Title
Essays in Financial Economics

Author
Zheng, Geoffery Xin

Publication Date
2020

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Three Essays in Financial Economics

A dissertation submitted in partial satisfaction

of the requirements for the degree

Doctor of Philosophy in Management

by

Geoffery Zheng

2020

ABSTRACT OF THE DISSERTATION

Three Essays in Financial Economics

by

Geoffery Zheng

Doctor of Philosophy in Management

University of California, Los Angeles, 2020

Professor Stavros Panageas, Chair

In the first chapter, I use a simple decomposition to distinguish between (a) inequality driven by wealth accumulation by old money, and (b) that driven by the entry of new money into the top of the wealth distribution. I make use of administrative real estate holdings data and hand-collected genealogies to create a novel panel dataset of wealth spanning 1982 through 2018. I find that (i) the contribution of new money is large, and (ii) this contribution increases with horizon. Over 80 percent of the increase in wealth inequality is attributable to new money households. Many theories of inequality explain the rapid increase in wealth inequality; relatively few can explain the relative importance of new money. I present a parsimonious model featuring market incompleteness and innovative young firms that is able to match both these moments while also generating additional asset pricing predictions. In the model, inequality increases not because the rich get richer, but because of the emergence of fast-growing young firms.

In the second chapter (with Sebastian Gryglewicz and Barney Hartman-Glaser), I address a puzzling stylized fact of executive compensation: Firms with better investment opportunities tend to have lower levels of managerial incentives. Managerial incentives are measured as pay-performance-sensitivity, the change in the manager's compensation for a 1% increase in firm value. While it seems odd that increased opportunities do not coincide with stronger managerial incentives, I write down a theoretical model in which this pattern naturally emerges. The result hinges on two important ingredients: the manager exerts costly, unob-

servable effort to improve the firm; and the manager is more risk averse than the investor. In the model, the firm's manager can exert effort to improve firm value, but this effort is costly to the manager. Thus, there are direct costs of effort, which come from incentivizing the manager to work hard, as well as indirect costs of effort, which come from the manager's risk aversion. Our model contributes to the literature on executive compensation by rationalizing a puzzling stylized fact using a parsimonious contracting model. This article has been published in *Management Science*.

In the third chapter (with Bruno Pellegrino), I ask the question: How much can a country grow its economy via better allocation of labor and capital? A robust prediction of neoclassical economics is that, in a market economy, labor and capital should flow to its best users. However, in the data, some firms seem to be using labor and physical capital very productively, while others squander these resources. The contribution of this chapter is to ask how much of this variation in productivity can be attributed to differences in managerial expectations. Using a novel dataset in which managers of over 8,000 European firms report on the constraints faced by their firms, I construct a dataset in which managerial expectations and accounting variables are jointly observed. I consider four different types of constraints: bureaucratic regulation, nepotism, limited access to financial capital, and labor market frictions in hiring workers. Building on prior theoretical work in Industrial Organization and Macro-economics, I develop a model that allows me to estimate the difference in productivity between constrained firms and their unconstrained counterparts using firm-level accounting and survey data. I find managers' expectations of financial constraints act like a 26% tax on physical capital in Spain, and that nepotism in Hungary and Spain acts like a 2% tax on firm profits. Taken together, these suggest that better allocation of resources could grow these economies by several percentage points. The amelioration of these frictions represent low hanging fruit for policy makers interested in promoting economic growth.

The dissertation of Geoffery Zheng is approved.

Andrew Granger Atkeson

Andrea Lynn Eisfeldt

Barney P. Hartman-Glaser

Stavros Panageas, Committee Chair

University of California, Los Angeles

2020

To my loving parents, Qinfen and Yan, and my amazing sister Melanie, for their support and encouragement from day one.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I am deeply grateful to my advisor, Stavros Panageas, for his continual guidance and support. He served as an unwavering beacon throughout the tumult and uncertainty of research. Through our innumerable conversations, I have learned what it means to be an economist, a researcher, and a scientist.

I am also indebted to Barney Hartman-Glaser, for his role as equal parts mentor, co-author, and friend. I am also very thankful to Andrea Eisfeldt, who always encouraged me to think harder about my research questions, and Andy Atkeson, who invariably understood my work better than I myself did.

My greatest joys during the Ph.D. program were the stimulating conversations I was able to have with faculty and fellow doctoral students. In no particular order, I would like to thank Tony Bernardo, Ivo Welch, Lars Lochstoer, Tyler Muir, Valentin Haddad, and Mark Garmaise, and all the faculty at the Anderson School of Management. I would also like to thank my fellow Ph.D. students, both those in the finance area as well as several from other areas: Bruno Pellegrino, Marco Testoni, and Dan Yavorsky. Together, they created an environment in which I always felt welcome to ask questions and offer up ideas.

Marvin Lieberman and Craig Jessen in the doctoral office were indispensable in helping me navigate the doctoral program. I would also like to thank the Anderson School of Management and the Fink Center, for their generous financial assistance.

VITA

2008–2012	B.A. Economics and Physics, University of Chicago
2009–2013	B.S. Mathematics, University of Chicago
2013–2014	Investment Analyst, Brookstone Partners
2014–2020	Ph.D. Candidate in Management (Finance), UCLA Anderson School of Management
2014–2018	Fink Center Scholar
2014–2018	Anderson Fellow
2015–2020	Teaching Assistant, UCLA Anderson School of Management. Taught Computational Finance; Empirical Methods in Finance; Financial Innovation, Foundations of Finance; Option Markets; Probability, Statistics, and Computational Methods for Econometrics; Stochastic Calculus; Theory and Application of Regression Analysis
2018	Instructor, UCLA Anderson School of Management. Taught Introduction to Python
2019-2020	Max L. Heyman Jr. Fellow

CHAPTER 1

Wealth Shares in the Long Run

1.1 Introduction

Over the last few decades, there has been a secular rise in top wealth shares. How did a small fraction of households accumulate so much wealth; not just in absolute terms, but relative to the rest of the economy? One explanation is that wealth begets wealth and the rise in wealth inequality is driven by high returns on wealth (Piketty and Goldhammer, 2014; Hubmer, Krusell, and Smith, 2016), resulting in increasing inequality and declining social mobility. Yet there is significant churn in the ranks of the ultra-wealthy. Few households manage to stay on the Forbes list of the 400 richest Americans over long periods of time; only 20 percent of the 1982 Forbes 400 list have family who appear on the list in 2018 (Gomez, 2018; Benhabib, Bisin, and Luo, 2015). Each year, ten percent of Forbes 400 members fall off the list and are replaced by newly wealthy households. In this paper, I study the relative contribution of old and new money to the rise in U.S. wealth inequality.

To isolate these distinct contributions, I present a decomposition of changes in wealth inequality that differentiates between intensive contribution of incumbent wealthy households, whose wealth growth reflects returns on incumbent wealth, and the extensive contribution of new entrants who enter top wealth percentiles by displacing other wealthy households. top wealth percentile absent the effects of entry. The decomposition allows me to quantify the relative importance of these contributions in explaining the secular increase in wealth inequality. When I apply my decomposition to a novel panel of wealthy households, I find that incumbent wealth grows at a rate close to that of aggregate household wealth, meaning that the contribution of the intensive margin is small. Consequently, the displacement term is responsible for over 80 percent of the increase in wealth inequality since 1986.

Understanding the drivers of rising wealth inequality is of clear policy importance. Inequality of realized outcomes can arise from inequality of opportunities, and one role of public policy is to promote equal opportunities and mobility. Many politicians have labeled the rise of wealth inequality as the unjust product of an unfair system and want to directly address wealth inequality through redistributive policies. On the other hand, many of today's Forbes 400 members got there by founding disruptive new firms. Wealth inequality can be celebrated as a sign of a dynamic economy that rewards innovation and entrepreneurship or vilified as a symptom of rent-seeking behavior. My paper helps to distinguish between these competing interpretations by quantifying how much of the rise in top wealth shares is the result of new entrants.

In addition to their implications for wealth inequality, the individual wealth dynamics underlying the increase in wealth inequality are an important quantity in many economic models. In any micro-founded model of consumption and investment behavior, it is agents' beliefs about their consumption and wealth dynamics that drive their decisions. As we typically lack data on agents' wealth, models have been evaluated using aggregate data. Empirical data on household wealth dynamics offer new ways to test existing asset pricing models. Wealthy agents hold a significant fraction of assets and are likely candidates for marginal investors in markets (Malloy, Moskowitz, and Vissing-Jørgensen, 2009). The wealth dynamics of these households are closely linked to the stochastic discount factor and impact asset prices. At a deeper level, differences in growth rates across households inform us about imperfect risk sharing and market frictions. In the presence of such market imperfections, heterogeneity among agents matters for asset prices (Constantinides and Duffie, 1996).

To understand the implications of heterogeneity for wealth inequality and asset prices, I develop an overlapping generations model in which borrowing constraints inhibit some agents' ability to borrow against their future dividend income. Thus, the distribution of wealth matters for the interest rate. I further show that high incumbent returns and wealthy new entrants have starkly different implications for the interest rate, despite both affecting wealth inequality. When wealth inequality increases due to wealthier new entrants, the interest rate falls. This is consistent with the secular decline in interest rates over the recent decades which occurred alongside the rise in wealth inequality.

In the model, measuring the wealth growth of incumbents using repeated cross-sections leads to erroneous conclusions about asset prices. This is because, even in the top percentiles of the wealth distribution, some agents are not marginal in financial markets. Instead, I propose measuring wealth growth of a fixed population of incumbents. I show that this measurement of cohort wealth growth recovers returns in the economy, even in the presence of borrowing constraints.

In order to measure the returns on incumbent wealth, I construct a panel data set of wealth for ultra-wealthy households. Starting from the time a household first appears on the Forbes 400 list, I track its wealth across multiple hand-collected data sources through to present day. A key challenge in estimating wealth dynamics from existing data sources is that the wealth of the formerly wealthy is unobserved. Without observing the left-tail of wealth dynamics, it is difficult to draw conclusions about expected growth rates. The value of my panel is that it follows each household, irrespective of their present-day wealth. My panel tracks wealth for individuals like Bill Gates and Jeff Bezos, who have stayed on the Forbes 400 list over many years; it also the first panel that tracks wealth estimates for individuals like Richard Adams, who appeared on the list from 1997 to 2000 and peaked at rank 174 with a net worth of $1.4 billion, only to fall off the list when the tech bubble burst. I provide the first estimates of growth rates of incumbent wealth in the United States at long horizons.

Using my panel data set, I am able to estimate growth rates of wealth by fixing a set of households and calculating their realized wealth growth at long horizons. I estimate growth rates for entry cohorts, comprised of households that entered the Forbes 400 population at the same time. Aggregating at the level of entry cohorts allows me to better estimate average growth rates by averaging over multiple households. At the same time, it allows me to test for differences in average growth rates between cohorts by comparing contemporaneous growth rates between entry cohorts. I also estimate growth rates for incumbent cohorts, comprised of households that appeared on the Forbes 400 list at the same time. Each incumbent cohort consists of multiple entry cohorts, and incumbent growth rates are the wealth-weighted average of entry cohort growth rates. Incumbent growth rates are the appropriate growth rate for studying changes in wealth shares because they measure the wealth growth of a fixed population of incumbent households, those who were members of the top wealth share at a given point in time. Increases in the top wealth share above and beyond the wealth growth of incumbent top wealth households must be the result of entrants.

The extensive margin accounts for roughly half of the increase in wealth inequality since 2006 and over 80 percent of the increase since 1986. At long horizons, wealthy households have grown at an annual rate of 6.3 percent compared to a growth rate of 5.7 percent for aggregate wealth. Thus, there is evidence in the data to support the view that incumbent wealth self-perpetuates. However, if incumbents were the only driver of increasing wealth inequality, the wealth share of the Forbes 400 would have increased from 0.9 percent of total wealth in 1986 to 1.1 percent in 2018. In reality, the wealth share of the Forbes 400 increased from 0.9 to 2.8 percent over that period.

The increasing role of displacement at longer horizons is the result of heterogeneous growth rates within the incumbent population, as groups with lower growth rates shrink relative to faster growing groups. These effects are difficult to observe in short time-samples. By tracking households that appeared on early Forbes 400 lists, I am able to identify a population of incumbent top wealth households and estimate their wealth growth over long horizons. For the earliest Forbes households, I observe over thirty-five years of wealth estimates. The long period covered by my data set allows me to observe both cross-sectional and time-series heterogeneity in growth rates.

I find that the observed heterogeneity in growth rates is well explained by a life cycle model of wealthy households. Newer entrants to the Forbes 400 list grow at a faster rate than older cohorts of entrants; households also grow at slower rates as they age. Over time, differences in growth rates lead to changes in the composition of wealthy households, so that the instantaneous growth rate overestimates the long-run growth rate of the incumbent wealth share. Ignoring the role of heterogeneous growth rates results in estimates of displacement that are biased downwards and roughly half as large in size.

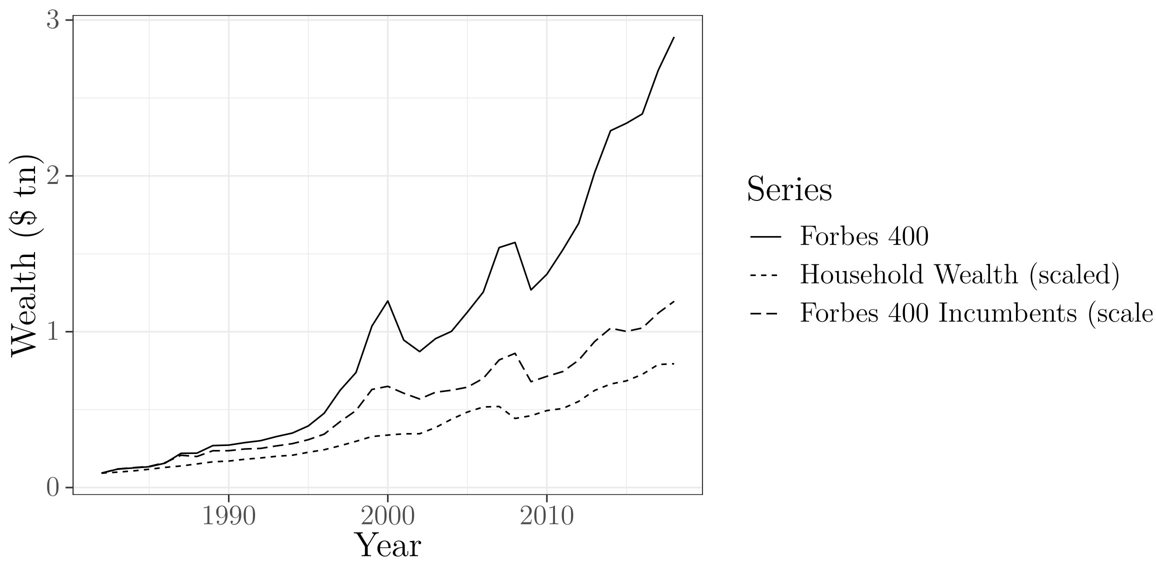


Figure 1.1: Comparison of Wealth Growth, 1982 - 2018. The total nominal wealth held by all members of the Forbes 400 is plotted in solid black. Aggregate household wealth, scaled to match the total Forbes wealth in 1982, is plotted in short dashes. The total nominal wealth held by incumbent Forbes 400 members, scaled to match the total Forbes wealth in 1982, is plotted in long dashes.

