \setminus \mathcal{B}} t^{\mu(w)} \rightarrow w,
\]
\[
= \sum_{w \in W} t^{\mu(w)} \rightarrow w.
\]

**Proof of Proposition 1**

First, we show that the arrangements stable under full taxation (\( \tilde{\tau} = 1 \)) cannot Pareto dominate those stable under tax \( \hat{\tau} < 1 \).

**Claim.** Suppose that \([\bar{\mu}; \bar{t}]\) is stable under tax \( \hat{\tau} < 1 \), and that \([\tilde{\mu}; \tilde{t}]\) is stable under tax \( \tilde{\tau} = 1 \). Then, \([\bar{\mu}; \bar{t}]\) (under tax \( \hat{\tau} = 1 \)) cannot Pareto dominate \([\tilde{\mu}; \tilde{t}]\) (under tax \( \hat{\tau} < 1 \)).

**Proof.** As no transfers get through under full taxation, an arrangement stable under full taxation is most likely to Pareto dominate some other arrangement when all transfers between
match partners are 0. Thus, we assume that \( \tilde{t}^{\mu(w) \rightarrow w} = 0 \) for each \( w \in W \), and suppose that \([\tilde{\mu}; \tilde{t}]\) (under full taxation) Pareto dominates \([\hat{\mu}; \hat{t}]\) (under tax \( \hat{\tau} \)). This would imply that
\[
\alpha_{m}^{\hat{\mu}(m)} = \alpha_{m}^{\tilde{\mu}(m)} - \hat{t}^{m \rightarrow \hat{\mu}(m)} \geq \alpha_{m}^{\tilde{\mu}(m)} - \tilde{t}^{m \rightarrow \tilde{\mu}(m)}, \quad (A.1)
\]
\[
\gamma_{w}^{\hat{\mu}(w)} = \gamma_{w}^{\tilde{\mu}(w)} + \xi_{\hat{\tau}}^{\text{prop}}(\tilde{\mu}(w) \rightarrow w) \geq \gamma_{w}^{\tilde{\mu}(w)} + \xi_{\tilde{\tau}}^{\text{prop}}(\hat{\mu}(w) \rightarrow w), \quad (A.2)
\]
with strict inequality for some \( m \) or \( w \). However, stability of \([\hat{\mu}; \hat{t}]\) under tax \( \hat{\tau} \) implies that
\[
\alpha_{m}^{\hat{\mu}(m)} - \hat{t}^{m \rightarrow \hat{\mu}(m)} \geq \alpha_{m}^{\tilde{\mu}(m)} - \tilde{t}^{m \rightarrow \tilde{\mu}(m)}, \quad (A.3)
\]
\[
\gamma_{w}^{\hat{\mu}(w)} + \xi_{\hat{\tau}}^{\text{prop}}(\hat{\mu}(w) \rightarrow w) \geq \gamma_{w}^{\tilde{\mu}(w)} + \xi_{\tilde{\tau}}^{\text{prop}}(\tilde{\mu}(w) \rightarrow w). \quad (A.4)
\]
Combining (A.1) and (A.3) gives
\[
0 \geq -\hat{t}^{m \rightarrow \hat{\mu}(m)}, \quad (A.5)
\]
for each \( m \in M \), while combining (A.2) and (A.4) gives
\[
0 \geq \xi_{\hat{\tau}}^{\text{prop}}(\hat{\mu}(w) \rightarrow w), \quad (A.6)
\]
for each \( w \in W \). Strict inequality must hold in (A.5) or (A.6) for some \( m \) or \( w \).

In the first of these cases, we have
\[
\hat{t}^{m' \rightarrow \hat{\mu}(m')} > 0
\]
for some \( m' \in M \); hence, there exists at least one \( w \in \hat{\mu}(m') \) for whom
\[
\hat{t}^{\mu(w) \rightarrow w} > 0. \quad (A.7)
\]
But (A.7) contradicts (A.6).

In the second case, we have
\[
0 > \xi_{\hat{\tau}}^{\text{prop}}(\hat{\mu}(w') \rightarrow w'), \quad (A.8)
\]
for some \( w' \in W \). If we take \( m = \hat{\mu}(w') \), then (A.8) and (A.6) together imply that
\[
0 > \sum_{w \in \hat{\mu}(m)} \hat{t}^{\mu(w) \rightarrow w} = \hat{t}^{m \rightarrow \hat{\mu}(m)},
\]
contradicting (A.5).

For \( \tilde{\tau} < 1 \), \( \xi_{\tilde{\tau}}^{\text{prop}}(\cdot) \) is strictly increasing and the conclusion of the proposition follows from the following more general result.

**Proposition 1'.** Suppose that \( \tilde{\xi}(\cdot) \) is strictly increasing, that \([\tilde{\mu}; \tilde{t}]\) is stable under \( \tilde{\xi}(\cdot) \), and that \([\hat{\mu}; \hat{t}]\) is stable under \( \hat{\xi}(\cdot) \), with \( \tilde{\xi}(\cdot) \leq \hat{\xi}(\cdot) \). Then, \([\tilde{\mu}; \tilde{t}] \) (under \( \tilde{\xi}(\cdot) \)) cannot Pareto
dominate \([\hat{\mu}; \hat{t}]\) (under \(\hat{\xi}(\cdot)\)).\(^3\)

**Proof.** The case for \(\tilde{\tau}\) Pareto dominance of \([\tilde{\mu}; \tilde{t}]\) (under \(\tilde{\xi}(\cdot)\)) over \([\hat{\mu}; \hat{t}]\) (under \(\hat{\xi}(\cdot)\)) would imply that

\[
\alpha_{\tilde{m}}^{\tilde{\mu}(m)} - \tilde{t}^{m \rightarrow \tilde{\mu}(m)} \geq \alpha_{m}^{\hat{\mu}(m)} - \hat{t}^{m \rightarrow \hat{\mu}(m)},
\]

\[
\gamma_{\tilde{w}}^{\tilde{\mu}(w)} + \tilde{\xi}(\tilde{t}^{\tilde{\mu}(w) \rightarrow w}) \geq \gamma_{w}^{\hat{\mu}(w)} + \hat{\xi}(\hat{t}^{\hat{\mu}(w) \rightarrow w}),
\]

with strict inequality for some \(m\) or \(w\). However, stability of \([\hat{\mu}; \hat{t}]\) under \(\hat{\xi}(\cdot)\) implies that

\[
\alpha_{\tilde{m}}^{\tilde{\mu}(m)} - \tilde{t}^{m \rightarrow \tilde{\mu}(m)} \geq \alpha_{m}^{\tilde{\mu}(m)} - \tilde{t}^{m \rightarrow \tilde{\mu}(m)},
\]

\[
\gamma_{\tilde{w}}^{\tilde{\mu}(w)} + \tilde{\xi}(\tilde{t}^{\tilde{\mu}(w) \rightarrow w}) \geq \gamma_{w}^{\tilde{\mu}(w)} + \tilde{\xi}(\tilde{t}^{\tilde{\mu}(w) \rightarrow w}),
\]

where the second inequality in (A.12) follows from the fact that \(\hat{\xi}(\cdot) \geq \tilde{\xi}(\cdot)\).

Combining (A.9) and (A.11) gives

\[
\tilde{t}^{m \rightarrow \tilde{\mu}(m)} \geq \hat{t}^{m \rightarrow \hat{\mu}(m)},
\]

for each \(m \in M\), while combining (A.10) and (A.12) gives

\[
\tilde{\xi}(\tilde{t}^{\tilde{\mu}(w) \rightarrow w}) \geq \tilde{\xi}(\hat{t}^{\tilde{\mu}(w) \rightarrow w})
\]

\[
\tilde{\xi}(\tilde{\mu}(w) \rightarrow w) \geq \hat{\xi}(\tilde{t}^{\tilde{\mu}(w) \rightarrow w}) \geq \hat{\xi}(\hat{t}^{\hat{\mu}(w) \rightarrow w}),
\]

(A.14)

for each \(w \in W\), where the second line of (A.14) follows from the fact that \(\tilde{\xi}(\cdot)\) is strictly increasing. Strict inequality must hold in (A.13) or (A.14) for some \(m\) or \(w\).