I find that the secular rise in top wealth shares is primarily the result of displacement. As shown in Figure 1.1, the wealth held by members of the Forbes 400 has increased from roughly $100 billion in 1982 to $3 trillion in 2018. This has significantly outpaced the growth in aggregate household wealth over the same period. However, it has also significantly outpaced the growth in wealth held by those initial Forbes members over the same period. The 1982 Forbes 400 members held $1 trillion in 2018, while the other $2 trillion of wealth is the result of displacement.

In the final section of my paper, I examine the implications of my findings for macro-finance models. These models typically have strong predictions for wealth distributions and wealth dynamics but have not been evaluated on their ability to match the data. My estimated wealth dynamics offer a new and important set of empirical moments for model selection. I start with the standard representative agent model and discuss the model's difficulties in reconciling wealth dynamics and asset prices. I show that several extensions to the model are still unable to jointly match my findings. The large role of displacement is strong evidence in support of models of disruptive growth and incomplete markets. The presence of heterogeneous growth rates and life cycle effects in wealth dynamics are consistent with portfolios featuring concentrated ownership in risky firms. Concentrated firm ownership also relates the life cycle effects I observe in wealth growth rates to life cycle effects in firm growth rates. My findings suggest that the underlying drivers of wealth inequality are the same as those underlying other macroeconomic phenomena such as the rise of superstar firms and the fall of the labor share.

1.1.1 Related Literature

My paper contributes to the growing literature on the rise in top wealth shares (Piketty and Goldhammer, 2014; Kuhn, Schularick, and Steins, 2017; Garbinti, Goupille-Lebret, and Piketty, 2017). This literature focuses on the overall increase in wealth shares, while my paper emphasizes the individual wealth dynamics that underlie the increase in wealth share. The motivation for focusing on the underlying dynamics is to distinguish between the role of incumbent growth rates and that of displacement. Concerns regarding the self-perpetuation of large fortunes are directly related to the relative magnitude of incumbent growth rates, which are distinct from the growth of top wealth shares.

Several papers have addressed the role of idiosyncratic wealth shocks in top wealthy households. My paper further extends the literature on the rise of top wealth shares by accounting for persistent differences in growth rates across households. Using detailed Swedish administrative data, Fagereng, Guiso, et al. (2016) document the importance of idiosyncratic risk for explaining dispersion in wealth growth of top wealth brackets. They find that heterogeneous returns can explain most of the time-variation in Swedish top wealth shares from 2000 to 2007. Consistent with their finding, I find the relative role of incumbent growth in driving changes in wealth inequality is larger at short horizons.

My paper also relates to the theoretical literature characterizing wealth inequality given an underlying stochastic process for wealth. My paper quantifies a qualitative insight of Gabaix et al. (2016), which is that the rapid increase in wealth inequality cannot be explained solely by changing growth rates of wealth. Less than one fifth of the increase in the Forbes wealth share is the result of high growth rates of wealth. An additional insight of their paper is that incorporating high-growth types that rapidly climb the ranks of the wealth distribution can generate fast transition dynamics. These high-growth individuals are analogous to the new entrants that I measure in my data set.

The paper closest to mine is Gomez (2018), which decomposes the rise in wealth inequality into within and displacement terms. Our theoretical frameworks differ in that he assumes a homogeneous growth rate of wealth among top wealth households, whereas I allow for persistent heterogeneity in growth rates. In Section 1.2.2, I elaborate on the differences in our methodologies. Conceptually, the differences stem from the fact that I follow a fixed population of households over time in calculating the growth rate of incumbent wealth. Thus, my growth rates can be interpreted as the growth rate of wealth for an incumbent wealthy household over time, rather than chained one-year growth rates of current Forbes 400 members. His paper also differs from mine in the statistical method used to impute wealth of missing households. I do this by creating a panel of wealth for Forbes 400 households and measuring realized growth rates in the panel. Gomez (2018) uses a Kaplan-Meier estimator to infer unobserved growth rates based on the distribution of observed returns in the Forbes 400. I show that, in the presence of heterogeneous growth rates, estimates of growth rates from repeated cross sections are biased estimates of individual household growth rates.

An empirical contribution of my paper is the construction of a panel data set of wealth for Forbes 400 households. I do this by merging observations from several existing data sets of wealth estimates. I impute missing observations using real estate ownership data from the LexisNexis public records data set. The data set has been used in the finance literature to investigate questions related to corporate leverage (Cronqvist, Makhija, and Yonker, 2012) and CEO succession (Yonker, 2017). Another paper that uses real estate value as a proxy for household wealth is Koudijs and Salisbury (2016). In a similar spirit, Civale, Diez-Catalan, and Salgado (2017) uses equity holdings as a proxy for household wealth within the Forbes 400.

My measured wealth dynamics are the realized value of an underlying portfolio. Previous papers including Calvet, Bach, and Sodini (2015), Fagereng, Guiso, et al. (2016), and

Fagereng, Holm, et al. (2019) have characterized wealth dynamics in European countries, whereas my focus is on American households and American wealth inequality. While these papers rely on administrative data, I construct a panel data set to estimate these dynamics for wealthy American households in the absence of analogous data. Earlier work on wealth inequality in the United States has used repeated cross sectional data sets such as the Survey of Consumer Finances (Benhabib, Bisin, and Luo, 2015) and estate tax filings (Kopczuk and Saez, 2004). Rather than estimate wealth dynamics from repeated cross-sections, I construct a panel to directly estimate growth rates of wealth, thereby avoiding the need for structural assumptions relating the cross-sectional wealth distribution to the underlying data-generating process. My results are the first estimates of long run wealth dynamics for wealthy American households. My work on rising wealth inequality is complementary to studies of rising income inequality in the United States (Guvenen, Ozkan, and Song, 2014; Song et al., 2018).

My paper contributes to the asset pricing literature that relates the wealth distribution to observed asset prices. Papers that discuss the effect of heterogeneity on asset prices include Gârleanu and Panageas (2015) and Gomez et al. (2016). In those papers, ex-ante heterogeneity drives changes in the wealth distribution as well as changes in risk premia due to time variation in risk-bearing capacity following strings of good and bad shocks. These models predict that wealthy individuals invest more aggressively and grow faster than aggregate wealth, resulting in increasing wealth inequality. This is at odds with my empirical finding that wealth inequality has increased significantly while wealthy households have outpaced aggregate wealth only modestly.

I find that wealth dynamics of wealthy households feature heterogeneous growth rates and idiosyncratic shocks. These features parallel those present in random growth models of firms, which have been used to explain the size distribution of firms (Luttmer, 2007). Furthermore, the large role of displacement is consistent with an increasingly skewed distribution of new firms (Gârleanu, Kogan, and Panageas, 2012; Gârleanu and Panageas, 2017) and an increase in idiosyncratic volatility (Herskovic et al., 2016; Hartman-Glaser, Lustig, and Xiaolan, 2017). At the aggregate level, it is also closely tied to the rise of superstar firms (Autor et al., 2017). Some papers that analyze the impact of concentrated ownership of firms on asset prices include Haddad (2012) and Di Tella (2019). Peter (2019) studies the role of firm dynamics and financing frictions in explaining cross-country differences in wealth inequality. Campbell, Ramadorai, and Ranish (2018) studies the equity portfolios of Indian households and find evidence of heterogeneous returns arising from concentrated equity positions.

My findings on family wealth dynamics are complementary to the literature on family firm dynamics (Bennedsen et al., 2007; Bertrand and Schoar, 2006). A majority of individuals in the Forbes 400 are associated with a family firm, and a number of papers study the impact of family ownership on firm outcomes (Anderson and Reeb, 2003). Pérez-González (2006), Villalonga and Amit (2006), and Morck, Stangeland, and Yeung (2000) show that lower performance of family firms arises in part due to within-family transition of managerial roles. My finding that incumbent wealthy households have quite ordinary growth rates of wealth echoes these results on firm management in a adjacent economic setting.

My work also contributes to the literature on inter-generational mobility (Clark and Cummins, 2013; Barone and Mocetti, 2016). In the long run, economic mobility is affected by both wealth dynamics within an individual's lifetime, as well as inter-generational transfers. I find that incumbent households' wealth share has increased over time, meaning that "old-money" has self-perpetuated over the past thirty years. However, I also find evidence that older cohorts of wealthy families under-perform newer cohorts. Overall, the self-perpetuation of wealth is not the driver of the sharp increase in wealth inequality over the past 30 years.

The rest of the paper is organized as follows: In Section 1.2, I present a model of wealth inequality and asset prices. In Section 1.3, I outline the data sources and methodology used to construct my panel data set, and then apply my framework to decompose the rise in the top wealth share. In Section 1.4, I discuss how my findings present challenges for standard macro-finance models and propose extensions.

1.2 Theory

To clarify concepts and motivate the measurement of cohort growth rates, I now layout an economy in which the long run wealth growth of agents, rather than the wealth growth of a percentile of the wealth distribution, determines the interest rate in the economy. I show that wealthy inequality driven by displacement leads to lower rates of return and higher asset prices, whereas superior incumbent growth rates leads to higher rates of return.

The key ingredients in the model are a life-cycle profile of firm dynamics and borrowing constraints. The borrowing constraint prevents entering agents with high expected wealth growth from fully borrowing against their future income, so that these agents are not marginal in determining interest rates.

The model features no aggregate risk and there is only a single traded asset, a riskless bond. As I show, even in this framework, the source of wealth inequality matters for asset prices. Thus, the effect of wealth inequality of asset prices likely generalizes to setups featuring a richer portfolio choice problem, but which I am forced to abstract from for the sake of tractability. My preferred interpretation is that the interest rate in this economy measures the returns on tradable wealth.

1.2.1 Model

At time t_0, the economy is populated by a continuum of agents i. Each agent owns a firm paying dividends at rate y_i. Initially, the dividend of each agents' firm grows at a high rate μ^H. However, each firm is risky in the sense that high growth firms can decay and become low growth firms with growth rate $\mu^L < \mu^H$. This decay occurs according to a Poisson process with instantaneous intensity $\lambda\, dt$. By the law of large numbers, aggregate dividends are deterministic and there is no aggregate risk. Agents in the economy have log preferences and seek to maximize expected utility given subjective discount parameter ρ

$$U\left(\{c_t\}\right) = \mathbb{E} \int_0^{\infty} e^{-\rho s} \log c_s \, ds$$

where the expectation is taken over both the transition time and the death time experienced by the agent.

Agents are born at rate δ owning firms whose initial dividend Y is drawn from a distribution with mean κY. The exact distribution of new firm dividends will affect the stationary wealth distribution in the economy, but not the main results presented, which hold for any positive support distribution with mean κY. Agents are in a state of perpetual youth and low-growth agents die at an i.i.d rate δ. Firms of deceased agents do not disappear, but instead continue to grow and produce output for consumption. Figure 1.2 plots a potential sample path for a firm with initial dividend y_0. Up to time t_λ, the firm grows at rate μ^H, and grows at rate μ^L forever after, even though the founder passes away at time t_δ.

As in Blanchard (1985), I assume that a competitive annuity market exists which redistributes the wealth of deceased agents proportionately among surviving agents according to their wealth. This assumption serves to allow agents to perfectly hedge their individual mortality risk. The law of motion of total output Y in this economy is given by

$$\frac{dY}{Y} = \left(\mu^H x + \mu^L (1-x) + \delta\kappa\right) dt \tag{1.1}$$

where x denotes the output share of high growth firms, $x = \frac{Y^H}{Y}$ and Y^H denotes total output of firms with high dividend growth rates. The laws of motion of total output of high- and low- growth firms are

$$\frac{dY^H}{Y^H} = \left(\mu^H + \frac{\delta\kappa}{x} - \lambda\right) dt \tag{1.2}$$

$$\frac{dY^L}{Y^L} = \left(\mu^L + \lambda\frac{x}{1-x}\right) dt \tag{1.3}$$

Over an interval dt, high type firms' output grows by μ^H and a fraction $\lambda\, dt$ of the high type firms transition and become low growth firms. Newly entering agents owning firms with aggregate dividends $\delta\kappa Y$ further increase the growth rate of high type output. For low type output, incumbent firms' output grows by μ^L, and output is further increased by the arrival of transitioning firms into the low growth state. As firms do not disappear upon the founder's death, δ does not appear in the growth rate of Y^L. Displacement in this economy proceeds deterministically, wherein new firms are born at a constant rate and comprise a constant share of aggregate output. I now introduce a borrowing constraint which limits the high type agents' participation in financial markets. Under no-trade, these agents consume the dividends of their firms, which grow at rate μ^H. The high-type agents would like to borrow against their firms in order to smooth consumption. The dividend yield is relatively low for high growth firms, and thus in autarky, these agents under-consume relative to their total wealth . On the opposite extreme, absent frictions, the low-type agents would lend to the high-type agents and expected consumption growth would be equalized across all agents. Agents owning high growth firms over-consume in the short term, finance their excess consumption with loans, and repay these loans once their firms transition to the low growth state. The constraint limits this by restricting high type agents' ability to borrow. Specifically, I impose that agents cannot sell their firms and cannot credibly promise to repay more than proportion α of their dividend income y. The problem of a high growth agent a firm paying dividend y and loan balance l is therefore

$$V^H(y,l) = \max_c \left\{ u(c)\,dt + e^{-\rho dt}\left(e^{-\lambda dt}V^H(y',l') + \left(1-e^{-\lambda\,dt}\right)V^L(y',l')\right)\right\} \tag{1.4}$$
$$\text{s.t.}\quad \mathbb{E}_t\int_0^\tau e^{-rs}\left(c_{t+s}-y_{t+s}\right)ds \le \alpha y_t \quad \forall t,\tau$$
$$y' = y\left(1+\mu^H dt\right)$$
$$l' = l\,(1+rdt) + (c-y)\,dt$$

Low type agents are repaying their loans and lending to current period high type agents. A low growth agent is therefore unconstrained by the borrowing limits and solves the problem

$$V_L(y,l) = \max_c \left\{ u(c)\,dt + e^{-(\rho+\delta)dt}V_L(y',l')\right\} \tag{1.5}$$
$$\text{s.t.}\quad y' = y\left(1+\mu^L dt\right)$$
$$l' = l\,(1+rdt) + (c-y)\,dt$$

Definition 1. *A symmetric steady-state equilibrium consists of agent masses* m_H*, and* m_L*,* *an interest rate r, and consumption policies c_i for $i \in \{H, L\}$ such that*

1. *Agent masses are constant over time*

2. *The borrowing and consumption policies solve the optimization problem of high- and low-type agents, taking agent masses and the interest rate as given*

3. *The consumption market clears*

4. *The lending market clears*

In steady state, a fraction $\frac{\delta}{\delta+\lambda}$ of firms will be in the high growth rate, and a fraction $\frac{\lambda}{\delta+\lambda}$ will be in the low growth state. For the economy to be stationary, the output of high- and low-growth firms must grow at the same rate. When low-growth firms make up a smaller fraction of the economy, their lower intensive growth rate is supplemented by a high extensive margin of growth coming from decaying high-growth firms.

Proposition 1. *The steady state output share of high growth firms is*

$$x = \frac{\sqrt{(\delta\kappa - \mu_H + \lambda + \mu_L)^2 + 4\delta\kappa(\mu_H - \mu_L)} - (\delta\kappa - \mu_H + \lambda + \mu_L)}{2(\mu_H - \mu_L)}, \tag{1.6}$$

and output growth is given by

$$g_Y = \mu_H - \lambda + \frac{\delta\kappa}{x}. \tag{1.7}$$

Furthermore, the steady state output share of high type firms is decreasing in μ^L and in λ, and increasing in κ:

$$\frac{dx}{d\mu^L} = -\frac{1-x}{\mu^H - \mu^L + \delta\kappa/x^2} < 0, \tag{1.8}$$

$$\frac{dx}{d\lambda} = -\frac{1}{\mu^H - \mu^L + \delta\kappa/x^2} < 0, \tag{1.9}$$

and

$$\frac{dx}{d\kappa} = \frac{\delta + \kappa/x}{\mu^H - \mu^L + \delta\kappa/x^2} > 0, \tag{1.10}$$

Intuitively, the steady state output share of high type firms is higher when low type firms grow slowly, $\frac{dx}{d\mu^L} < 0$, and is lower when new firms spend less time as high type firms, $\frac{dx}{d\lambda} < 0$.

Solution to the Low Type's Problem Under the assumption of log preferences, the low type agent finds it optimal to consume a constant fraction $\rho + \delta$ of her total wealth, given by the value of her firm plus her financial wealth

$$w = \frac{y}{r - \mu^L} + l$$

so that her net growth rate of total wealth is $r - (\rho + \delta)$. Even though firms cannot be bought or sold among living agents, firms can be priced via a no-arbitrage relationship which implies that the value of a low growth firm is the discounted present value of a growing perpetuity. The results are entirely unchanged if the setup is modified to allow for the purchase and sale of low type firms. High type agents are constrained and would not purchase these firms, while low type agents are indifferent between owning their personalizing growing perpetuity or a basket of identical growing perpetuities.

Solution to the High Type's Problem High-growth agents will find it optimal to always be at the leverage constraint αy. An agent who does so has locally deterministic consumption growth of μ^H whereas the borrowed amount grows at rate r. Thus, she will borrow the maximum amount as long as $\mu^H > r - \rho$. For an agent who does not borrow up to the constraint, they can increase their utility by borrowing ε more today at rate r, consuming it, and repaying $\varepsilon e^{r\,dt}$ out of tomorrow's dividend. I provide a formal proof in Appendix 1.7.

Given that the decay rate is i.i.d and the leverage constraint is proportional to dividends y, this argument is independent of the current level of dividends and holds for all high-growth agents. Thus, high growth agents consume in excess of their income. Conditional on remaining a high-type, they have consumption growth equal to μ^H, and their consumption-dividend ratio is given by $1 + \alpha\left(r + \lambda - \mu^H\right)$.

The leverage constraint will bind as long as it prevents agents from consuming their optimal amount. This optimal amount corresponds to the consumption-income ratio of a newborn agent who was free to sell her firm and reinvest at the prevailing interest rate r.

Under log preferences, the agent will consume a constant fraction ρ of her wealth, which is given by

$$\frac{y}{r-\mu^L}\frac{\lambda+r-\mu^L}{\lambda+r-\mu^H} \tag{1.11}$$

Therefore, the constraint always binds in equilibrium as long as

$$1+\alpha\left(r+\lambda-\mu^H\right) < \frac{\rho}{r-\mu^L}\frac{\lambda+r-\mu^L}{\lambda+r-\mu^H} \tag{1.12}$$

and the consumption of a high type agent owning firm with current dividend y is given by ϵy, where

$$\epsilon := \min\left\{1+\alpha\left(r+\lambda-\mu^H\right), \frac{\rho}{r-\mu^L}\frac{\lambda+r-\mu^L}{\lambda+r-\mu^H}\right\}$$

In either case, consumption conditional on remaining in the high growth state grows at rate μ^H. The equilibrium interest rate r affects the consumption-income ratio of high type agents, but not the growth rate of consumption.

Financial Markets New loans are made to finance high type consumption at rate $(\epsilon-1)\,Y^H dt$. These loans accrue interest and are repaid by agents after their firms transition to the low growth state. By the law of large numbers, fraction $\lambda\, dt$ of high type firms decay over interval dt, and the agents owning those firms have loans in aggregate totaling $\lambda L\, dt$. The law of motion for loans outstanding to high type agents is given by

$$dL = \left((r-\lambda)\,L + (\epsilon-1)\,xY\right)dt \tag{1.13}$$

and total wealth of low type agents is the sum of loans outstanding and value of all current low type firms

$$W^L = \frac{(1-x)\,Y}{r-\mu_L} + L \tag{1.14}$$

Proposition 2. *In steady-state, the net wealth of low type agents is given by*

$$W^L = \left(\frac{1-x}{r-\mu_L} + \frac{\epsilon-1}{g+\lambda-r}x\right)Y \tag{1.15}$$

The interest rate r^* satisfies the market clearing condition

$$\epsilon x + (\rho + \delta) \frac{W_t^L}{Y} = 1 \tag{1.16}$$

By Walras' law, once the consumption market clears, the lending market will also clear.

Proposition 2 states that the wealth of low type agents is given by the value of the low growth firms in the economy, plus the value of loans outstanding to high type agents. In equilibrium, there are also some outstanding loans made to former high type agents that have yet to be repaid, but these are simply transfers among the low type agents and do not affect the aggregate wealth of low type agents. The interest rate in the constrained economy is determined by consumption market clearing. Too low an interest rate results in excess demand, as low types seek to consume a constant fraction of their wealth. As the interest rate increases, these agents prefer to save and enjoy the higher rate of return. In the opposite case, too high an interest results in a demand shortage as low types prefer to save rather than consume, and thus the interest rate needs to decline to encourage low type agents to consume more.

1.2.2 Growth Rates

Fixing a cohort, the log growth rate of cohort wealth, which I refer to as a *cohort growth rate*, is given by

$$\frac{1}{t} \log \frac{W_t}{W_0} = e^{-\lambda t} \left(\mu^H - \rho\right) + \frac{1}{t} \int_0^t \left(rt - (\rho + \delta) t + \left(\mu^H + \delta - r\right) s - \ln \varphi\right) \lambda e^{-\lambda s} ds \tag{1.17}$$

$$= r - (\rho + \delta) + \frac{1 - e^{-\lambda t}}{\lambda t} \left(\mu^H - \rho - r - \lambda \ln \varphi\right) \xrightarrow[t \to \infty]{} r - (\rho + \delta) \tag{1.18}$$

where

$$\varphi = \frac{\lambda + r - \mu^L}{\lambda + r - \mu^H}$$

is the ratio of the value of a high-growth firm to the value of a low-growth firm with the same level of dividends. Thus, in this economy the cohort growth rate reveals the returns on financial wealth and tradable assets.

Agents' wealth growth is characterized by a period of high initial growth followed by modest growth in the long term. Figure 1.3 plots the stationary distribution of wealth in this economy, distinguished between high type and low type agents. High type agents are wealthier on average and over-represented in the upper tails of the wealth distribution. Measurements of wealth growth done using repeated cross-sections of a top percentile above threshold q, as in Gomez (2018), can be written as

$$\sum_{j\in\{H,L\}} \mu^j \int_{q_t}^{\infty} f^j\left(w_t\right) dw_t$$

where f is the joint density of growth rates and wealth. In a stationary economy, the distribution f is invariant over time and thus the measured wealth growth is constant over time. In particular, these measured growth rates are a function of μ^H, whereas the cohort growth rate, and asset prices, are only a function of r. In this economy, $\mu^H > r$, and so measurements based on repeated cross sections will over-estimate the long-run growth rate of wealth due to the transition dynamics. Wealthiest households today will always have a high growth rate, yesteryear's wealthy households have transitioned and now grow at a rate reflecting asset returns. This is equivalent to saying that the wealth dynamics captured by cohort growth rate converge to the wealth dynamics of the marginal agent, while measurements from repeated cross sections are biased at all horizons. Figure 1.4 plots the relative population of high type agents above a cutoff level of wealth, $\mathbb{P}\left[\mu = \mu^H \mid W \geq w\right]$, when the distribution of new firm dividends is exponential. The growth rate estimated from repeated cross sections is significantly higher than the true long run growth rate of a cohort, and this bias is independent of horizon.

Importantly, the use of cohort growth rates is valid even in the absence of transition dynamics. Such a case can be modeled either by eliminating the motive to borrow through equating μ^L and μ^H. In this case, every agents' wealth grows at r less consumption as every agent is marginal in the bond market. Thus, the use of cohort growth rates is a more robust method of determining the stochastic discount factor, as it recovers the right discount factor in the friction-less case as well as in the case of constrained agents. In addition, cohort growth rates are valid even when the econometrician cannot directly observe which agents are constrained. When agents know their type but the econometrician does not, cohort growth rates are robust to selection biases, as all agents decay to the low type in the long run.

1.2.3 Wealth Inequality and Asset Prices

Within the model, the contribution of new entrants and incumbents to rising wealth inequality is governed by two parameters. Wealth accumulation by new entrants is increasing in κ, the output share of new firms. When new firms are more valuable, the agents who own those firms are the wealthiest agents in the economy. On the other hand, wealth accumulation by incumbents is increasing in μ^L, the growth rate of old firms. In the model, these firms' growth rates determine the investment opportunities available to low type agents. When these investment opportunities are comparatively valuable, the wealthiest agents in the economy are those who were born with valuable firms and had the good fortune to live for a long time, accumulating wealth all the while. I now examine the effect that a relative shift in these parameters has on interest rates. I show that while rising wealth inequality driven by new entrants results in a decline in the interest rate, increasing wealth inequality driven by incumbents results in an increase in the interest rate. Both an increase in κ and an increase in μ^L have the effect of increasing estimates of wealth growth constructed using repeated cross sections, but, as shown in Figure 1.5, the interest rate falls when displacement, captured by κ, increases. Cohort growth rates accurately reflect this decline in incumbent households' wealth growth. This further motivates my empirical methodology of estimating cohort growth rates using panel data.

1.3 Empirics

In this section, I detail the construction of my data set and present the results of analysis using that data set. I present the data sources used in Section 1.3.1. I detail how I combine

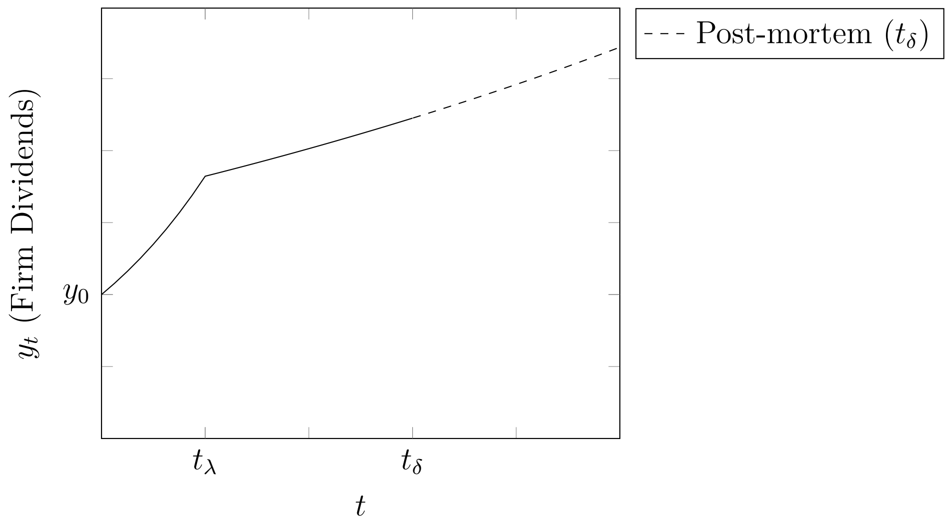


Figure 1.2: Illustrative Firm Dynamics. The figure plots a representative draw of dividends y_t for a firm owned by an agent born at time $t = 0$. At time t_λ, the firm transitions to the low growth state. At time t_δ, the owner dies. It is important to note that the firm continues to produce output and grow at rate μ^L after the owner passes away.

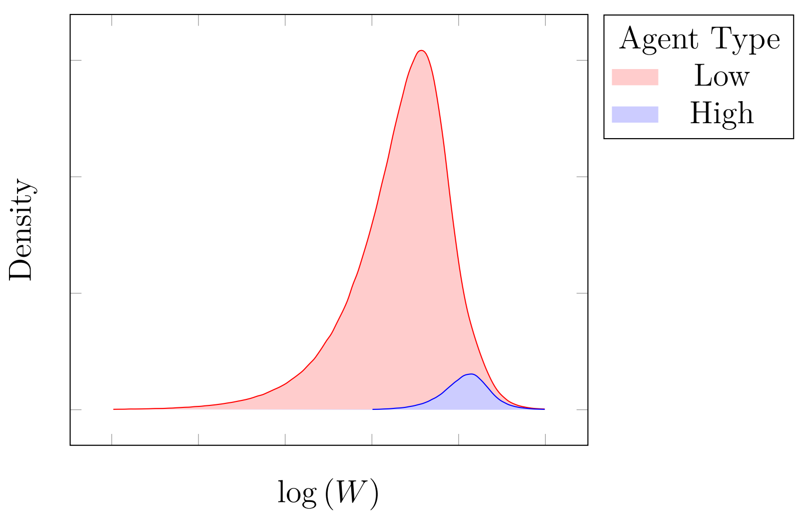


Figure 1.3: Stationary Wealth Distribution. New agents are born with wealth drawn from an exponential distribution with scale parameter κ.

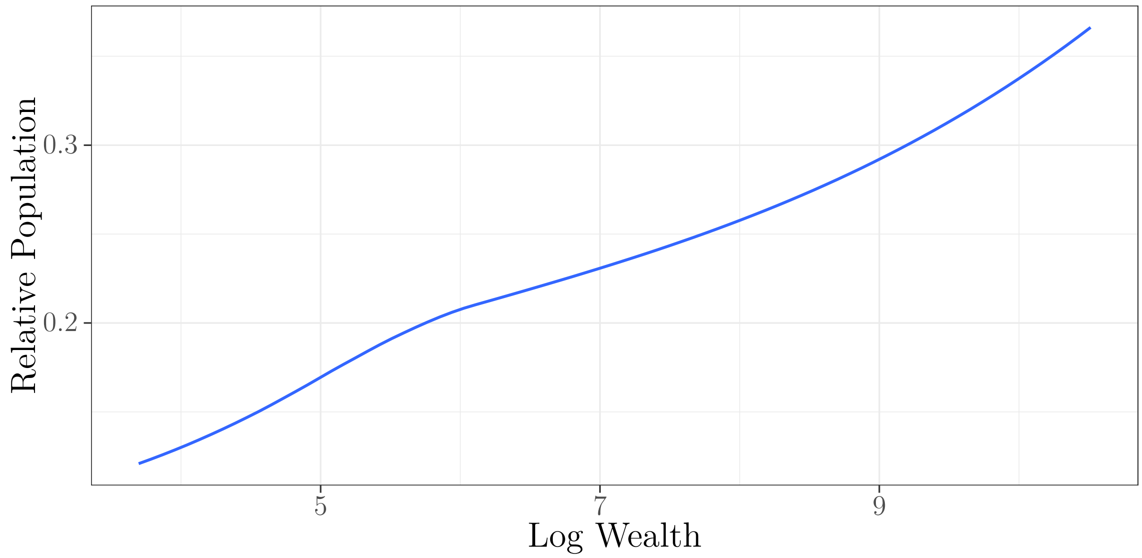


Figure 1.4: Fraction of High Type Agents. The figure plots the fraction of high type agents among the population of agents with wealth greater than cutoff q. Agents in the upper percentiles of the wealth distribution are more likely to be in the high growth state.

the different data sources into a single panel in Section 1.3.2. I detail the aggregation of individual observations into populations of entry and incumbent cohorts in Sections 1.3.3 and 1.3.4, respectively. I present findings in Section 1.3.5.