In the first of these cases, we have

\[
\hat{t}^{m \rightarrow \tilde{\mu}(m')} > \hat{t}^{m \rightarrow \hat{\mu}(m')}
\]

for some \(m' \in M\); hence, there exists at least one \(w \in \hat{\mu}(m')\) for whom

\[
\hat{t}^{\tilde{\mu}(w) \rightarrow w} > \hat{t}^{\hat{\mu}(w) \rightarrow w}.
\]

(A.15)

But (A.15) contradicts (A.14).

In the second case, we have

\[
\tilde{\mu}(w') \rightarrow w' > \hat{\mu}(w') \rightarrow w'
\]

(A.16)

\(^3\)We say that an arrangement \([\tilde{\mu}; \tilde{t}]\) (under \(\tilde{\xi}(\cdot)\)) Pareto dominates arrangement \([\hat{\mu}; \hat{t}]\) (under \(\hat{\xi}(\cdot)\)) if

\[
\alpha_{\tilde{m}}^{\tilde{\mu}(m)} - \tilde{t}^{m \rightarrow \tilde{\mu}(m)} \geq \alpha_{m}^{\hat{\mu}(m)} - \hat{t}^{m \rightarrow \hat{\mu}(m)},
\]

\[
\gamma_{\tilde{w}}^{\tilde{\mu}(w)} + \tilde{\xi}(\tilde{t}^{\tilde{\mu}(w) \rightarrow w}) \geq \gamma_{w}^{\hat{\mu}(w)} + \hat{\xi}(\hat{t}^{\hat{\mu}(w) \rightarrow w})
\]

with strict inequality for some \(i \in M \cup W\).
for some \( w' \in W \). If we take \( m = \hat{\mu}(w') \), then (A.16) and (A.14) together imply that

\[
\sum_{w \in \tilde{\mu}(m)} \tilde{t}_m(w) \rightarrow w > \sum_{w \in \hat{\mu}(m)} \hat{t}_m(w) \rightarrow w;
\]

hence, we find that

\[
\tilde{t}^m \rightarrow \hat{\mu}(m) > \hat{t}^m \rightarrow \hat{\mu}(m),
\]

contradicting (A.13).

\[ \square \]

**Proof of Theorem 1**

If \( \hat{\mu} = \hat{\mu} \), then the theorem is trivially true. Thus, we consider a wage market in which \([\hat{\mu}; \hat{t}] \) is stable under tax \( \hat{\tau} \), \([\hat{\mu}; \hat{t}] \) is stable under tax \( \hat{\tau} \), \( \hat{\tau} > \hat{\tau} \), and \( \hat{\mu} \neq \hat{\mu} \).

The stability conditions for the managers imply that

\[
\alpha_{m}^\mu(m) - \tilde{t}^m \rightarrow \tilde{\mu}(m) \geq \alpha_{m}^\mu(m) - \tilde{t}^m \rightarrow \tilde{\mu}(m),
\]

(A.17)

\[
\alpha_{m}^\mu(m) - \hat{t}^m \rightarrow \hat{\mu}(m) \geq \alpha_{m}^\mu(m) - \hat{t}^m \rightarrow \hat{\mu}(m),
\]

(A.18)

these inequalities together imply that

\[
\sum_{m \in M} (\tilde{t}^m \rightarrow \tilde{\mu}(m) - \tilde{t}^m \rightarrow \tilde{\mu}(m)) \geq \sum_{m \in M} (\hat{t}^m \rightarrow \hat{\mu}(m) - \hat{t}^m \rightarrow \hat{\mu}(m)).
\]

(A.19)

As the market is a wage market, we have

\[
\xi_{\tau}^{\text{prop}}(\tilde{\mu}(w) \rightarrow w) = (1 - \hat{\tau})\tilde{t}(w) \rightarrow w \quad \text{and} \quad \xi_{\tau}^{\text{prop}}(\hat{\mu}(w) \rightarrow w) = (1 - \hat{\tau})\hat{t}(w) \rightarrow w,
\]

hence, the stability conditions for the workers imply that

\[
\gamma_{w}^\mu(w) + (1 - \hat{\tau})\tilde{t}(w) \rightarrow w \geq \gamma_{w}^\mu(w) + (1 - \hat{\tau})\hat{t}(w) \rightarrow w,
\]

(A.20)

\[
\gamma_{w}^\mu(w) + (1 - \hat{\tau})\tilde{t}(w) \rightarrow w \geq \gamma_{w}^\mu(w) + (1 - \hat{\tau})\hat{t}(w) \rightarrow w.
\]

(A.21)

Summing these inequalities and applying Lemma 1, we obtain

\[
(1 - \hat{\tau}) \sum_{m \in M} (\tilde{t}^m \rightarrow \tilde{\mu}(m) - \hat{t}^m \rightarrow \hat{\mu}(m)) \geq (1 - \hat{\tau}) \sum_{m \in M} (\tilde{t}^m \rightarrow \tilde{\mu}(m) - \hat{t}^m \rightarrow \hat{\mu}(m)).
\]

(A.22)

Combining (A.19) and (A.22), we find that

\[
(1 - \hat{\tau}) \sum_{m \in M} (\tilde{t}^m \rightarrow \tilde{\mu}(w) - \hat{t}^m \rightarrow \hat{\mu}(m)) \geq (1 - \hat{\tau}) \sum_{m \in M} (\tilde{t}^m \rightarrow \tilde{\mu}(w) - \hat{t}^m \rightarrow \hat{\mu}(m)).
\]

(A.23)
Since $\hat{\tau} < \tilde{\tau}$, (A.23) implies that
\[
\sum_{m \in M} \left( \hat{t}_m^{m \rightarrow \hat{\mu}(m)} - \tilde{t}_m^{m \rightarrow \tilde{\mu}(m)} \right) \geq 0.
\] (A.24)

Next, using (A.18) and (A.21), we find that
\[
\Xi(\hat{\mu}) - \Xi(\tilde{\mu}) = \sum_{m \in M} \left( \alpha_m^{\hat{\mu}(m)} - \alpha_m^{\tilde{\mu}(m)} \right) + \sum_{w \in W} \left( \gamma_w^{\hat{\mu}(w)} - \gamma_w^{\tilde{\mu}(w)} \right)
\geq \sum_{m \in M} \left( \hat{t}_m^{m \rightarrow \hat{\mu}(m)} - \tilde{t}_m^{m \rightarrow \tilde{\mu}(m)} \right) - (1 - \hat{\tau}) \sum_{w \in W} \left( \hat{t}_w^{\hat{\mu}(w) \rightarrow w} - \tilde{t}_w^{\tilde{\mu}(w) \rightarrow w} \right),
\]
\[
= \hat{\tau} \sum_{m \in M} \left( \hat{t}_m^{m \rightarrow \hat{\mu}(m)} - \tilde{t}_m^{m \rightarrow \tilde{\mu}(m)} \right) \geq 0,
\]
where the final inequality follows from (A.24).

Proof of Proposition 2 and Derivation of Equation (1.11)

Summing (A.20) across women, we find that
\[
\sum_{w \in W} \left( \gamma_w^{\tilde{\mu}(w)} - \gamma_w^{\hat{\mu}(w)} \right) \geq (1 - \tilde{\tau}) \sum_{w \in W} \left( \tilde{t}_w^{\tilde{\mu}(w) \rightarrow w} - \tilde{t}_w^{\hat{\mu}(w) \rightarrow w} \right)
\geq (1 - \tilde{\tau}) \sum_{w \in W} \left( \tilde{t}_w^{\hat{\mu}(w) \rightarrow w} - \hat{t}_w^{\hat{\mu}(w) \rightarrow w} \right),
\]
\[
\geq 0,
\]
where the inequality (A.26) follows from (A.19), and the inequality (A.27) follows from (A.24). Thus, we see Proposition 2 – the workers receive higher match utility under $\tilde{\mu}$ than under $\hat{\mu}$.