1.3.1 Data

The initial construction of my panel begins with the Forbes 400 data set, published annually since 1982. By starting with Forbes 400 lists, I have a number of repeated observations for the same individual over many years. The data collection challenge of this paper is to fill in wealth observations missing in the Forbes 400 lists.

Forbes Dropoff Lists In order to account for dropouts from the Forbes 400, I employ a number of data sources. The first auxiliary data set is Forbes Magazine's own published list of drop offs, beginning in 2012. For all subsequent Forbes 400 lists, Forbes Magazine reported the wealth of individuals who were removed from the list on the grounds that they were no longer among the 400 richest Americans. I manually collect these reports from archives of Forbes' website. The weaknesses of this data set are that: (i) it only exists since

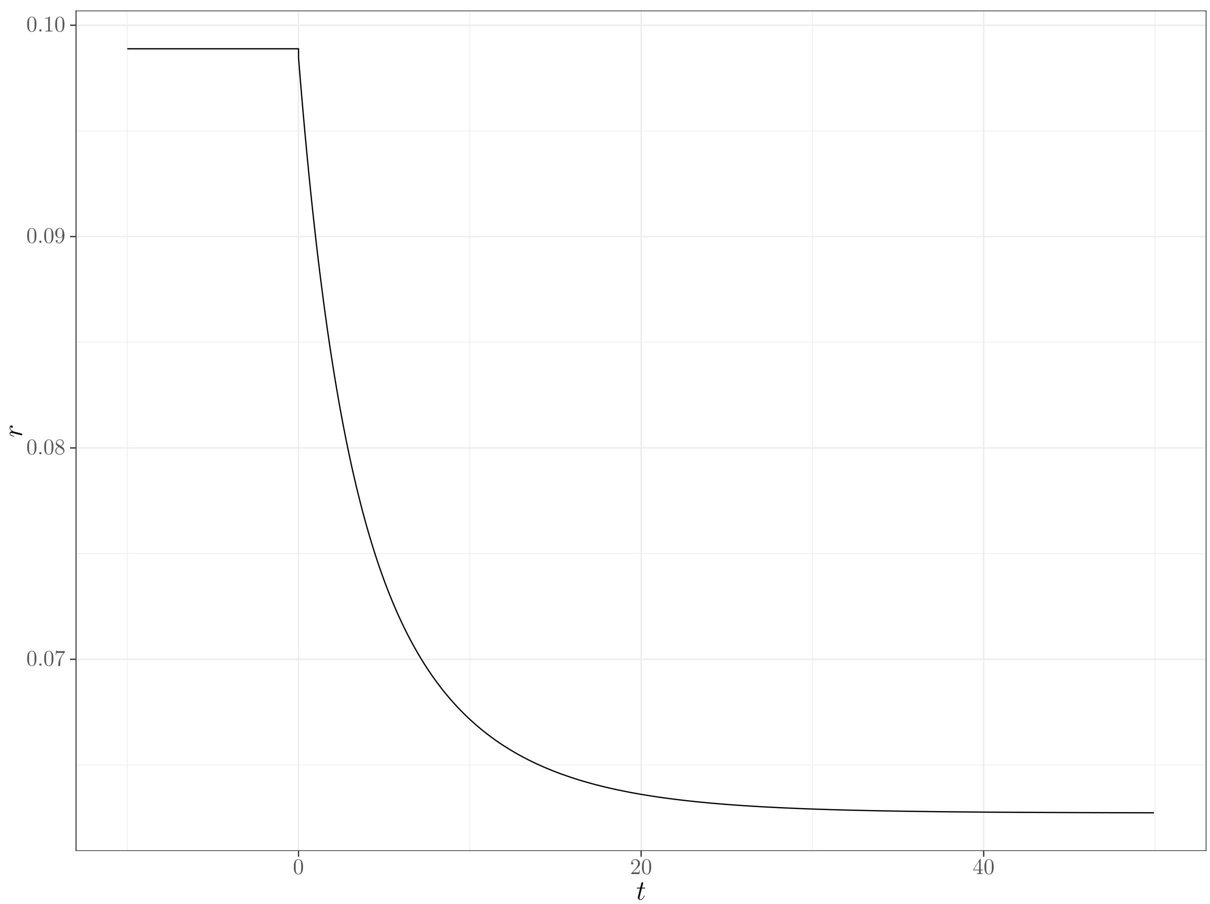


Figure 1.5: Transition path of interest rates. Prior to time $t = 0$, the economy is in steady-state. Following an increase in κ and a decrease in μ_L, the interest rate r_t experiences an immediate discontinuous drop, following by a protracted smooth decline to the steady interest rate under the new parameters.

2012, (ii) it only contains wealth for dropoffs in the year immediately following their exit from the Forbes 400 list, and (iii) it does not report wealth for deceased individuals.

Forbes Billionaire Lists The second auxiliary data set is Forbes Magazine's published list of world-wide billionaires. This list was first compiled in 1996, and continues to this day. I scraped the historical Forbes Billionaire lists from archives of Forbes' website. Individuals who fall off the Forbes 400 list, but who remain billionaires, stay in the Forbes Billionaire data set This is the case for a number of individuals, and I am able to combine the data sets to create a balanced panel of wealth for these individuals extending through to 2018. Another advantage of the Forbes Billionaire list is that it assists me in estimating the wealth of deceased Forbes 400 individuals.

Family Structures for Forbes 400 members In order to identify family members, I manually collect data on the names and, where possible, age and location of children and spouses of Forbes 400 individuals. Consistent with Bernstein and Swan (2008), I find that the average Forbes 400 individual has three children. I hand collect data on the number and the names of children using a variety of internet data sources. For deceased Forbes 400 members, their obituaries often contain information on surviving family members. Even for surviving individuals, or individuals for whom I could not locate an obituary, it is possible to obtain names of family members using obituaries of close relatives. [1] In total, I identified 4,843 children of Forbes 400 members, and found names and other information for 4,578 of those children. A detailed list of sources used in the construction of this data set is available upon request.

---

[1] In some cases, Forbes 400 members or their spouses have written books and included dedications to their children. This is the case for, among others, Robert and Janice Davidson, as well as David Shaw. The Davidsons wrote *Genius Denied: How to Stop Wasting Our Brightest Young Minds.* David Shaw's wife Beth Kobliner wrote *Make Your Kid A Money Genius (Even If You're Not): A Parents' Guide for Kids 3 to 23.* More esoteric examples include Pincus Green, whose children jointly wrote a letter to then-president Bill Clinton requesting a presidential pardon for their father.

LexisNexis Property Records In order to account for individuals not found in the Forbes data sets, due either to dropping off prior to 2006 or dropping to below $1 billion in net worth, I make use of the LexisNexis Public Records data set. LexisNexis offers a search interface through which I can observe basic biographical information, along with address history and property records, for a significant proportion of the American population. Starting with the biographical information included in the Forbes 400 lists, I search for individuals in the LexisNexis database based on name, approximate age, and state of residence. From there, I reject potential matches based on employment history and family information. Through this process, I manually link 1,565 Forbes 400 individuals to a unique `LexID`.

For each of the 1,565 Forbes 400 individuals that I am able to uniquely identify in LexisNexis, I download all property deeds and property assessments pertaining to that individual, as well as the names and addresses of all likely family members. For each likely family member, I then find the most likely matched `LexID` corresponding to that individual in the LexisNexis database, based on biographical information, and download all property deeds and assessments pertaining to these potential family members. I aggregate property records at the family unit, so that all family members' property records are grouped together. I further process the property records data to account for duplicates and potentially mis-labeled records using two methods. First, I exclude non-apartment properties sharing identical GPS coordinates. Second, I exclude any remaining properties which feature substantially similar parcel numbers. Finally, I use textual analysis to exclude commercially zoned properties.

Wealth-X Profiles The final non-standard data set that I use to produce my panel consists of Wealth-X profiles on ultra-wealthy individuals, defined here as individuals with net worth exceeding $30 million as of 2018. The profiles are maintained by dedicated staff employed by Wealth-X, and contain information derived from publicly disclosed transactions, holdings, philanthropy, conspicuous purchases, board memberships, professional and family ties, and other biographical information. I first extract a list of all ultra-wealthy individuals, both foreign and domestic, in the Wealth-X database. Based on this list of individuals, I then collect each profile and extract family details and portfolio holdings. Thus, my data set contains every individual Wealth-X has identified as having a net worth exceeding $30 million in 2018. In this paper, I principally focus my attention on domestic ultra-wealthy individuals, and thus discard all individuals with no business or residential addresses within the United States. I then manually match these individuals to Forbes 400 family units based on the hand-collected family structure information.

1.3.2 Methodology

In this Section, I discuss the procedure by which I combine different data sets into a single panel of household wealth. I begin with the Forbes 400 lists, and combine family units so as to minimize the contribution of death and bequests. Thus, in any year that a family appears in the Forbes 400, I take the Forbes 400 wealth to be the total wealth of that family. I now begin filling in missing observations from the panel of wealth. Starting from the year a family first enters the Forbes 400 list, I impute missing observations using the Forbes Dropoff lists, the Forbes Billionaire lists, and my estimates based on the family's residential property holdings.

For individuals and families who exit the Forbes 400 after 2012, the Forbes Dropoff list contains a single additional observation in which Forbes Magazine staff estimate their wealth. This estimate is the basis of Forbes' decision to exclude the individual. Whenever available, I fill in missing observations using these reported values. For the remaining observations, I first attempt to fill in missing wealth observations using data from the Forbes Billionaire lists, which go back as far as 1996. Even in the post 2012 period, the Forbes Billionaire list contributes to filling in missing observations for individuals who exit due to death and for any years following the year of immediate exit from the Forbes 400.[2]

To fill in the remaining missing observations, I use estimates of the family's residential portfolio holdings, collected from LexisNexis, to impute a wealth value for each missing observation. I impute missing wealth observations from housing value observations. For each household i, I collect the first record date t_{ij}^{start} and final record date t_{ij}^{end} for each piece

[2]Forbes Magazine's Dropoff lists do not report wealth for deceased members of the Forbes 400.

of residential property j associated with any household members identified in LexisNexis. I then collect housing values h_{ij} for years $t \in \left[t_{ij}^{\text{start}}, t_{ij}^{\text{end}}\right]$ by using the most recent property valuation. In cases in which purchase and/or sale price from deed records are available, I exclusively use those prices, rather than relying on more recent property assessments.

I aggregate property values at the household level by summing the value of all properties j owned by household i in year t to arrive at a total housing value H_{it}:

$$H_{it} := \sum_j h_{ijt} \mathbb{I}_{\left[t_{ij}^{\text{start}}, t_{ij}^{\text{end}}\right]} (t) \, .$$

I then use the panel of total housing values to impute unobserved wealth observations based on the following definition

$$\hat{W}_{it} := W_{is}^* \left(\frac{H_{it}}{H_{is}} \right)^{\varepsilon}, \quad s := \max \left\{ \tau \leq t \mid W_{i\tau}^* \text{ exists} \right\}, \tag{1.19}$$

where $\varepsilon = 1$ in my primary specification. In Appendix 1.9, I discuss the economic assumptions motivating this imputation and elaborate on the strengths and weaknesses of this imputation method. The imputation procedure using real estate can be described in simple terms: for a given year t in which I observe housing values H_{it} for household i, but not wealth W_{it}^*, I estimate that the unobserved household wealth is equal to last known value of wealth from year τ, multiplied by the percentage increase in the household's housing value between years τ and t.

1.3.3 Cohort Identification and Aggregation

Using the methodology described above, I construct a survival bias-free panel of wealthy individuals. A primary contribution of my paper is the decomposition of wealth inequality, and in particular documenting heterogeneous contributions to increasing wealth inequality. To focus on this heterogeneity, I group Forbes 400 households by their year of entry into the Forbes 400. This corresponds to the "birth" of the cohort in the model, and represents the earliest point in time for which I have wealth estimates for each household. I further

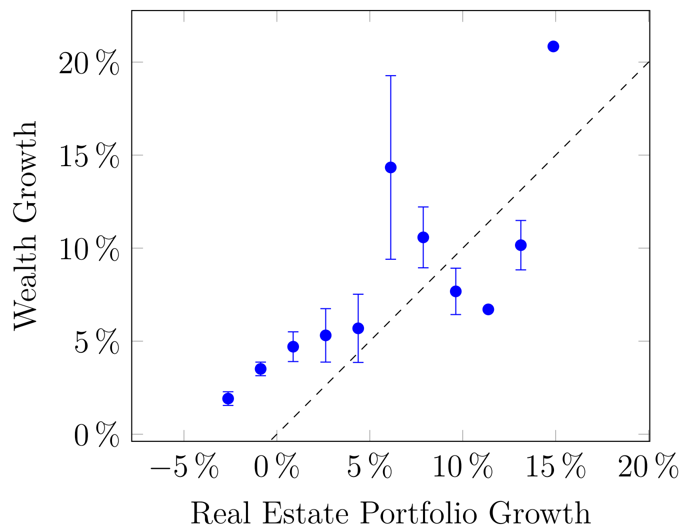


Figure 1.6: Real Estate Portfolio Growth and Total Wealth Growth, 5-Year Horizon. For individuals who remained on the Forbes 400 list, I plot the annualized growth rate of real estate portfolio growth, obtained from LexisNexis, against the annualized growth rate of wealth, obtained from Forbes 400 lists.

aggregate cohorts at the five year horizon, so that my first cohort corresponds to households which entered the Forbes 400 between 1982 and 1986, the second cohort contains households which entered between 1987 and 1991, and so on. Summary statistics on the coverage of my panel are show in Table 1.1. Figures 1.6 and 1.7 present binned scatter plots of cohort-level wealth growth against cohort-level real estate growth at the five- and ten-year horizon, respectively, for individuals within each entry cohort that remained on the Forbes 400 list. The relatively good fit motivates the assumption of a constant portfolio share in real estate. These figures are conditional on the household remaining on the Forbes 400 list, so that it is possible to calculate their realized growth rate of wealth, independent of any imputation procedures.

As discussed in Section 1.2, cohort growth rates are a robust means of measuring household growth rates of wealth in the presence of constraints and heterogeneity. In addition, there are two statistical reasons to aggregate households at the cohort level. First, while there are approximately 1,400 distinct households in my panel, almost 400 of these households entered the list in the inaugural publication of the Forbes 400 list. Thus, there are on average less than 30 households which enter the Forbes 400 in a given year. In such a small

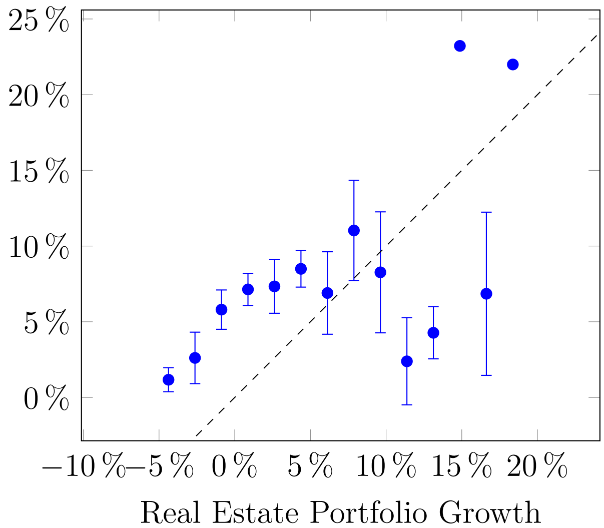


Figure 1.7: Real Estate Portfolio Growth and Total Wealth Growth, 10-Year Horizon. For individuals who remained on the Forbes 400 list, I plot the annualized growth rate of real estate portfolio growth, obtained from LexisNexis, against the annualized growth rate of wealth, obtained from Forbes 400 lists.

population, idiosyncratic wealth shocks still play a large role, and thus the wealth dynamics of small cohorts are imprecisely measured in the data. The concern here is a cross-sectional one; I want to compare long run growth rates across cohorts, and thus my estimates need to be precise enough to distinguish trends across cohorts. The second reason is that my focus is on long run growth rates of wealth, and thus combining cohorts simplifies the time-series analysis. An alternative specification featuring overlapping yearly fixed effects and cohort effects would both complicate the analysis and limit my statistical power by introducing many more degrees of freedom.

1.3.4 Incumbent Identification and Aggregation

In every year, the top wealth percentile is populated by individuals from multiple entry cohorts. To measure the contribution of this incumbent population to rising top wealth shares, I group households based on the years they appear in the Forbes 400 list. In year t, all households that appeared on the Forbes 400 list in the prior five years are included in the year t incumbent population, and I measure incumbent wealth at time t as of year T as the total wealth of this fixed set of households as of year T. With the exception of the 1986

Cohort	Size	Forbes	Augmented Forbes	Imputed Panel
1982–1986	552	61	71	163
1987–1991	159	30	39	75
1992–1996	145	35	46	94
1997–2001	181	43	52	125
2001–2006	107	40	51	92
2006–2011	112	57	83	106
2011–2016	88	56	76	85
2017–2018(*)	28	28	28	28

Table 1.1: Summary statistics for Entry Cohorts. Entry Cohorts are defined based on the first year the family is observed in the Forbes 400. The total number of the individuals in the cohort is listed under Size. The number of cohort members with wealth estimates in the 2018 Forbes 400 list is presented under Forbes. The number of cohort members with wealth estimates in some Forbes Magazine publication is presented under Augmented Forbes. The number of cohort members with estimates in my panel is presented under Imputed Panel. Due to its small relative size, I exclude the 2017-2018 Forbes 400 cohort in my empirical analysis.

incumbent population, this is different from the year t entry cohort.

Using entry cohorts as the unit of analysis is useful when the interest is in documenting heterogeneous growth rates. There the identification strategy is essentially to compare contemporaneous realized growth rates across entry cohorts. However, for estimating the displacement term, the incumbent wealth growth rate is a sufficient statistic for the joint distribution of cohort growth rates and wealth shares. I present results based on yearly incumbent populations and compare the growth rate of several incumbent cohorts to the rise in wealth inequality over five-year staggered periods.

1.3.5 Results

I now present cohort growth rates estimated from my panel data set. A key finding is that older cohorts have lower growth rates compared to newer cohorts. I show this in several ways. I first compare average rates of return over the entire sample. I then compare contemporaneously estimated rates of return between different cohorts. This heterogeneity is consistent with the life cycle dynamics introduced in my model, in which younger households own the high growth firms, but decay to the low growth state over time.

	Entry Cohort					
Period	1986	1991	1996	2001	2006	2011
1986–1991	8.7 %					
1991–1996	6.5 %	10.3 %				
1996–2001	11.4 %	16.6 %	10.0 %			
2001–2006	2.9 %	2.1 %	7.6 %	7.5 %		
2006–2011	1.2 %	1.5 %	2.5 %	-0.4 %	3.3 %	
2011–2016	6.4 %	4.8 %	6.0 %	5.3 %	8.6 %	8.3 %
2016–2018	7.7 %	3.3 %	6.3 %	13.0 %	8.6 %	6.0 %
Whole Sample	6.3 %	6.8 %	6.6 %	5.2 %	6.4 %	7.7 %

Table 1.2: Period Growth rates, by Entry Cohort. Annualized wealth growth rates of different Entry Cohorts of the Forbes 400, measured across five-year periods. Entry Cohorts are as defined in the text. Whole Sample growth rates are annualized wealth growth rates of wealth from the first year of the Cohort to 2018.

Table 1.2 presents the five year wealth growths rate of each entry cohort of Forbes 400 households, along with the long-term growth rate of that cohort from its inaugural year through 2018. Older cohorts tend to have lower growth rates than newer cohorts. A notable exception is the 2001 cohort, which featured a number of dot com entrepreneurs who remained on the Forbes 400 for only a short period of time.

Later cohorts in my panel are only observed in the period after entering the Forbes 400. This makes comparing full-sample growth rates insufficient for identifying heterogeneous growth rates, as the sample averages are confounded by aggregate market returns in periods prior to a cohort's appearance in my panel. A potential explanation for these differences in sample averages could be that wealthy households all have a growth rate of wealth, driven by equity holdings, and that stock market returns were low in the late 1980's, and have progressively improved since then. Such a data-generating process would be consistent with common growth rates of wealth, yet different observed sample averages.

I account for time-varying drivers of growth rates in two different ways. The first approach is consistent with a concern that equity holdings, along with time-varying stock market returns, are driving my estimates. I run regressions on residuals of wealth growth after controlling for a time-invariant market loading. The specification is

$$\mu_{st} = \beta^{\text{Mkt}}\text{Mkt}_t + \beta_s + \varepsilon_{st},$$

where Mkt corresponds to the July through June Fama-French market factor return.[3] Effectively, I subtract 0.4 times the periods' Fama-French market factor return from each cohort-year observation. The coefficient of 0.4 comes from regressing my estimates of wealth growth against the market factor, and explains a substantial component of the time-variation in growth rates. Wealth individuals saw their wealth decline in down market periods such as the late 1990's and late 2000's. Results from regressing market-neutral wealth growth on cohort fixed effects are presented in Table 1.3, Column (3). I still find that older cohorts' wealth grows at a slower rate than that of younger cohorts.

The second way in which I account for time-varying common growth rates is the inclusion of year fixed effects in my regressions. The specification is

$$\mu_{st} = \alpha_t + \beta_s + \varepsilon_{st}$$

This method is silent on the factors driving time-varying growth rates of wealth. A limitation of this approach is that the level of the cohort fixed effects β cannot be disentangled from the level of the time fixed effects α. Results of these regressions, run at both one- and five-year horizons, are also reported in Table 1.3. With the exception of the five-year returns, all specifications estimate that younger cohorts grow faster than older cohorts. As a consequence of my normalization, the level of the coefficients in Columns (2) and (4) are not informative, and the appropriate test for heterogeneous growth rates is to look at the relative ordering of the cohorts, as well as the magnitude of the differences in returns, rather than the levels of the returns. In the case of Column (2), in which I conduct my analysis at the five-year horizon, it is only the newest 2006-2011 cohort that under-performs the oldest cohort. The overall trend is consistent across these different specifications.

[3] I use July through June to better line up with the publication of the Forbes 400 lists.

	Wealth Growth			
Cohort	(1)	(2)	(3)	(4)
1982–1986	6.3 %	1.9 %	1.9 %	0.0 %
1987–1991	6.8 %	1.9 %	2.5 %	1.0 %
1992–1996	6.5 %	2.3 %	2.7 %	1.2 %
1997–2001	5.2 %	2.5 %	1.8 %	1.8 %
2002–2006	6.4 %	2.3 %	2.4 %	3.2 %
2006–2011	7.6 %	1.5 %	2.4 %	2.4 %
New – Old	1.4 %	-0.3 %	0.5 %	2.4 %
Mkt	N	N	Y	N
Year F.E	N	Y	N	Y
N	126	28	28	126

Table 1.3: Decomposition of Entry Cohort Growth Rates. I decompose annualized entry cohort growth rates into common time-varying components and cohort-specific, time-invariant components. Column (1) reports realized growth rates for each entry cohort. Column (2) reports entry cohort growth rates, controlling for five-year binned fixed effects, Column (3) reports the residual growth rates after projecting entry cohort growth rates onto Market returns. Column (4) reports residual entry cohort growth rates, controlling for single year fixed effects.

The observed trend across the regressions suggests that age could be a likely contributor to this effect. I investigate the role of age by augmenting my regression specification to include cohort age effects alongside the common time-varying component and the cohort-specific growth rates. I report results of these specifications in Table 1.4. Column (1) corresponds to the specification

$$\mu_{st} = \beta^{\text{Mkt}}\text{Mkt}_t + \text{Age}_{st} + \varepsilon_{st},$$

where age is the number of years since that cohort's birth year, $s - t$. The economic interpretation of the coefficient is that the cohort that entered the Forbes 400 list at time $t - 5$ under-performs the time t cohort by 0.2 percentage points per annum. The oldest cohort in my panel entered in 1986, and the youngest cohort entered in 2011. From these results, I would predict that the 1986 cohort grows 1 percent slower each year than the 2011 cohort. This is compared to a difference in growth rates of 1.4 percentage points when comparing

	(1)	(2)	(3)	(4)
Age	$-0.2\,\%^{**}$		$-0.2\,\%^{**}$	
$I_{(\ 1,\ 5]}$ (Age)		$0.7\,\%$		$0.0\,\%$
$I_{(\ 6,10]}$ (Age)		$0.6\,\%$		$-0.2\,\%$
$I_{(11,15]}$ (Age)		$0.4\,\%$		$-0.5\,\%$
$I_{(16,20]}$ (Age)		$0.4\,\%$		$-0.5\,\%$
$I_{(21,25]}$ (Age)		$0.0\,\%^{*}$		$-1.0\,\%^{*}$
$I_{(26,30]}$ (Age)		$-0.2\,\%^{*}$		$-1.1\,\%^{*}$
$I_{(31,35]}$ (Age)		$0.2\,\%$		$-0.8\,\%$
Cohort F.E	N	N	Y	Y
Mkt	Y	Y	Y	Y
N	28	28	28	28

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

Table 1.4: Decomposition of Entry Cohort Growth Rates. I decompose entry cohort growth rates into common time-varying components, cohort-specific, time-invariant components, and common, cohort age dependent components. Column (1) reports the effect on age, controlling for contemporaneous market returns. Column (2) reports effects of age, estimated non-parametrically using a series of age buckets. Column (3) reports the age coefficient, controlling for cohort effects and market returns. Column (4) reports the same non-parametric estimates as in Column (2), where I additionally control for cohort-specific effects.

the sample averages reported in Table 1.2, and is within the range of estimates presented in Table 1.3. The effect is not driven by the choice of the 1986 and 2011 cohorts. To show this, I substitute the linear age effect for a sequence of age indicator variables, binned at the five year level. The results are reported in Column (2). With the exception of the very oldest cohort, I find a stable monotonically decreasing relationship in age. Furthermore, the magnitude of the difference in age fixed effects is similar to the coefficient from the linear specification. The coefficients are unchanged when I re-introduce cohort fixed effects. I report results for the linear specification including cohort fixed effects in Column (3), and for the fixed effects specification including cohort fixed effects in Column (4).

The presence of heterogeneous growth rates has quantitative implications for the estimation of long term growth rates and displacement. The growth rate of wealthy households at time t, and consequently the growth of the wealth share of wealthy households, depends on the relative wealth shares inside of the top wealth percentile. Different growth rates cause this composition to vary over time. In the results that follow, I show that the moderate heterogeneity in growth rates across cohorts results is substantively different conclusions regarding the sources of increasing wealth inequality. This can be seen visually in Figure 1.8, which plots the cumulative increase in wealth inequality since 1986 as well as the contributions due to incumbent growth and displacement.

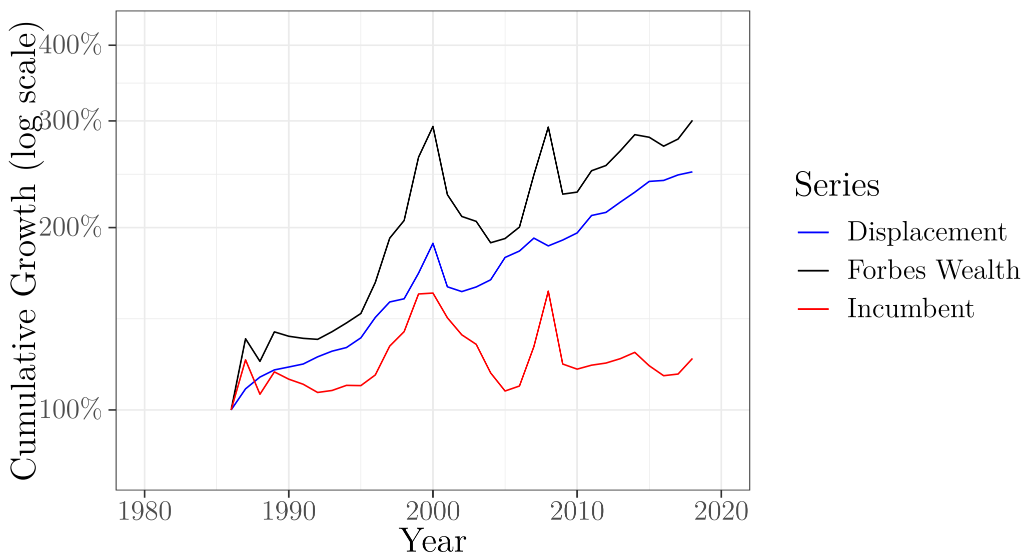


Figure 1.8: Decomposition of Wealth Inequality, 1986–2018. I plot the cumulative wealth growth of the 1986 Incumbent Cohort (Red), the Forbes 400 (Black), and the implied contribution of Displacement (Blue). Growth rates of incumbent cohort wealth and top wealth are deflated by the growth of aggregate wealth and should be interpreted as growth rates of incumbent cohort and top wealth shares.

In Table 1.5, I present estimates of wealth growth for ex ante wealthy households. I do this by fixing a population of Forbes 400 households who have appeared on the list prior to a given year t, and following that population of households through 2018. I refer to these as incumbent growth rates to distinguish from the cohort growth rates discussed earlier. A population of incumbent households as of year t includes households who entered the Forbes list anywhere between 1982 and year t, whereas the year t cohort of households only includes households who first entered the Forbes list within the five years prior to t. Therefore, the incumbent growth rate is the wealth-weighted average of cohort growth rates.