Furthermore, this implies that
\[
\sum_{m \in M} \left( \alpha_m^{\hat{\mu}(m)} - \alpha_m^{\tilde{\mu}(m)} \right) \geq 0,
\] (A.28)
so that we may calculate the lowest tax under which a given inefficient match $\tilde{\mu}$ can be stable. Combining (A.17) and (A.25), we find that
\[
\sum_{w \in W} \left( \gamma_w^{\tilde{\mu}(w)} - \gamma_w^{\hat{\mu}(w)} \right) \geq (1 - \tilde{\tau}) \sum_{m \in M} \left( \alpha_m^{\hat{\mu}(m)} - \alpha_m^{\tilde{\mu}(m)} \right).
\] (A.29)

The inequality in (A.28) allows us to rearrange (A.29) to obtain
\[
\frac{\sum_{w \in W} \left( \gamma_w^{\tilde{\mu}(w)} - \gamma_w^{\hat{\mu}(w)} \right)}{\sum_{m \in M} \left( \alpha_m^{\hat{\mu}(m)} - \alpha_m^{\tilde{\mu}(m)} \right)} \geq (1 - \tilde{\tau}),
\]
so that we find

$$
\tilde{\tau} \geq \frac{\sum_{m \in M} \left( \alpha_m \hat{\mu}(m) - \alpha_m \tilde{\mu}(m) \right)}{\sum_{m \in M} \left( \alpha_m \hat{\mu}(m) - \alpha_m \tilde{\mu}(m) \right)} + \frac{\sum_{w \in W} \left( \gamma_w \hat{\mu}(w) - \gamma_w \tilde{\mu}(w) \right)}{\sum_{m \in M} \left( \alpha_m \hat{\mu}(m) - \alpha_m \tilde{\mu}(m) \right)}
$$

$$
= \frac{\mathcal{M}(\hat{\mu}) - \mathcal{M}(\tilde{\mu})}{\sum_{m \in M} \left( \alpha_m \hat{\mu}(m) - \alpha_m \tilde{\mu}(m) \right)}. \quad (A.30)
$$

**Proof of Proposition 3**

Assume a matching $\hat{\mu}$ is stable under tax $\tilde{\tau}$. In a wage market, if we re-normalize the workers utilities by dividing by $(1 - \tilde{\tau})$, then a match that is stable with the renormalized utilities and no taxation is also stable with the original utilities and tax $\tilde{\tau}$

$$
\gamma_{w} \tilde{\mu}(w) + (1 - \tilde{\tau}) \tilde{t}_{m \to w} \geq \gamma_{w} \hat{\mu}(w) + (1 - \tilde{\tau}) \hat{t}_{m \to w},
$$

$$
\iff \frac{1}{1 - \tilde{\tau}} \gamma_{w} \tilde{\mu}(w) + \tilde{t}_{m \to w} \geq \frac{1}{1 - \tilde{\tau}} \gamma_{w} \hat{\mu}(w) + \hat{t}_{m \to w}. \quad (A.31)
$$

Combining (A.31) with the standard manager stability conditions,

$$
\alpha_m \hat{\mu}(m) - \hat{t}_{m \to \hat{\mu}(m)} \geq \alpha_m \hat{\mu}(m) - \hat{t}_{m \to \hat{\mu}(m)},
$$

$$
\alpha_m \tilde{\mu}(m) - \tilde{t}_{m \to \tilde{\mu}(m)} \leq \alpha_m \tilde{\mu}(m) - \tilde{t}_{m \to \tilde{\mu}(m)}, \quad (A.32)
$$

$$
\alpha_m \hat{\mu}(m) - \hat{t}_{m \to \hat{\mu}(m)} \leq \alpha_m \hat{\mu}(m) - \hat{t}_{m \to \hat{\mu}(m)}, \quad (A.33)
$$

so that

$$
\tilde{t}_{m \to \hat{\mu}(m)} - \hat{t}_{m \to \hat{\mu}(m)} \geq \hat{t}_{m \to \hat{\mu}(m)} - \hat{t}_{m \to \hat{\mu}(m)}. \quad (A.34)
$$

**Proofs of Proposition 4 and Corollary 1**

Suppose that in a wage market, both $[\hat{\mu}, \hat{\mu}]$ and $[\hat{\mu}, \tilde{\mu}]$ are stable under tax $\tau$. The stability conditions for the managers imply that

$$
\alpha_m \hat{\mu}(m) - \hat{t}_{m \to \hat{\mu}(m)} \geq \alpha_m \hat{\mu}(m) - \hat{t}_{m \to \hat{\mu}(m)},
$$

$$
\alpha_m \hat{\mu}(m) - \hat{t}_{m \to \hat{\mu}(m)} \leq \alpha_m \hat{\mu}(m) - \hat{t}_{m \to \hat{\mu}(m)}, \quad (A.32)
$$

so that

$$
\tilde{t}_{m \to \hat{\mu}(m)} - \hat{t}_{m \to \hat{\mu}(m)} \geq \hat{t}_{m \to \hat{\mu}(m)} - \hat{t}_{m \to \hat{\mu}(m)}. \quad (A.34)$$
Meanwhile, the stability conditions for the workers imply that
\begin{align}
\gamma_{\tilde{w}}(w) + (1 - \tau)\tilde{t}_{\tilde{w}}(w) - w &\geq \gamma_{\hat{w}}(w) + (1 - \tau)\hat{t}_{\hat{w}}(w), \\
\gamma_{\hat{w}}(w) + (1 - \tau)\hat{t}_{\hat{w}}(w) - w &\leq \gamma_{\tilde{w}}(w) + (1 - \tau)\tilde{t}_{\tilde{w}}(w),
\end{align}
(A.35) (A.36)
so that
\begin{align}
(1 - \tau)(\hat{\mu}(w) - \tilde{\mu}(w)) &\leq (1 - \tau)(\hat{\mu}(m) - \tilde{\mu}(m)).
\end{align}
(A.37)

Summing (A.34) and (A.37) across agents and using Lemma 1, we find that
\begin{align}
\sum_{m \in M} \left( \tilde{t}_{\tilde{m} \to \tilde{\mu}(m)} - \hat{t}_{\hat{m} \to \hat{\mu}(m)} \right) &= \sum_{m \in M} \left( \hat{t}_{\hat{m} \to \hat{\mu}(m)} - \tilde{t}_{\tilde{m} \to \tilde{\mu}(m)} \right),
\end{align}
(A.38)

If the managers are not indifferent in aggregate between $\tilde{\mu}$ and $\hat{\mu}$, so that
\begin{align}
\sum_{m \in M} \left( \alpha_{\hat{m}}(m) - \alpha_{\tilde{m}}(m) \right) &\neq 0,
\end{align}
(A.39)
we have,
\begin{align}
\tau &= 1 + \frac{\sum_{w \in W} \left( \gamma_{\tilde{w}}(w) - \gamma_{\hat{w}}(w) \right)}{\sum_{m \in M} \left( \alpha_{\tilde{m}}(m) - \alpha_{\hat{m}}(m) \right)}.
\end{align}
(A.40)

This shows Proposition 4.

To see Corollary 1, it suffices to observe that (A.40) pins down a unique tax rate in the case that (A.39) holds. Thus, if there are two tax rates under which matchings $\tilde{\mu}$ and $\hat{\mu}$ are both stable, then we must have
\begin{align}
\sum_{m \in M} \left( \alpha_{\hat{m}}(m) - \alpha_{\tilde{m}}(m) \right) &= 0.
\end{align}
(A.41)

But then, we also have
\begin{align}
\sum_{w \in W} \left( \gamma_{\tilde{w}}(w) - \gamma_{\hat{w}}(w) \right) &= 0,
\end{align}
(A.42)
by (A.38). Combining (A.41) and (A.42), we find that

\[ \mathcal{T} (\hat{\mu}) - \mathcal{T} (\tilde{\mu}) = \sum_{m \in M} (\alpha_m (m) - \alpha_m (m)) + \sum_{w \in W} (\gamma_w (w) - \gamma_w (w)) = 0, \]

as desired.

**Proof of Proposition 5**

Follows directly from the arguments presented the text.

**Proof of Lemma 2**

In a strictly positive wage market, all matches are accompanied by a strictly positive transfer; hence, a lump sum tax on transfers is equivalent to a flat fee for matching. Thus, Lemma 2 follows from the following slightly more general result.

Here and hereafter, we say that an arrangement or matching is stable under flat fee \( f \) if it is stable given transfer function \( \xi_f (\cdot) \).

**Lemma 2'. Reduction of a flat fee for matching (weakly) increases the number of workers matched in stable matchings.** That is, if matching \( \tilde{\mu} \) is stable under flat fee \( \tilde{f} \), matching \( \hat{\mu} \) is stable under flat fee \( \hat{f} \), and \( \hat{f} < \tilde{f} \), then

\[ \#(\hat{\mu}) \geq \#(\tilde{\mu}), \]

where \( \#(\mu) \) denotes the number of workers matched in matching \( \mu \).

**Proof.** As \([\tilde{\mu}; \tilde{t}]\) is stable under flat fee \( \tilde{f} \), we have

\[ \alpha_m (\tilde{m} \rightarrow \tilde{\mu}) \geq \alpha_m (\hat{m} \rightarrow \hat{\mu}) \]

\[ \gamma_w (w) + \tilde{t}_{\tilde{\mu}(w) \rightarrow w} - \tilde{f} \cdot \{1_{\tilde{\mu}(w) \neq w}\} \geq \gamma_w (w) + \tilde{t}_{\hat{\mu}(w) \rightarrow w} - \hat{f} \cdot \{1_{\hat{\mu}(w) \neq w}\} \]

where \( \{1_{\mu(w) \neq w}\} \) is an indicator function that equals 1 if \( w \) is matched in matching \( \mu \) and 0 if \( w \) is unmatched in matching \( \mu \). Summing these inequalities across agents, and using Lemma 1, we find that

\[ \sum_{m \in M} (\alpha_m (\tilde{m} \rightarrow \tilde{\mu}) - \alpha_m (\hat{m} \rightarrow \hat{\mu})) + \sum_{w \in W} (\gamma_w (\tilde{\mu}(w)) - \gamma_w (\hat{\mu}(w))) + \tilde{f} \cdot (\#(\tilde{\mu}) - \#(\hat{\mu})) \geq 0. \]

(A.43)

Similarly, as \([\hat{\mu}; \hat{t}]\) is stable under flat fee \( \hat{f} \),

\[ \alpha_m (\hat{m} \rightarrow \hat{\mu}) \geq \alpha_m (\tilde{m} \rightarrow \tilde{\mu}) \]

\[ \gamma_w (w) + \hat{t}_{\hat{\mu}(w) \rightarrow w} - \hat{f} \cdot \{1_{\hat{\mu}(w) \neq w}\} \geq \gamma_w (w) + \tilde{t}_{\tilde{\mu}(w) \rightarrow w} - \tilde{f} \cdot \{1_{\tilde{\mu}(w) \neq w}\}; \]
these inequalities yield
\[ \sum_{m \in M} (\alpha^\hat{\mu}_m(m) - \alpha^\tilde{\mu}_m(m)) + \sum_{w \in W} (\gamma^\hat{\mu}_w(w) - \gamma^\tilde{\mu}_w(w)) + \hat{f} \cdot (\#(\hat{\mu}) - \#(\tilde{\mu})) \geq 0. \] (A.44)

upon summation.

Adding (A.43) and (A.44) shows that
\[ (\hat{f} - \tilde{f})(\#(\hat{\mu}) - \#(\tilde{\mu})) \geq 0. \]

Thus, if \( \tilde{f} > \hat{f} \), we must have \( \#(\hat{\mu}) \geq \#(\tilde{\mu}) \); this proves the result.

\[ \square \]

**Proof of Theorem 2**

As in the proof of Lemma 2, Theorem 2 follows from the following slightly more general result.

**Theorem 2’.** A reduction in a flat fee for matching (weakly) increases the total match utility of stable matchings. That is, if \( \tilde{\mu} \) is stable under flat fee \( \tilde{f} \), \( \hat{\mu} \) is stable under flat fee \( \hat{f} \), and \( \hat{f} < \tilde{f} \), then
\[ \mathfrak{T}(\hat{\mu}) \geq \mathfrak{T}(\tilde{\mu}). \]

**Proof.** Using (A.44) and Lemma 2’, we find that
\[ \mathfrak{T}(\hat{\mu}) - \mathfrak{T}(\tilde{\mu}) = \sum_{m \in M} (\alpha^\hat{\mu}_m(m) - \alpha^\tilde{\mu}_m(m)) + \sum_{w \in W} (\gamma^\hat{\mu}_w(w) - \gamma^\tilde{\mu}_w(w)) \geq \hat{f} \cdot (\#(\hat{\mu}) - \#(\tilde{\mu})) \geq 0; \] (A.45)

this proves Theorem 2’.

\[ \square \]

**Proof of Proposition 6**

As in the proof of Lemma 2, Proposition 6 follows from the following slightly more general result.

**Proposition 6’.** Let \( \hat{\mu} \) be an efficient matching, and let \( \tilde{\mu} \) be stable under flat fee \( \tilde{f} \). Then,
\[ 0 \leq \mathfrak{T}(\hat{\mu}) - \mathfrak{T}(\tilde{\mu}) \leq \tilde{f} \cdot (\#(\hat{\mu}) - \#(\tilde{\mu})). \]

**Proof.** This is immediate from (A.43).

\[ \square \]

**Proof of Proposition 7**

As in the proof of Lemma 2, Proposition 7 follows from the following slightly more general result.
Proposition 7'. A matching $\tilde{\mu}$ can be stable under a flat fee only if

$$\tilde{\mu} \in \arg \max_{\mu : \#(\mu) \leq \#(\tilde{\mu})} \{T(\mu)\}.$$  

Proof. From (A.43), we see that if $[\tilde{\mu}; \tilde{t}]$ is stable under flat fee $\tilde{f}$, then for any matching $\hat{\mu} \neq \tilde{\mu}$,

$$T(\tilde{\mu}) - T(\hat{\mu}) + \tilde{f} \cdot (\#(\tilde{\mu}) - \#(\hat{\mu})) \geq 0. \quad (A.46)$$

If fewer workers are matched in $\tilde{\mu}$ than in $\hat{\mu}$ (i.e. $\#(\tilde{\mu}) \geq \#(\hat{\mu})$), (A.46) implies that

$$T(\tilde{\mu}) - T(\hat{\mu}) \geq \tilde{f} \cdot (\#(\tilde{\mu}) - \#(\hat{\mu})) \geq 0,$$

so that $\tilde{\mu}$ must have higher total match utility than $\hat{\mu}$. \hfill $\Box$

A.2 Simulation Details

For each market, we draw fixed and idiosyncratic components of match utilities; the combination of these terms into match utilities depends on the simulation. We use the Woodcock (2008) estimates of the variance of worker fixed effects, manager fixed effects, and match-specific effects. Woodcock (2008) finds mean wages of $41107. Since the labor share of GDP is about 2/3, we multiply $41107 by 3/2 to estimate productivities. As we think workers have some surplus from working, we multiplied $41107 by 2/3 to get worker match utilities.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Interpretation</th>
<th>Mean</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x_m$</td>
<td>Manager FE of productivity</td>
<td>0</td>
<td>.198</td>
</tr>
<tr>
<td>$x_w$</td>
<td>Worker FE of productivity</td>
<td>0</td>
<td>.102</td>
</tr>
<tr>
<td>$y_m$</td>
<td>Manager FE of worker (dis)utility</td>
<td>0</td>
<td>.198</td>
</tr>
<tr>
<td>$y_w$</td>
<td>Worker FE of worker (dis)utility</td>
<td>0</td>
<td>.102</td>
</tr>
<tr>
<td>$x_{mw}$</td>
<td>Idiosyncratic component of productivity</td>
<td>0</td>
<td>.079</td>
</tr>
<tr>
<td>$y_{mw}$</td>
<td>Idiosyncratic component of worker (dis)utility</td>
<td>0</td>
<td>.079</td>
</tr>
<tr>
<td>$\mu_1$</td>
<td>Mean productivity</td>
<td>$\frac{3}{2} \cdot 41107$</td>
<td>$-$</td>
</tr>
<tr>
<td>$\mu_2$</td>
<td>Mean disutility of work</td>
<td>$-\frac{3}{2} \cdot 41107$</td>
<td>$-$</td>
</tr>
</tbody>
</table>