Using the estimates of incumbent cohort wealth growth, I can decompose the rise in wealth inequality into a within term and a displacement term. The within term captures

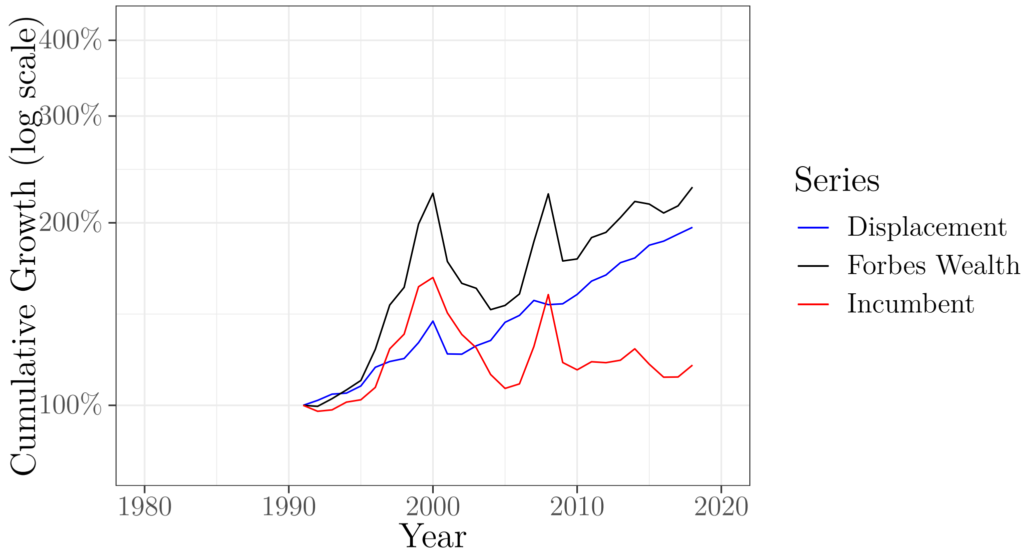


Figure 1.9: Decomposition of Wealth Inequality, 1991–2018. I plot the cumulative wealth growth of the 1991 Incumbent Cohort (Red), the Forbes 400 (Black), and the implied contribution of Displacement (Blue). Growth rates of incumbent cohort wealth and top wealth are deflated by the growth of aggregate wealth and should be interpreted as growth rates of incumbent cohort and top wealth shares.

the growth rate of already-wealthy individuals, while the displacement term captures the contribution of newly-wealthy individuals replacing previously-wealthy individuals in top wealth percentiles. Table 1.6 presents data from standard sources on the wealth growth of top wealth, captured by the Forbes 400; aggregate household wealth; and the relative increase in top wealth shares over a period of time. To decompose the component attributable to the within term, I compare aggregate household wealth growth to the wealth growth of households who entered the Forbes 400 in the five years prior to the period of interest. By comparing the Cohort column and the Forbes 400 column, we see that no cohort has outperformed the Forbes 400 as a whole over long periods and that the role of displacement is consistently large. Cohort growth rates are the estimates of the long-term growth rate of the cohort of newly wealthy households, and the results in the table indicate that high growth rates of wealth among newly wealthy households *after entering the Forbes 400* cannot explain the rise in wealth inequality.

In addition to the the population of newly wealthy households, we can also analyze the growth rates of ex-ante wealthy households. Table 1.7 presents the results of the same

Period	Incumbent Cohort 1986	1991	1996	2001	2006	2011
1986–1991	8.7 %					
1991–1996	6.5 %	7.2 %				
1996–2001	11.4 %	12.7 %	12.1 %			
2001–2006	2.9 %	2.7 %	3.5 %	4.3 %		
2006–2011	1.2 %	1.3 %	1.6 %	1.3 %	2.1 %	
2011–2016	6.4 %	6.0 %	5.8 %	5.8 %	6.9 %	7.4 %
2016–2018	7.7 %	6.7 %	6.6 %	8.1 %	8.8 %	8.5 %
Whole Sample	6.3 %	6.0 %	5.8 %	4.3 %	5.2 %	7.7 %

Table 1.5: Period Growth rates, by Incumbent Cohort. Annualized wealth growth rates of different Incumbent Cohorts of the Forbes 400, measured across five-year periods. Incumbent Cohorts are as defined in the text. Whole Sample growth rates are annualized wealth growth rates of wealth from the first year of the Cohort to 2018.

	Long-Run Growth Rates				Relative Contribution	
Period	Forbes 400	Cohort	Household	Inequality	Entry Cohort	Displacement
1986–2018	9.1 %	6.3 %	5.7 %	3.4 %	18 %	82 %
1991–2018	8.5 %	6.8 %	5.5 %	3.1 %	43 %	57 %
1996–2018	8.2 %	6.6 %	5.4 %	2.8 %	41 %	59 %
2001–2018	6.6 %	5.2 %	4.9 %	1.7 %	16 %	84 %
2006–2018	7.0 %	6.4 %	3.6 %	3.4 %	82 %	18 %
2011–2018	9.1 %	7.7 %	6.4 %	2.7 %	45 %	55 %

Table 1.6: Decomposition of Wealth Inequality, by Entry Cohort. For five year staggered periods, I present annualized growth rates of the Forbes 400, the most recent Entry Cohort as of the start of the Period, and Aggregate household wealth. The difference between the growth rates of the Forbes 400 and Aggregate household wealth is the increase in Inequality. The relative contributions to wealth inequality of the Entry Cohort and of Displacement are presented in the last two columns. Entry Cohort is calculated as $(\text{Cohort} - \text{Household})\,(\text{Forbes 400} - \text{Household})^{-1}$

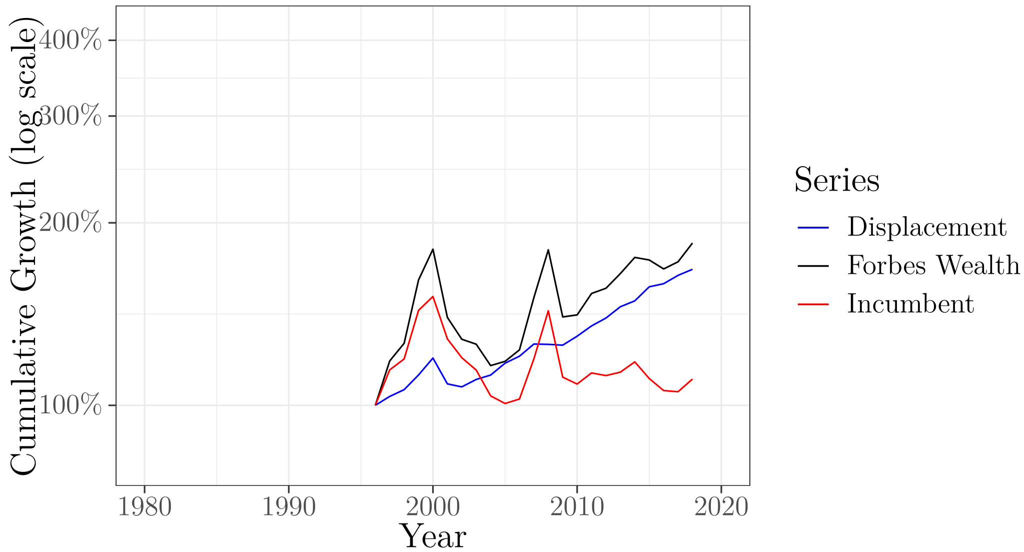


Figure 1.10: Decomposition of Wealth Inequality, 1996–2018. I plot the cumulative wealth growth of the 1996 Incumbent Cohort (Red), the Forbes 400 (Black), and the implied contribution of Displacement (Blue). Growth rates of incumbent cohort wealth and top wealth are deflated by the growth of aggregate wealth and should be interpreted as growth rates of incumbent cohort and top wealth shares.

decomposition, where incumbent growth rates are used rather than cohort growth rates. Incumbent households are those who were on the Forbes 400 at any point in the 5 years prior to the start of the period. From the consistently high relative contribution of displacement, we see that it is also not the wealth accumulation of ex-ante wealthy households that explains the bulk of the rise in wealth inequality. Both proxies for the within term lead to the conclusion that over 80 percent of the increase in wealth inequality since 1986 is the result of displacement.

With the exception of the 2001 Incumbent Cohort, rising inequality is the result of both a growing incumbent wealth share as well as displacement. The 2001 Incumbent Cohort is the only cohort for which the incumbent wealth share has declined, and this is likely attributable to tech bubble, which led to many one-time appearances on the Forbes 400. Those households suffered large drops in their wealth and exited the Forbes 400 list, leading to low estimates of the present day wealth of that Incumbent Cohort. The fact that the tech bubble as a industry-specific wealth shock is likely the reason that the 2006 incumbent cohort has grown their wealth share over time despite the Financial Crisis. Timing considerations

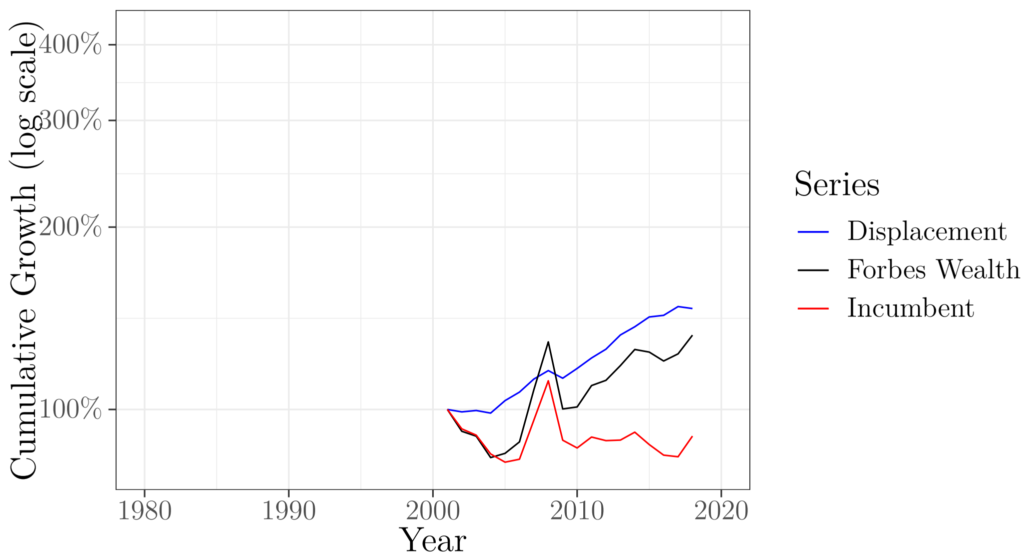


Figure 1.11: Decomposition of Wealth Inequality, 2001–2018. I plot the cumulative wealth growth of the 2001 Incumbent Cohort (Red), the Forbes 400 (Black), and the implied contribution of Displacement (Blue). Growth rates of incumbent cohort wealth and top wealth are deflated by the growth of aggregate wealth and should be interpreted as growth rates of incumbent cohort and top wealth shares.

also play a role. The relative wealth share of the Forbes 400 attained high water marks in the years 2000 and 2008. Thus, the 2006 incumbent cohort's initial wealth estimates do not reflect a fall from this local maximum.

The contribution of displacement has declined over the sample period, from over 80 percent since the late 1980's to just over 50 percent over the last 10 years. My decomposition of the cohort growth rates suggests that life-cycle effects play a role in this relative decline. More recent incumbent populations are earlier in the life cycle, so that their growth rates of wealth are still relatively high.

For comparison, I plot the cumulative contribution of displacement, estimated using one-year incumbent growth rates, in Figures 1.12 and 1.13. The chained one year growth rates overestimate the long term wealth growth of wealthy households, and consequently underestimates the contribution of displacement. For the full sample, starting in 1982, the relative contribution of the within and displacement terms are roughly equal, consistent with the results of Gomez (2018). For the sample starting in 1986, the within term calculated

Period	Long-Run Growth Rates				Relative Contribution	
	Forbes 400	Incumbent	Household	Inequality	Incum. Cohort	Disp.
1986–2018	9.1 %	6.3 %	5.7 %	3.4 %	18 %	82 %
1991–2018	8.5 %	6.0 %	5.5 %	3.1 %	18 %	82 %
1996–2018	8.2 %	5.8 %	5.4 %	2.8 %	16 %	84 %
2001–2018	6.6 %	4.3 %	4.9 %	1.7 %	-36 %	136 %
2006–2018	7.0 %	5.2 %	3.6 %	3.4 %	48 %	52 %
2011–2018	9.1 %	7.7 %	6.4 %	2.7 %	47 %	53 %

Table 1.7: Decomposition of Wealth Inequality, by Incumbent Cohort. For five year staggered periods, I present annualized growth rates of the Forbes 400, the most recent Incumbent Cohort as of the start of the Period, and Aggregate household wealth. The difference between the growth rates of the Forbes 400 and Aggregate household wealth is the increase in Inequality. The relative contributions to wealth inequality of the Incumbent Cohort and of Displacement are presented in the last two columns. Incum. Cohort is calculated as $(\text{Incumbent} - \text{Household})\,(\text{Forbes 400} - \text{Household})^{-1}$

using chained one year estimates of incumbent wealth growth outweighs the importance of the displacement term, and explains the bulk of the increase in the wealth share of the Forbes 400.

1.4 Implications for Economic Models

A primary reason for economists to be aware of facts regarding wealth inequality is that many standard economic models make strong predictions about agents' wealth growth. This includes both static models of cross-sectional heterogeneity among agents as well as dynamic models which explicitly address the evolution of the wealth distribution. By documenting new facts regarding household wealth dynamics, my empirical results serve as informative benchmarks against which to evaluate many economic models. In this section, I discuss several classes of models and their implied moments of wealth inequality. I explain why representative agent models which ignore heterogeneity and market incompleteness produce predictions inconsistent with the data. Finally, I outline a model that can jointly address many of my empirical facts and discuss the implications of the model for asset prices.

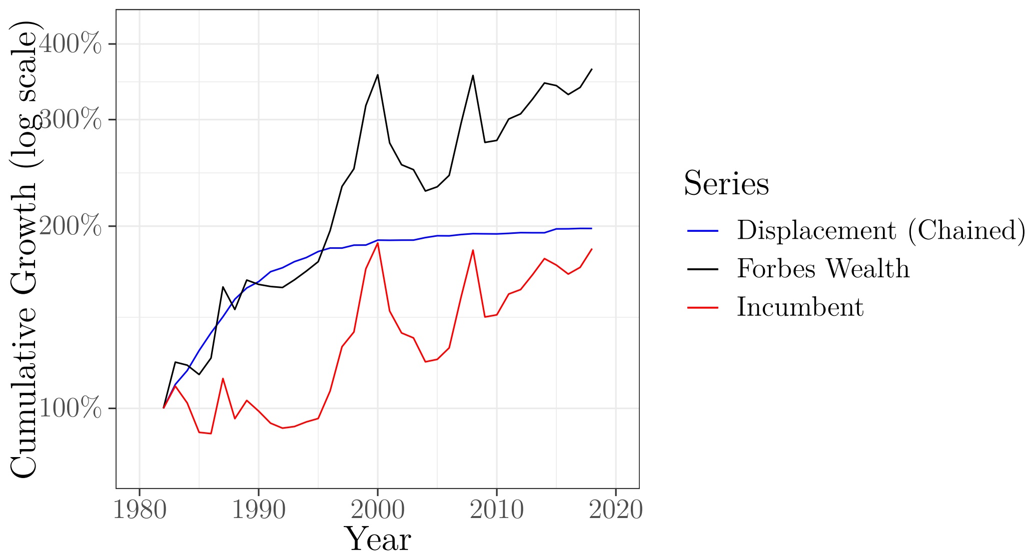


Figure 1.12: Decomposition of Wealth Inequality, 1982–2018. I plot the chained one year estimates of each Incumbent Cohort (Red). I plot the cumulative wealth growth of the Forbes 400 (Black), and the implied contribution of Displacement (Blue). Growth rates of incumbent cohort wealth and top wealth are deflated by the growth of aggregate wealth and should be interpreted as growth rates of incumbent cohort and top wealth shares.