For the baseline specification, we use:

$$\alpha_m^w = \mu_1 \exp(x_m + x_w + x_{mw}),$$
$$\gamma_m^w = \mu_2 \exp(y_m + y_w + y_{mw}).$$

To analyze the impact of raising mean utility in the market (Figures 1.12 and 1.13 and the discussion thereof), we compare the baseline simulation markets to two markets, one with

$$\alpha_m^w = 10000 + \mu_1 \exp(x_m + x_w + x_{mw}),$$
$$\gamma_m^w = \mu_2 \exp(y_m + y_w + y_{mw}),$$

126
and the other with
\[
\begin{align*}
\alpha_m^w &= \mu_1 \exp(x_m + x_w + x_{mw}), \\
\gamma_m^w &= \min\{10000 + \mu_2 \exp(y_m + y_w + y_{mw}), 0\}.
\end{align*}
\]

To analyze the impact of changes in the variance of match utilities, (Figure 1.14 and the ensuing discussion), we used
\[
\begin{align*}
\alpha_m^w &= \mu_1 \exp(\theta \cdot (x_m + x_w + x_{mw})), \\
\gamma_m^w &= \mu_2 \exp(\theta \cdot (y_m + y_w + y_{mw})),
\end{align*}
\]
evaluated at different values of the variance multiplier \(\theta \in \{\frac{1}{4}, 1, 4\}\).

To analyze the impact of correlation between productivity and work disutility for a given match pair (Figure 1.16 and the related discussion), we used
\[
\begin{align*}
\alpha_m^w &= \mu_1 \exp(x_m + x_w + x_{mw}), \\
\gamma_m^w &= \mu_2 \exp(\eta(x_m + x_w + x_{mw}) + (1 - \eta^2)^{\frac{1}{2}}(y_m + y_w + y_{mw})),
\end{align*}
\]
evaluated at different values of the correlation parameter \(\eta \in \{-0.8, 0, 0.8\}\).

In order to keep the total variance constant while varying the weights on the idiosyncratic component and the fixed effect component (Figure 1.15 and related discussion), we drew new values for fixed effects and idiosyncratic match utilities, each with variance 1. We used match utilities
\[
\begin{align*}
\alpha_m^w &= \mu_1 \exp((.198 + .102 + .079)^{\frac{1}{2}}(\beta \cdot x_w + (1 - \beta^2)^{\frac{1}{2}} \cdot x_{mw})), \\
\gamma_m^w &= \mu_2 \exp((.198 + .102 + .079)^{\frac{1}{2}}(\beta \cdot y_m + (1 - \beta^2)^{\frac{1}{2}} y_{mw})),
\end{align*}
\]
evaluated at different values of the fixed-effect weight, \(\beta \in \{0, .5, 1\}\). In each case, the overall variance of the corresponding normal distribution was \(0.198 + 0.102 + 0.079 = 0.379\)—the same as in the other simulations. These utilities with \(\beta = 0\) are used in Figure 1.6 and the corresponding discussion of the Feldstein (1999) estimate of deadweight loss.
B.1 Deriving GePP

Proof of Proposition 1. Writing $Q_i$ for $Q_i(P)$ for conciseness, the firm’s first order conditions are

$$Q_i + \left( \frac{\partial Q_i}{\partial P_i} + \frac{\partial Q_i}{\partial P_i}^T \frac{\partial P_i}{\partial P_i} \right) (P_i - mc_i(Q_i)) = 0.$$  

Remembering that $\frac{dQ_i}{dP_i} = \frac{\partial Q_i}{\partial P_i}^T + \left( \frac{\partial Q_i}{\partial P_i} \right)^T \frac{\partial P_i}{\partial P_i}$, and then multiplying by $-\left( \frac{dQ_i}{dP_i}^T \right)^{-1}$ the firm’s first-order conditions can be rewritten as:

$$0 = -\left( \frac{dQ_i}{dP_i}^T \right)^{-1} Q_i - (P_i - mc_i(Q_i)) \equiv f_i(P).$$

After a merger of firms $i$ and $j$, the newly formed firm takes into account the effect of $P_i$ on $\pi_j$ and no longer expects $P_j$ to react to $P_i$ since the two are chosen jointly. The merged firm’s first-order derivatives with respect to $P_i$ can be written:

$$h_i(P) = - (P_i - mc_i(Q_i)) \left( \frac{dQ_i}{dP_i}^T - \frac{\partial Q_j}{\partial P_j} \frac{\partial P_j}{\partial P_i} \right)^{-1} Q_i \left( \frac{dQ_j}{dP_i} - \frac{\partial Q_j}{\partial P_j} \frac{\partial P_j}{\partial P_i} \right)^T (P_j - mc_j(Q_j)),$$

where the last term equals $\tilde{D}_{ij}(P_j - mc_j(Q_j))$. Using the definition $g_i(P) = h_i(P) - f_i(P)$, we have:

$$g_i(P) = \tilde{D}_{ij}(P_j - mc_j(Q_j)) - \left( \left( \frac{dQ_i}{dP_i}^T - \frac{\partial Q_j}{\partial P_j} \frac{\partial P_j}{\partial P_i} \right)^{-1} (P_j - mc_j(Q_j)) - \left( \frac{dQ_i}{dP_i}^T \right)^{-1} Q_i. $$

\[
\square
\]

B.2 Cournot

Example 2 Continued:. Note that post-merger-to-monopoly, the firm is just a multiproduct monopoly which we can think of as choosing prices. Beginning from symmetry, in the limit as the products become undifferentiated, a unit tax on each of the two goods will then increase prices according to the pass-through rate facing the monopolist. Yet note that $\tilde{D}$, the diversion ratio, is always 1 since the goods are undifferentiated and thus any sales lost by
one on the margin are picked up by the other as we saw above. Let the actual pass-through rate for this common cost shock (which must be the same across products as within, given that products are homogeneous) post-merger be \( \rho \); then by equation (2.9), which is valid so long as \( \left( \frac{\partial h}{\partial P} \right)^{-1} \) is strictly negative definite

\[
0 \neq \rho = - \left[ \left( \frac{\partial h}{\partial P} \right)^{-1} \right]_{11} + \left[ \left( \frac{\partial h}{\partial P} \right)^{-1} \right]_{21} = 0,
\]

a contradiction. Thus it must be that \( \left( \frac{\partial h}{\partial P} \right)^{-1} \) is not in fact strictly negative definite at symmetric prices and thus pass-through must locally be infinite. Thus, in the undifferentiated limit, the 0 value of GePP is misleading: the net effect any common cost shock will have on incentives will apparently be 0, but because pass-through is so large, this cancels out. Away from the limit in the symmetric case, exact calculations of merger pass-through are a simplification of the formula in Appendix B.5.

### B.3 Conjectural variations examples

Often, the two merging firms are closer competitors (and potential accommodators) with each other than with other firms in the industry. Therefore, we now consider a three firm model, with the two merging firms being symmetric but the third-firm being asymmetric, representing a reduced form for the rest of the industry. To keep things simple, though, we assume that the quantity of all firms \( q \) and all firms’ (Bertrand) demand slopes are the same, but now we have two diversion ratios: \( d \), the (Bertrand) diversion to and from the third firm from and to each of the two merger partners and \( \delta \), the diversion from each merger partner to the other. The mark-ups of the two merger partners are \( m \). We assume that conjectures are in proportion to diversion: each merger partner anticipates an accommodating reaction of \( \lambda \delta \) from its partner and \( \lambda d \) from the third firm, while the third firm expects \( \lambda d \) from each of the merger partners.

**Proposition 10.** In the three-firm example, GePP from a merger of the two close firms is

\[
m \frac{\delta + \tilde{\lambda} (d^2 - \delta^2) - (d^2 + \delta^2) \frac{\delta^2}{1-\lambda}}{1-d^2\tilde{\lambda}} \approx m \frac{\delta + \tilde{\lambda} (d^2 - \delta^2)}{1-d^2\tilde{\lambda}},
\]

where \( \tilde{\lambda} \equiv \frac{\lambda}{1+\delta} \) and again the approximation is valid for small \( \lambda \). Approximate GePP is thus increasing (decreasing) in \( \lambda \) if and only if \( d \) is greater (less) than \( \frac{\delta}{\sqrt{1+\delta}} \). Approximate GePP is constant in \( \lambda \) if and only if \( d = \frac{\delta}{\sqrt{1+\delta}} \). Precise GePP decreases in strictly more cases than does approximate GePP.

If the strength of the within-merger interaction is small compared to that outside the merger, GePP increases with anticipated accommodation. Conversely, if the strength of within-merger interaction is sufficiently greater than the total outside interaction then accommodation decreases GePP. Some relevant cases may be close to the point where the degree of accommodation anticipated has little effect. It is reassuring for the theory of
Proof. Our proof here is almost entirely analogous to that of Example 5. The first-order condition now requires that for the merging firms

\[
m = -\frac{\partial Q_1}{\partial P_1} \left( 1 - [d^2 + \delta^2] \lambda \right), \quad \Rightarrow \frac{\partial Q_1}{\partial P_1} = -\frac{q}{m \left( 1 - [d^2 + \delta^2] \lambda \right)}.
\]

On the other hand by the logic of conjectures discussed in the proof of Example 5, if \( l \) represents the pre-merger merging-firm-to-non-merging-firm conjecture, \( L \) represents the same between the merging firms, and \( \tilde{l} \) represents the post-merger version of \( l \) then

\[
l = \tilde{l} (1 + L) \iff \tilde{l} = \frac{l}{1 + L}.
\]

Plugging in our definitions of \( l = d\lambda \) and \( L = \delta \lambda \) we obtain \( \tilde{l} = \frac{d\lambda}{1+\delta\lambda} \equiv \delta \tilde{\lambda} \). Now we can compute

\[
\frac{d^M Q_1}{dP_1} = \frac{\partial Q_1}{\partial P_1} \left( 1 - \tilde{l}d \right) = -\frac{q \left( 1 - d^2 \tilde{\lambda} \right)}{m \left( 1 - [d^2 + \delta^2] \lambda \right)}, \quad \Rightarrow \quad mD_{12} = \frac{\delta + d^2 \tilde{\lambda}}{1 - d^2 \lambda}, \quad (B.2)
\]

and

\[
Q_1 \left( \frac{1}{\frac{d^2 Q_1}{dP_1}} - \frac{1}{\frac{dQ_1}{dP_1}} \right) = m \left( -\frac{1 - [d^2 + \delta^2] \lambda}{1 - d^2 \tilde{\lambda}} + 1 \right) = -m \frac{d^2 \tilde{\lambda} - (d^2 + \delta^2) \lambda}{1 - d^2 \lambda}.
\]

With a little algebra, subtracting (B.3) from (B.2) yields the formula in the text, given that \( \tilde{\lambda} - \lambda = \frac{\delta^2}{1 - \delta \lambda} \). As before, the more sophisticated formula is decreasing in \( \lambda \) whenever the simpler version is, but also decreases in some cases (for larger \( \lambda \)) when the simpler version does not.

Returning to the simpler formula and taking the derivative with respect to \( \lambda \) yields an expression proportional to \( d^2 (1 + \delta) - \delta^2 \), which is clearly positive or negative depending on the sign of the inequality in the proposition.

\[
\]

130
B.4 Taylor Series Error Term

For notational convenience let $x = h^{-1}$. The error term is

$$
\frac{1}{2} \left( \sum_i \sum_j \frac{\partial^2 x_i}{\partial h_i \partial h_j} g_i(P^0) g_j(P^0) \right) = \frac{1}{2} \left[ \sum_i \left( \begin{array}{c} \frac{\partial^2 x_1}{\partial h_1 \partial h_1} \\ \vdots \\ \frac{\partial^2 x_n}{\partial h_1 \partial h_n} \end{array} \right) \right] g_i(P^0) g_i(P^0) = \frac{1}{2} \left[ \sum_i \left( \frac{\partial^2 x}{\partial h_i \partial h_i} \right) \right] g_i(P^0) \equiv \frac{1}{2} \left[ \sum_i \left( \frac{\partial^2 x}{\partial h_i \partial h_i} \right) \right] g_i(P^0).
$$

We know $\frac{\partial x}{\partial h} = I$. Differentiating with respect to $h_i$ gives

$$
\frac{\partial^2 x}{\partial h_i \partial h} \frac{\partial x}{\partial h} + \frac{\partial x}{\partial h} \left( \begin{array}{c} \sum_k \frac{\partial^2 h_1}{\partial x_k \partial x_k} \frac{\partial x}{\partial h_i} \\ \vdots \\ \sum_k \frac{\partial^2 h_n}{\partial x_k \partial x_k} \frac{\partial x}{\partial h_i} \end{array} \right) = 0.
$$

Solving for $\frac{\partial^2 x}{\partial h_i \partial h}$, using $\frac{\partial x}{\partial h} = (\frac{\partial h}{\partial x})^{-1}$ and substituting into (B.4) gives

$$
E = -\frac{1}{2} \sum_i \frac{\partial x}{\partial h} \left( \begin{array}{c} \sum_k \frac{\partial^2 h_1}{\partial x_k \partial x_k} \frac{\partial x}{\partial h_i} \\ \vdots \\ \sum_k \frac{\partial^2 h_n}{\partial x_k \partial x_k} \frac{\partial x}{\partial h_i} \end{array} \right) \frac{\partial x}{\partial h} g_i(P^0) g_i(P^0).
$$
If we look at just the $a$th entry of the vector, we have

$$E_a = -\frac{1}{2} \sum_i \sum_j \frac{\partial x_a}{\partial h_j} \left( \sum_k \frac{\partial^2 h_j}{\partial x_k \partial x_k} \frac{\partial x_k}{\partial h_i} \right) \frac{\partial x}{\partial h} g_i (P^0) g (P^0)$$

$$= -\frac{1}{2} \sum_i \sum_j \sum_k \sum_l \frac{\partial x_a}{\partial h_j} \left( \sum_m \sum_n \frac{\partial x_a}{\partial h_j} \frac{\partial^2 h_j}{\partial x_k \partial x_k} \frac{\partial x_k}{\partial h_i} \right) \frac{\partial x}{\partial h} g_i (P^0) g_m (P^0)$$

$$= -\frac{1}{2} \sum_i \sum_j \sum_k \sum_l \frac{\partial x_a}{\partial h_j} \left( \sum_m \sum_n \frac{\partial x_a}{\partial h_j} \frac{\partial^2 h_j}{\partial x_k \partial x_k} \frac{\partial x_k}{\partial h_i} \right) D^2 h_j \frac{\partial x}{\partial h_m} g_m (P^0)$$

$$= -\frac{1}{2} \sum_i \sum_j \frac{\partial x_a}{\partial h_j} g^T (P^0) \left( \frac{\partial x}{\partial h} \right)^T D^2 h_j \frac{\partial x}{\partial h} g (P^0),$$

where the Hessian and derivatives are evaluated at some price in $[P^0, P^M]$. Letting $[A]_{ij}$ indicate the $ij$ element of matrix $A$,

$$E_a = -\frac{1}{2} \sum_j \left[ \left( \frac{\partial h}{\partial x} \right)^{-1} \right]_{aj} g^T (P^0) \left( \left( \frac{\partial h}{\partial x} \right)^T \right)^{-1} \left( D^2 h_j \right) \left( \frac{\partial h}{\partial x} \right)^{-1} g (P^0),$$

where $D^2 h_j$ denotes the Hessian. Letting $[A]_{ij}$ indicate the $ij$ element of matrix $A$,