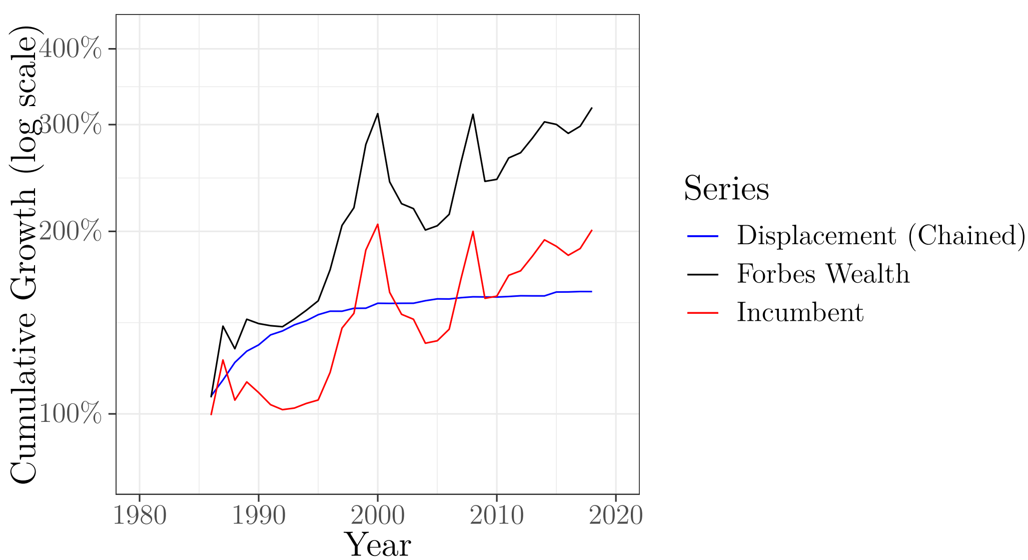


Figure 1.13: Decomposition of Wealth Inequality, 1986–2018. I plot the chained one year estimates of each Incumbent Cohort (Red). I plot the cumulative wealth growth of the Forbes 400 (Black), and the implied contribution of Displacement (Blue). Growth rates of incumbent cohort wealth and top wealth are deflated by the growth of aggregate wealth and should be interpreted as growth rates of incumbent cohort and top wealth shares.

Additional Dimensions of Heterogeneity A limitation of my methodology is that I cannot identify differences in dispersion across cohorts. By aggregating at the cohort level, idiosyncratic shocks are diversified. While my panel is constructed at the household level, estimating dispersion based on imputed estimates of wealth leads to low statistical power tests of heterogeneous dispersion. At the same time, differences in dispersion are distinct from differences in growth rates and do not affect my decomposition of inequality into the within and displacement terms.

Models of the Wealth Distribution The single asset, representative agent model is a work horse model in macro-finance. In this model, agents face inter-temporal investment and savings decisions and trade in time-zero complete markets to hedge future consumption risk. A robust prediction of these models is that post-trade consumption and wealth growth are equalized across all agents. As preferences are typically assumed to be homothetic, there are no wealth effects and aggregate wealth in the economy is a sufficient statistic for the wealth distribution. This means that the representative agent model is consistent with any observed wealth distribution. The challenge for these models is that, after agents trade and equalize wealth growth, the scaled wealth distribution is constant over time. Thus, these models are inconsistent with the rise of wealth inequality. Consistent with the model's prediction of a constant scaled wealth distribution, the models also predict no displacement in the ranks of top household wealth.[4] A household's rank in the wealth distribution at time t is identical to their rank in the wealth distribution at time $t+1$. Thus, the model is able to rationalize neither increasing wealth inequality nor the observed level of displacement.

An extension of the representative agent model that is able to rationalize time-varying wealth inequality is the introduction of heterogeneous agents. When differences between agents lead to differences in investment and saving decisions, aggregate wealth is no longer a sufficient statistic for the wealth distribution. In contrast to the representative agent model, in which today's wealthy households are identical to those of yesterday and even yesteryear, the heterogeneous agent model features churn in the wealth distribution. Today's wealthy

[4]A small amount of displacement can be attributed to death and demography.

are a mixture of those who were born wealthy and those who held high growth rate portfolios. Thus, a heterogeneous agent model can rationalize increasing wealth inequality as the result of heterogeneous growth rates of wealth across agents.

The challenge for the heterogeneous agent model is to rationalize relatively low growth rates of wealth for ex-ante wealthy households will also rationalizing increasing wealth inequality. The puzzle is explaining why wealth inequality increases (decreases) over time if wealthy households are those with lower (higher) average growth rates of wealth? In my empirical results, I find that Forbes 400 households have wealth growth rates similar to aggregate household wealth. While these ex-ante wealthy households do outgrow aggregate wealth slightly, the growth of these incumbents can only explain 20 percent of the rise in wealth inequality. The heterogeneous agent model would predict incumbent growth drives changes in wealth inequality.

I find that displacement is responsible for 80 percent of the rise in wealth inequality. Incumbent wealth households have continued growing their wealth, but have been displaced in the top wealth percentile by new households entering. Furthermore, in the data, these households growth rates are not relatively high after entering the top wealth percentile. These facts suggests are consistent with a heterogeneous agent model in which changes in the distribution of new household wealth, rather than cross-sectional differences in investment and savings decisions, drives increases in wealth inequality.

Model Selection The motivation for selecting models on the basis of their ability to simultaneously match wealthy agents' wealth dynamics and the aggregate wealth distribution is that wealthy agents hold a large fraction of the wealth in the economy and are a likely candidate for marginal agents who impact prices. This has two sets of implications useful for selecting asset pricing models. The first is in rationalizing observed prices in financial markets. The second is in rationalizing realized wealth dynamics.

Any arbitrage-free model of asset prices features a stochastic discount factor that correctly prices all traded assets. Equivalently, prices are considered "fair" by all marginal agents in the economy. Thus, observed asset prices should be consistent with the stochastic discount factor of wealthy households.

Furthermore, the wealth dynamics of wealth agents should be interpreted as the equilibrium decisions of a marginal economic agent. Qualitatively, these wealth dynamics do not look like the dynamics of a passive index investor who loads on the market. Wealthy households' wealth dynamics feature idiosyncratic dispersion and heterogeneous growth rates. Models that predict investment behavior inconsistent with observed wealth dynamics are thus likely to be mis-specified.

There is an additional, practical, consideration that makes wealth dynamics a desirable diagnostic tool. For top wealthy households, wealth dynamics are almost identical to returns on their investment portfolio. It is well known that the income distribution features a thinner tail than the wealth distribution. Wealth dynamics for top wealthy households are driven by their portfolios, not their incomes.

Models of Firm Dynamics and Ownership What kinds of assets can explain these wealth dynamics? A model featuring concentrated firm ownership is a parsimonious model of wealthy households portfolio holdings that can rationalize the observed wealth dynamics and also the large role of displacement in the rise of wealth inequality. Increasingly skewed firm size distributions have been discussed in Hartman-Glaser, Lustig, and Xiaolan (2017) and Autor et al. (2017) and offer an explanation for the economic mechanism explaining how new households can accumulate significant wealth in a short period of time. My observed life cycle effects across wealthy cohorts mirror those posed in Luttmer (2007) as an explanation for the observed size distribution of firms. Surviving firms gradually decline in growth rates over time. Persistent firm percentage ownership and a constant dividend-yield are sufficient conditions for firm dynamics to drive wealth dynamics. This is distinct from the model of Kogan, Papanikolaou, and Stoffman (2013), which features a skewed distribution of innovation and displacement. In that model, firms differ in their growth rates, but incumbent investors are diversified and thus there is no cross-sectional heterogeneity in wealth dynamics.

1.5 Conclusion

I present a model relating wealth inequality and asset prices. In the model, the rise in wealth inequality, coupled with the decline in interest rates, points to increased displacement as the primary driver of increasing wealth inequality. This is consistent with my empirical results, in which I find that over 80 percent of the rise in wealth inequality is driven by the entry of new wealthy households displacing incumbents. This speaks to the importance of "new money" in understanding the rapid rise of wealth inequality in the United States. At the same time, I find that the relative importance of displacement is smaller at shorter time horizons. I show that this can be explained by heterogeneous growth rates across cohorts. I find evidence that growth rates differ across cohorts and can be explained by life cycle effects wherein older cohorts' wealth accumulates at a slower rate than newer cohorts' wealth.

My findings have significant implications beyond understanding the rise in wealth inequality. Wealthy households are a likely candidate to be marginal in financial markets, and understanding their portfolio decisions and realized wealth dynamics offer a powerful tool for model selection. I explain that my empirical results cannot be rationalized by standard macro-finance models featuring a representative agent and complete markets. Models incorporating heterogeneous portfolio holdings and idiosyncratic firm dynamics as in Piketty, Saez, and Zucman (2017) are a promising direction. More generally, asset pricing models ought to incorporate the impact of entry of new agents and investment opportunities that cannot be invested in by incumbent agents. Finally, returns and individual wealth dynamics are linked by the portfolio decisions of households. Understanding these dynamics and the portfolio problem faced by wealthy agents are important directions for future research and offers the potential to combine insights from household finance, asset pricing, and macroeconomics.

1.6 Appendix: Transition Dynamics

Starting from a steady state featuring a relatively low level of κ, upon a regime change to a higher level of κ, the interest rate *falls*. The higher value of κ also implies a decrease in the growth rate of incumbent firms μ^L. This can be interpreted as a relative increase in the importance of displacement for economic growth. Absent a drop in the interest rate r, a decrease in μ^L reduces the value of all existing firms in the economy. The high-type agents are constrained and unable to consume more, while the low-type agents have now received a negative wealth shock due to the decrease in μ^L, which has the effect of reducing their consumption. Thus, the interest rate must drop in order to clear the consumption market. Increased displacement leads to higher wealth inequality and lower interest rates. Symmetrically, an increase in μ^L and a decrease in κ leads to *higher* interest rates. When wealth inequality is the result of high rates of return, the interest rate rises to induce the low type agents to continue lending to high type agents.

Aggregate dividends and output are deterministic following the regime change, enabling me to fully characterize the transition path $\{r_t\}$. Figure 1.5 plots the decline in interest rates r_t following an increase in κ and a decline in μ^L that keeps the long-run growth rate of the economy g constant. Following an increase in κ and a decrease in μ_L, the interest rate r_t experiences an immediate discontinuous drop, following by a protracted smooth decline to the steady interest rate under the new parameters.

All agents in the economy know the future path of interest rates, which implies a time varying price-dividend ratio p for the low type firms

$$\int_0^\infty \exp\left\{\int_0^s \left(r_{t+u} - \mu^L\right) du\right\} ds$$

This implies a no-arbitrage condition relating the current interest rate and price-dividend ratio to tomorrow's price dividend ratio

$$dp_t = \left(\left(r_t - \mu^L\right) p_t - 1\right) dt \tag{1.20}$$

Equation (1.20) states that the net return on a low type firm is equal to the dividend flow, plus capital gains accrued by virtue of dividend growth, plus capital gains accrued via changes in valuation ratios.

Solving for the transition path is done via a shooting method procedure, which I describe below. The economy begins in steady state with output $Y_0 = 1$ and a pool of outstanding loans L_0. A guess of the price-dividend ratio following the regime change, p_0 implies both the wealth of low type agents W_0^L and the consumption of low type agents under log preferences. The market clearing condition, restated below,

$$\epsilon_t x_t + (\rho + \delta) \frac{W_t^L}{Y_t} = 1$$

implies a consumption-income ratio ϵ_t for the high type agents.

Here, I make the assumption that the high type agents remain constrained following the transition path. Intuitively, this will always be the case following an increase in displacement κ, as the motives for smoothing consumption are made stronger by the lower value of μ_L. In this situation, the borrowing constraint implies that the consumption ratio is linear in the interest rate r_t

$$\epsilon_t = 1 + \alpha \left(r_t + \lambda - \mu^H \right)$$

and can be solved for r_0. Thus, the full economy can be characterized at time 0 just after the regime change. I then use Equations (1.20) and (1.13) to calculate next period's price-dividend ratio and outstanding loans, respectively. For the appropriate choice of p_0, this economy converges to the steady state economy under the new regime, and thus the asymptotic interest rate implied by the choice of p_0 must equal the steady state interest rate $r^{*,\text{new}}$.

1.7 Appendix: Proofs

Solution to the High Type Agents' Problem The problem of a constrained agent can be converted into an unconstrained problem by attaching Lagrange multipliers $\lambda, \xi_t \geq 0$ to obtain

$$\mathcal{L} = \mathbb{E}\left[\int_0^\tau e^{-\rho t} u\left(c_t\right) dt + e^{-\rho\tau} V_L\left(\frac{y_\tau}{r-\mu_L} - \int_0^\tau e^{-r(s-\tau)}\left(c_s - y_s\right) ds\right)\right]$$
$$+ \lambda \mathbb{E}\left[\alpha y_0 + \int_0^\tau e^{-rs}\left(y_s - c_s\right) ds\right] + \mathbb{E}\left[\int_0^\tau \xi_t \left(\alpha y_0 + \int_0^t e^{-rs}\left(y_s - c_s\right) ds\right) dt\right] \quad (1.21)$$

I claim that the optimal consumption process takes either the form

$$c_t = \begin{cases} \rho W_t, & t < \tau, \\ \left(\rho + \delta\right) W_t, & t \geq \tau \end{cases}$$

in the case that α is sufficiently large so that the constraint is not binding; or the form

$$c_t = \begin{cases} \epsilon y_t, & t < \tau, \\ \left(\rho + \delta\right) W_t, & t \geq \tau \end{cases}$$

in the case that the constraint binds. The agent's wealth W_t is given by

$$W_t = \int_0^t \left(y_s - c_s\right) e^{r(t-s)} ds + \frac{y_t}{r-\mu_L}\left(1 + \mathbb{I}_{\mu_H}\left(\mu\right) \frac{\mu_H - \mu_L}{r + \lambda - \mu_H}\right)$$

and the marginal propensity to consume out of income is given by

$$\epsilon = 1 + \alpha\left(r + \lambda - \mu_H\right)$$

It is straightforward to show that the agent, upon decaying to the low growth state, consumes a constant fraction of wealth $\rho+\delta$. In the event that α is sufficiently large that the borrowing constraint does not bind, log preferences and i.i.d dividend growth imply that the agent will consume fraction ρ of her wealth, which follows the process

$$dW = \left(y - \rho W + \mu_H P^H\left(y\right) + r\left(W - P^H\left(y\right)\right)\right) dt + \left(P^L\left(y\right) - P^H\left(y\right)\right) dN.$$

Let $\{c_t\}$ be the optimal consumption process. If c_t prescribes that the borrowing constraint is tight until stopping time τ, then we have that

$$\alpha y_t = (y_t - c_t)\, dt + (1 - (r + \lambda)\, dt)\, (\alpha\, (y_t + dy_t))$$

Substituting in the definition of dy and dropping higher order dt terms gives

$$\frac{c_t}{y_t} = 1 + \alpha\, (r + \lambda - \mu_H)\, .$$

It remains to be shown that the optimal c_t process keeps the agent at the borrowing constraint. The proof that this is optimal proceeds by contradiction. Assume that an agent who follows c_t expects her borrowing constraint to be slack over some interval of time $\mathcal{T} = (t', t' + \Delta t)$. This implies that her Euler equation holds with equality. Under the assumption of log preferences, it must be that both her wealth and consumption growth during period $\mathcal{T}$ are equal to $r - \rho$. Plugging into the dynamic budget constraint gives

$$(r - \rho)\, W = y - \rho W + rW + (\mu_H - r)\, P^H\, (y)$$

This simplifies to

$$rP^H\, (y) = y + \mu_H P^H\, (y)$$

which forms a contradiction, given that y is positive and $\mu_H > r$. Thus there are no such intervals $\mathcal{T}$ and the constraint is binding almost surely. As shown above, under a binding constraint it is optimal to consume ϵy_t, completing the proof.