$$-\frac{\partial f(P)}{\partial P} = \begin{pmatrix} 2 - \frac{Q_i \partial^2 Q_i}{\pi_i^2} \left( \frac{\partial Q_i}{\partial P} \right)^2 & \frac{\partial Q_i}{\partial P} \frac{\partial^2 Q_i}{\partial P \partial P} - Q_i \frac{\partial^2 Q_i}{\partial P \partial P} \left( \frac{\partial Q_i}{\partial P} \right)^2 \\ \frac{\partial Q_j}{\partial P} \frac{\partial^2 Q_j}{\partial P \partial P} - Q_j \frac{\partial^2 Q_j}{\partial P \partial P} \left( \frac{\partial Q_j}{\partial P} \right)^2 & 2 - \frac{Q_j \partial^2 Q_j}{\pi_j^2} \left( \frac{\partial Q_j}{\partial P} \right)^2 \end{pmatrix} = \rho^{-1} \equiv - \begin{pmatrix} m_1 & m_2 \\ m_3 & m_4 \end{pmatrix}.$$
Also,

\[
\frac{\partial g(P)}{\partial P} = \begin{pmatrix}
-\frac{\partial^2 Q_i}{\partial P_i^2} \frac{\partial Q_i}{\partial P_i} \frac{\partial Q_j}{\partial P_j} + \frac{\partial^2 Q_i}{\partial P_i^2} \frac{\partial^2 Q_j}{\partial P_i \partial P_j} \frac{\partial Q_j}{\partial P_j} - \frac{\partial^2 Q_i}{\partial P_i^2} \frac{\partial^2 Q_j}{\partial P_i \partial P_j} \frac{\partial Q_j}{\partial P_j} \\
-\frac{\partial Q_i}{\partial P_j} - (P_i - C_i) \frac{\partial^2 Q_i}{\partial P_i^2} \frac{\partial Q_i}{\partial P_i} \frac{\partial Q_j}{\partial P_j} - (P_i - C_i) \frac{\partial^2 Q_i}{\partial P_i^2} \frac{\partial^2 Q_j}{\partial P_i \partial P_j} \frac{\partial Q_j}{\partial P_j} - (P_i - C_i) \frac{\partial^2 Q_i}{\partial P_i^2} \frac{\partial^2 Q_j}{\partial P_i \partial P_j} \frac{\partial Q_j}{\partial P_j} \\
\end{pmatrix}
\]

so, using Slutsky symmetry of \( \frac{\partial^2 Q_i}{\partial P_i^2} = \frac{\partial^2 Q_i}{\partial P_i \partial P_j} \) and \( \frac{\partial^2 Q_j}{\partial P_j^2} = \frac{\partial^2 Q_j}{\partial P_i \partial P_j} \), we have

\[
\frac{\partial g(P)}{\partial P} = \begin{pmatrix}
v_i \\
v_j
\end{pmatrix},
\]

where

\[
v_i = \begin{pmatrix}
-(P_j - C_j) \frac{\partial Q_i}{\partial P_i} \frac{\partial Q_i}{\partial P_i} - m_2 (\frac{\partial Q_i}{\partial P_i})^2 \frac{\partial Q_i}{\partial P_i} - (\frac{\partial Q_i}{\partial P_i})^2 (2 - m_1) \frac{\partial Q_j}{\partial P_j} \\
-\frac{\partial Q_i}{\partial P_j} - (P_i - C_i) \frac{\partial Q_i}{\partial P_i} \frac{\partial Q_j}{\partial P_j} - m_2 (\frac{\partial Q_i}{\partial P_i})^2 \frac{\partial Q_j}{\partial P_j} - (\frac{\partial Q_i}{\partial P_i})^2 (2 - m_1) \frac{\partial Q_j}{\partial P_j}
\end{pmatrix},
\]

\[
v_j = \begin{pmatrix}
-(P_j - C_j) \frac{\partial Q_i}{\partial P_i} \frac{\partial Q_i}{\partial P_i} - m_2 (\frac{\partial Q_i}{\partial P_i})^2 \frac{\partial Q_i}{\partial P_i} - (\frac{\partial Q_i}{\partial P_i})^2 (2 - m_1) \frac{\partial Q_j}{\partial P_j} \\
-\frac{\partial Q_i}{\partial P_j} - (P_i - C_i) \frac{\partial Q_i}{\partial P_i} \frac{\partial Q_j}{\partial P_j} - m_2 (\frac{\partial Q_i}{\partial P_i})^2 \frac{\partial Q_j}{\partial P_j} - (\frac{\partial Q_i}{\partial P_i})^2 (2 - m_1) \frac{\partial Q_j}{\partial P_j}
\end{pmatrix},
\]
Chapter C: Appendix to Chapter 3

We formalize the process of playing rock-papers-scissors over Facebook as a sequence of best-of-three matches described by the game $\Gamma$ nested inside a larger game $\hat{\Gamma}$, which includes the matching process that pairs the players. We do not specify the matching process, as it turns out that it does not matter and the following holds for any matching process.\(^1\) Players may exit the game (and exit may not be random) after any subgame, but not in the middle of one. All players have symmetric payoffs and discount factor $\delta$ across subgames. Our results are similar to Wooders and Shachat (2001) for two-outcome games.

Each nested game $\Gamma$ is a “best-of-three” match of rock-paper-scissors played in rounds, which we will call “throws.” For each throw, both players simultaneously choose actions from $A = \{r, p, s\}$ and the outcome for each player is a win, loss, or tie; $r$ beats $s$, $s$ beats $p$, and $p$ beats $r$. A player wins $\Gamma$ by winning two throws. The winner of $\Gamma$ receives a pay-off of 1 and the loser gets -1. Note that $\Gamma$ is zero-sum. Therefore, at any stage of $\hat{\Gamma}$ the sum across players of all future discounted payoffs is zero.

Each match consists of at least two throws. Because of the possibility of ties, there is no limit on the length of a match. Let

$$K^l = \underbrace{A^2 \times A^2 \ldots \times A^2}_{l \text{ times}}$$

be the set of all possible sequences of $l$ throws by two players. Let $K^l \subset K^l$ be the set of possible complete matches of length $l$: sequences of throw pairs such that no player had 2 wins after $l - 1$ throws, but a player had two wins after the $l^{th}$ throw. Let $K = \cup K^l$ be the set of possible complete matches of any length. Let $\hat{K}^l \subset K^l$ be the set of possible incomplete matches of length $l$: sequences of throw pairs such that no player has 2 wins.

A player’s overall history after playing $t$ matches is the sequence of match histories for all matches he has played,

$$h_t^i \in H^t = \underbrace{K \times K \ldots \times K}_{t \text{ times}}.$$  

Players may not observe their opponents’ exact histories. Instead a player observes some public summary information of his opponent’s history. Let $f : H^t \to S^t \forall t$ be the function that maps histories into summary information. Denote by $s^t$ an element of $S^t$.

A strategy for a player having played $t$ complete matches, facing an opponent with $m$ complete matches after $l$ throws of the current match is a mapping from the player’s history, the information the player has about his opponent’s history, and partial history of the current match to a distribution of actions $\sigma_{t,m,l} : H^t \times S^r \times \hat{K}^l \to \Delta A$.

\(^1\)In the actual game players could challenge a specific player to a game or be matched by the software to someone else who was looking for an opponent.
It is helpful to define a function \( \#\text{win}_i : \hat{K}^t \cup K^t \to \{0, 1, 2\} \) \( \forall t \), which denotes the number of wins for player \( i \) after a match history. Similarly \( \#\text{win}_j \) is the number of wins for player \( j \) and \( \#\text{win} = \max\{\#\text{win}_i, \#\text{win}_j\} \) is the number of wins of the player with the most wins.

In Nash Equilibrium, at any throw of any match, the distribution chosen must maximize the expected payoff. The payoff consists of the flow payoff plus the continuation value from future matches if the match ends on this throw, or the expected payoff from continuing the match with the updated match history if the match does not end on this throw. So,

\[
\sigma_{t,m,l}(\hat{k}^t_{ij}, h^t_i, s^l_j) \in \arg\max_{\sigma_t} \mathbb{E}_{\sigma_t,\sigma_j}[u(\hat{k}^t_{ij}, a_i, a_j)],
\]

where

\[
u_i((\hat{k}^t_{ij}, (a_i, a_j)), h^t_i, s^l_j) = \begin{cases} 1 + \delta \cdot \eta((h^t_i, (\hat{k}^t_{ij}, (a_i, a_j)))) & \#\text{win}_i((\hat{k}^t_{ij}, (a_i, a_j))) = 2, \\ -1 + \delta \cdot \eta((h^t_i, (\hat{k}^t_{ij}, (a_i, a_j)))) & \#\text{win}_j((\hat{k}^t_{ij}, (a_i, a_j))) = 2, \\ v_i((\hat{k}^t_{ij}, (a_i, a_j)), h^t_i, s^l_j) & \#\text{win}((\hat{k}^t_{ij}, (a_i, a_j))) < 2. \end{cases}
\]

In the first two cases there is the immediate payoff from the match ending plus the inter-match continuation value, \( \eta(h^{t+1}_i) \) (the value of going back into the pool and playing future matches with the updated player history). In the last case there is just the intra-match continuation value, \( v(\hat{k}^{t+1}, h^t_i, s^l_j) \) (the value of continuing this match with the updated match history), which includes the later inter-match continuation value from when the match eventually ends. Note that we have implicitly set the intra-match discount factor to 1.\(^2\)

Because each match is symmetric and zero sum, the inter-match continuation values are unimportant.

**Lemma 3.** Under Nash equilibrium play, \( \eta(h^t_i) = 0 \ \forall h_i^t \in H^t \ \forall t. \)

**Proof.** Suppose that \( \eta(h^t_i) \neq 0 \). If \( \eta(h^t_i) < 0 \), then a player with that history could deviate to always playing \((\frac{1}{3}, \frac{1}{3}, \frac{1}{3})\). The player would win half their matches and lose half their matches and have continuation value 0. If \( \eta(h^t_i) > 0 \), then since \( \hat{\Gamma} \) is zero sum at every stage, there must exist \( h^*_j \) such that \( \eta(h^*_j) < 0 \), which by the same logic cannot happen in equilibrium. \( \square \)

Since inter-match continuation values are all zero, the intra-match continuation value is just the probability of eventually winning the match minus the probability of eventually losing the match. This means that continuation values are zero-sum: \( v_i(\hat{k}_{ij}, h_i, s_j) = -v_j(\hat{k}_{ij}, h_j, s_i) \).

The symmetry of the match also implies that, regardless of history, whenever players are tied for the number of wins in the match, the continuation value going forward is 0 for both players.

**Lemma 4.** In Nash equilibrium if \( \#\text{win}_i(\hat{k}^t_{ij}) = \#\text{win}_j(\hat{k}^t_{ij}) \), then \( v_i(\hat{k}^t_{ij}, h_i, s_j) = v_j(\hat{k}^t_{ij}, h_j, s_i) = 0. \)

---

\(^2\)Because what matters for the result is the symmetry across strategies at all stages, having an intra-match discount factor does not change the result, but substantially complicates the proof.
Proof. Assume \( \# \text{win}_i(\hat{k}^l_{ij}) = \# \text{win}_j(\hat{k}^l_{ij}) \). Suppose \( v_i(\hat{k}_{ij}, h_i, s_j) \neq 0 \). If \( v_i(\hat{k}_{ij}, h_i, s_j) < 0 \), player \( i \) has a profitable deviation to play \( \left( \frac{1}{3}, \frac{1}{3}, \frac{1}{3} \right) \) for the remainder of the match and they will win with probability one-half, giving them a continuation value of zero. Similarly for \( j \).

We therefore have \( v_i(\hat{k}_{ij}, h_i, s_j) \geq 0 \) and \( v_j(\hat{k}^l_{ij}, h_j, s_i) \geq 0 \) for all match histories.

If \( v_i(\hat{k}_{ij}, h_i, s_j) > 0 \) then since the match is zero-sum, there must exist an \( \hat{h}_i, \hat{h}_j \) such that \( v_i(\hat{k}_{ij}, \hat{h}_i, f(\hat{h}_j)) < 0 \) or \( v_j(\hat{k}_{ij}, \hat{h}_j, f(\hat{h}_i)) < 0 \), which contradicts the above result. Similarly for \( v_j > 0 \).

Lemma 5. In Nash equilibrium, if \( \# \text{win}_i(\hat{k}^l_{ij}) = 1 \) and \( \# \text{win}_j(\hat{k}^l_{ij}) = 0 \) then \( v_i(\hat{k}^l_{ij}, h_i, s_j) = \frac{1}{2} \) and \( v_j(\hat{k}^l_{ij}, h_j, s_i) = -\frac{1}{2} \).

Proof. Assume \( \# \text{win}_i(\hat{k}^l_{ij}) = 1 \) and \( \# \text{win}_j(\hat{k}^l_{ij}) = 0 \). Suppose \( v_i(\hat{k}_{ij}, h_i, s_j) < \frac{1}{2} \). This implies that player \( i \) has a profitable deviation to play \( \left( \frac{1}{3}, \frac{1}{3}, \frac{1}{3} \right) \) for the remainder of the match. If they do so they will win the match with probability

\[
\frac{1}{3} \cdot 1 + \frac{1}{2} + \frac{1}{3} \left( \frac{1}{3} + \frac{1}{3} \cdot \frac{1}{2} + \frac{1}{3} (\cdots) \right) = \frac{1}{3} \cdot \frac{3}{2} \left( 1 + \frac{1}{3} \cdots \right) = \frac{3}{4}.
\]

and lose with probability \( \frac{1}{4} \), giving them an intra-match continuation value of \( \frac{1}{2} \). Therefore \( v_i(\hat{k}^l_{ij}, h_i, s_j) \geq \frac{1}{2} \). Similar logic guarantees that \( v_j(\hat{k}^l_{ij}, h_j, s_i) \geq -\frac{1}{2} \).

If \( v_i(\hat{k}_{ij}, h_i, s_j) > \frac{1}{2} \) then since the match is zero-sum, there must exist an \( \hat{h}_i, \hat{h}_j \) such that \( v_i(\hat{k}_{ij}, \hat{h}_i, f(\hat{h}_j)) < \frac{1}{2} \) or \( v_j(\hat{k}_{ij}, \hat{h}_j, f(\hat{h}_i)) < -\frac{1}{2} \), which contradicts the above result. Similarly for \( v_j(\hat{k}^l_{ij}, h_j, s_i) > -\frac{1}{2} \).

Since the number of wins each player has after any history that has not ended the match is either 0 or 1, Lemmas 4 and 5 encompass all the possible \( \hat{k}^l_{ij} \). We therefore see that the continuation value depends only on the number of wins each player has

\[
v_i(\hat{k}^l_{ij}, h_i, s_j) = \bar{v}_i(\# \text{win}_i(\hat{k}^l_{ij}), \# \text{win}_j(\hat{k}^l_{ij})).
\]

So, for every \( \hat{k}^l_{ij} \) we can calculate \( u_i((\hat{k}_{ij}, a_i, a_j)) \) for each own throw and opponent throw. Figure C.1 gives these payoffs for each possible stage in a match. For each of these stages, it can only be a Nash Equilibrium of each player expects their opponent to play \( \left( \frac{1}{3}, \frac{1}{3}, \frac{1}{3} \right) \).
<table>
<thead>
<tr>
<th>(a) Start of Match</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A\B</strong></td>
</tr>
<tr>
<td>----------</td>
</tr>
<tr>
<td><strong>R</strong></td>
</tr>
<tr>
<td><strong>P</strong></td>
</tr>
<tr>
<td><strong>S</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>(b) A has 1 win, B has zero</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A\B</strong></td>
</tr>
<tr>
<td>----------</td>
</tr>
<tr>
<td><strong>R</strong></td>
</tr>
<tr>
<td><strong>P</strong></td>
</tr>
<tr>
<td><strong>S</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>(c) B has 1 win, A has zero</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A\B</strong></td>
</tr>
<tr>
<td>----------</td>
</tr>
<tr>
<td><strong>R</strong></td>
</tr>
<tr>
<td><strong>P</strong></td>
</tr>
<tr>
<td><strong>S</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>(d) A has 1 win, B has 1 win</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A\B</strong></td>
</tr>
<tr>
<td>----------</td>
</tr>
<tr>
<td><strong>R</strong></td>
</tr>
<tr>
<td><strong>P</strong></td>
</tr>
<tr>
<td><strong>S</strong></td>
</tr>
</tbody>
</table>

**Figure C.1:** Total payoffs (flow + continuation) for each stage of the match.