1.8 Appendix: Data Sources

The initial construction of my panel begins with the Forbes 400 data set. Forbes Magazine publishes a list of the wealthiest 400 Americans. The list is compiled by dedicated staff using a mix of public and private information. The first list was compiled in 1982, and has since been updated annually. By starting with Forbes 400 lists, I have a number of repeated observations for the same individual over many years. Perhaps unsurprisingly, the Forbes 400 list exhibits substantial persistence. From 1982 to 2018, Forbes Magazine published 37 lists of the 400 wealthiest Americans. There could be as many as 14,800 unique names published across those lists. However, the actual Forbes 400 lists feature less than 1,600 unique individuals, corresponding to an average attrition rate of just over 10 percentage points per annum. Equivalently, the average tenure on the Forbes 400 list is roughly 10 years. The data collection challenge of this paper is to fill in wealth observations missing in the Forbes 400 lists.

In order to account for dropouts from the Forbes 400, I employ a number of data sources. As these data sources are unfamiliar to the typical reader, I first enumerate the data sets before discussing each at length below. The data to be described are:

1. Forbes Dropoff List: Annual wealth estimates for displaced Forbes 400 members

2. Forbes Billionaire List: Annual wealth estimates for billionaires

3. Family Structures for Forbes 400 members

4. LexisNexis Property Records for family of Forbes 400 members

5. Wealth-X profiles for individuals exceeding $30 million net worth

Forbes Dropoff Lists The first auxiliary data set is Forbes Magazine's own published list of drop offs, beginning in 2012. For all subsequent Forbes 400 lists, Forbes Magazine reported the wealth of individuals who were removed from the list on the grounds that they were no longer among the 400 richest Americans. I manually collect these reports from archives of Forbes' website. Starting from the 14,800 observations in the Forbes 400, the published dropoff lists add an additional 175 observations. These observations are useful in that they are relatively simple to collect and match by name. The weaknesses of this data set are that: (i) it only exists since 2012, (ii) it only contains wealth for dropoffs in the year immediately following their exit from the Forbes 400 list, and (iii) it does not report wealth for deceased individuals. For the purposes of estimating long run trends in top wealth shares, such dropoff data is of limited use. Nevertheless, I present it first because it is the "cleanest" measure of wealth for dropoffs. The wealth estimates are compiled by the same Forbes Magazine staff that publish the main Forbes 400 lists, and thus the methodology for estimating the wealth of these individuals is likely to be consistent. The wealth estimates also feature no selection-bias at the one year horizon, in that all surviving dropoffs have their wealth reported.

Forbes Billionaire Lists The second auxiliary data set is Forbes Magazine's published list of world-wide billionaires. This list was first compiled in 1996, and continues to this day. The cutoff for inclusion in the Forbes 400, which I infer from the wealth of the lowest-ranked member in each annual list, has exceeded $1 billion since 2006.[5] Therefore, for many individuals who dropped off the Forbes 400 post-2006, the magazine staff continues to use a similar methodology to estimate their wealth. I scraped the historical Forbes Billionaire lists from archives of Forbes' website. Importantly, individuals who fall off the Forbes 400 list, but who remain billionaires, stay in the Forbes Billionaire data set. This is the case for a number of individuals, and I am able to combine the data sets to get a balanced panel of wealth for these individuals extending through to 2018. It would be impossible to do this using only the Forbes Dropoff data set for the simple reason that the wealth of dropoffs is only reported for a single year.

Another advantage of the Forbes Billionaire list is that it assists me in estimating the wealth of deceased Forbes 400 individuals. Given my focus on long term trends, my unit of analysis, wherever possible, is the family of a Forbes 400 member.[6] For a number of deceased Forbes 400 individuals, a family member continues to remain on the Forbes 400 list. This is the case, for example, for Dagmar Dolby, the widow of Ray Dolby. Even though Ray Dolby passed away in 2013, Dagmar Dolby survives to this day and continues to be on the Forbes 400. In 2012, the year immediately preceding his death, Roy Dolby was estimated to have a net worth of $2.4 billion. In 2013, the year Dagmar Dolby first appeared on the Forbes 400,

[5]The one exception was the cutoff of $950 million in 2009.

[6]Specifically, I include spouses, ex-spouses, children, and step-children.

her wealth was estimated to be, again, $2.4 billion. In other cases, a Forbes 400 member has numerous family members who divide up their wealth, but who nonetheless appear on the Forbes 400 list and for whom the *total* wealth is of similar magnitude to the wealth of the single original family member. This is the case for the Cargill sisters, consisting of Alexandra Daitch, Sarah MacMillan, Lucy Stitzer, and Katherine Tanner, who were the four daughters of W. Duncan MacMillan, who died in 2006. While these cases are relatively easy to identify and account for in the Forbes 400, the Forbes Billionaire data set allows me to identify those cases where the surviving family members are found across the two data sets. Roughly 700 additional observations of family unit wealth are obtained by joining together the Forbes 400 and Forbes Billionaire lists.

As I will elaborate upon later, conducting analysis at the family unit can have a drastic impact on conclusions regarding long terms wealth trends. As a simple example, the 2018 Forbes 400 list features 25 individuals who were on the inaugural 1982 Forbes 400 list, and a total of 68 individuals who first entered the ranks of the Forbes 400 prior to 1990. If, instead, one considers the inaugural year of the family unit, these numbers increase substantially. Eighty-two members of the 2018 Forbes 400 are members of families who were on the inaugural 1982 Forbes 400 list, more than three times the previous number. A total of 130 individuals are members of families that first entered the ranks of the Forbes 400 prior to 1990. This is all despite the fact that, across the 1,580 distinct members of the Forbes 400, there are 1,373 distinct family units. While this is merely a suggestive feature of the data, there are also methodological reasons to conduct analysis at the level of the family unit. For the purpose of understanding long term trends in top wealth shares and top wealth inequality, inter-generational transfers become increasingly important as one extends the time horizon.

Family Structures for Forbes 400 members In order to identify family members, I manually collect data on the names and, where possible, age and location of children and spouses of Forbes 400 individuals. Consistent with Bernstein and Swan (2008), I find that the average Forbes 400 individual has three children. The identification of family members of Forbes 400 individuals is a non-trivial task. While, in recent years, Forbes Magazine attempts to report the marital status for each member, along with the number of children they have, this number is often inaccurate. Common reasons are that the number provided is the number of surviving children, or that the number excludes numerous step-children. Taking Forbes Magazines' estimate as a starting point, I hand collect data on the number and the names of children using a variety of internet data sources. For deceased Forbes 400 members, their obituaries often contain information on surviving family members. Even for surviving individuals, or individuals for whom I could not locate an obituary, it is possible to obtain names of family members using obituaries of parents or siblings. In some cases, Forbes 400 members or their spouses have written books and included dedications to their children. This is the case for, among others, Robert and Janice Davidson, as well as David Shaw.[7] More esoteric examples include Pincus Green, whose children jointly wrote a letter to then-president Bill Clinton requesting a presidential pardon for their father. In total, I identified 4,843 children of Forbes 400 members, and found names and other information for 4,578 of those children. A detailed list of sources used in the construction of this data set is available upon request.

LexisNexis Property Records Thus far, the auxiliary sources of wealth information have relied upon wealth estimates produced by Forbes Magazine staff. In order to account for individuals not found in the Forbes data sets, due either to dropping off prior to 2006 or dropping to below $1 billion in net worth, I make use of the LexisNexis Public Records data set. LexisNexis offers a search interface through which I can observe basic biographical information, along with address history and property records, for a significant proportion of the American population. For property records, the key feature of the data set for this analysis, LexisNexis provides access to property deed records for 3,017 counties in the United States, out of a total possible 3,144. This is a coverage ratio of 96.0 percentage points. For now, I describe the characteristics of the LexisNexis data set and postpone discussion of how

[7] The Davidsons wrote *Genius Denied: How to Stop Wasting Our Brightest Young Minds*. David Shaw's wife Beth Kobliner wrote *Make Your Kid A Money Genius (Even If You're Not): A Parents' Guide for Kids 3 to 23*.

I estimate wealth using the data until Section 1.3.2. For property assessment records, which are filed more regularly, the coverage ratio is even higher, and covers all but three counties. Biographical information provided includes names of likely family members, employment history, and date of birth. All of this information is linked to an encoded version of a Social Security Number, as well as to a unique database identifier, a `LexID`. Starting with the biographical information included in the Forbes 400 lists, I search for individuals in the LexisNexis database based on name, approximate age, and state of residence. From there, I reject potential matches based on employment history and family information. Through this process, I manually link 1,565 Forbes 400 individuals to a unique `LexID`. For the less than 1 percent of Forbes 400 individuals who I am unable to link to a `LexID`, the reason is typically that the individual has no domestic residences. This is the case for, among others, Victor Fung, J Paul Getty Jr, and Tor Peterson. For each of the 1,565 Forbes 400 individuals that I am able to uniquely identify in LexisNexis, I algorithmically download all property deeds and property assessments pertaining to that individual, as well as the names and addresses of all likely family members. For each likely family member, I then algorithmically find the most likely matched `LexID` corresponding to that individual in the LexisNexis database, based on biographical information, and download all property deeds and assessments pertaining to these potential family members. The Python code I wrote to automate the extraction of information from the LexisNexis database into a format conducive to empirical analysis is available upon request.

I aggregate property records at the family unit, so that all family members' property records are grouped together. The property records contain geographic identifiers for the property in the form of street address, zoning, and parcel number, as well as some information regarding the value of that property. For property deeds, this valuation information consists of a sale value, a transaction date, names for the buyer and seller, as well as mortgage amount. For property assessments, this valuation consists of an assessed value for the stated tax year. I further process the property records data to account for duplicates and potentially mis-labeled records using two methods. First, I exclude non-apartment properties sharing identical GPS coordinates. Second, I exclude any remaining properties which feature substantially similar parcel numbers. For the bulk of my empirical analysis, I restrict attention to residential properties and exclude properties whose land usage indicates commercial zoning. In Section 1.3.2, I elaborate on the methodology used to produce a panel of wealth estimates using LexisNexis data.

Wealth-X Profiles The final non-standard data set that I use to produce my panel consists of Wealth-X profiles on ultra-wealthy individuals, defined here as individuals with net worth exceeding $30 million as of 2018. The profiles are maintained by dedicated staff employed by Wealth-X, and contain information derived from publicly disclosed transactions, holdings, philanthropy, conspicuous purchases, board memberships, professional and family ties, and other biographical information. I first extract a list of all ultra-wealthy individuals, both foreign and domestic, in the Wealth-X database. Based on this list of individuals, I then collect each profile and extract family details and portfolio holdings. Thus, my data set contains every individual Wealth-X has identified as having a net worth exceeding $30 million in 2018. For this paper, I principally focus my attention on domestic ultra-wealthy individuals, and thus discard all individuals with no business or residential addresses within the United States. I then manually match these individuals to Forbes 400 family units based on the hand-collected family structure information.

Wealth-X is a private corporation that maintains profiles on wealth individuals. While the methodology employed by Wealth-X is unlikely to be identical to that employed by Forbes magazine, the wealth estimates are highly correlated on the overlapping sample. For the population of United States billionaires, Wealth-X's reported list of billionaires slightly exceeds that of Forbes for the year 2018.[8] For the population of ultra-wealthy individuals with net worths exceeding $30 million, Wealth X reports roughly 20,000 such individuals in the United States for the year 2018. For comparison, the Survey of Consumer Finances estimated that 50,000 ultra-wealthy households, and 640 billionaire households existed in 2016. This is consistent with the characterization that Wealth-X has relatively comprehen-

[8] I attribute these discrepancies to differences in methodology and within-calendar year changes in individuals' net worth.

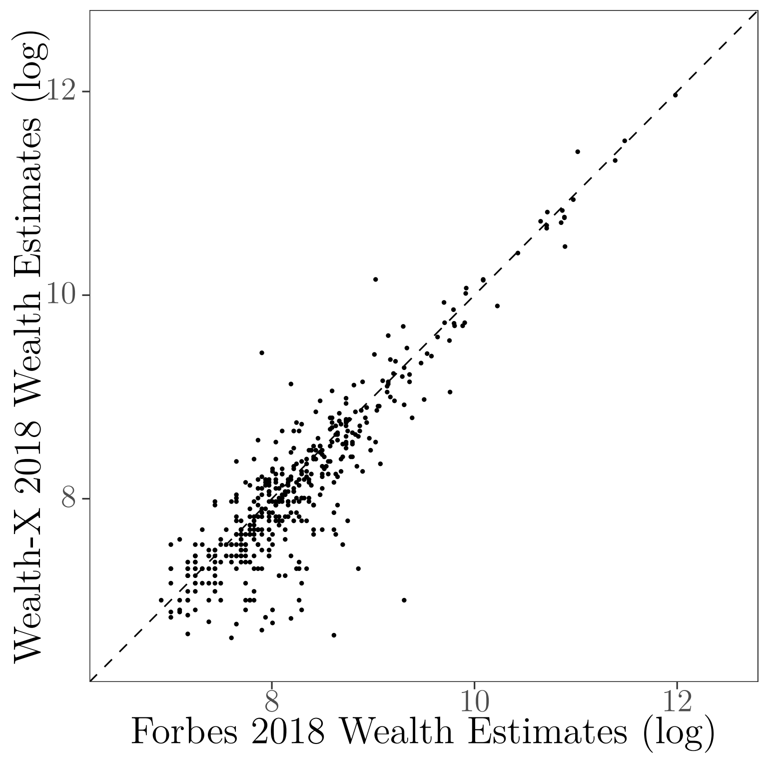


Figure 1.14: Plot of 2018 estimates of wealth from Forbes Magazine (x-axis) and Wealth-X (y-axis) for matched individuals, in logs.

sive coverage of individuals with net worths as low as \$100 million (a population numbering roughly 7,000), and a random sample of net worths between \$30 million and \$100 million, covering approximately 30 percent of that population.

One limitation of the Wealth-X database is that the portfolio holdings and valuation are as of 2018. Thus, Wealth-X data can only be used to fill in 2018 wealth levels for Forbes 400 individuals. For this reason, I use Wealth-X as a robustness check for both my hand-collected family structure data, as well as my 2018 wealth estimates for Forbes 400 dropoffs. When comparing family structure data, my dataset contains a superset of family members enumerated in Wealth-X. When comparing 2018 wealth estimates between the Forbes 400 list, my 2018 wealth panel, and Wealth-X profile estimates, I find that the estimates are highly correlated at the individual level ($\rho = 0.8$) and similar in terms of implications for aggregate quantities.

1.9 Housing Imputation

I assume that household preferences for Forbes 400 families are of the form

$$V_{it} = \log\left(C_{it}^{\psi_i} H_{it}^{\phi_i}\right) + \mathbb{E}_{it}\left[e^{-\rho_i} V_{i,t+1}\right],$$

where C_{it} denotes non-housing consumption, H denotes housing consumption, and ρ captures the subjective discount of household i. Under these assumptions, the household myopically consumes a constant proportion ρ of their wealth, of which a fraction $\phi_i / (\psi_i + \phi_i)$ consists of expenditures on housing. Abstracting from cross-sectional heterogeneity in financing, I further assume that housing consumption is simply the product of a common rental rate on housing p^H and the value of the household's residential housing stock.[9] Therefore, housing consumption and period wealth are related by

$$W_{it} = \frac{1}{\rho_i} \frac{\psi_i + \phi_i}{\phi_i} \frac{H_{it}}{p^H}.$$

Under this framework, the fraction of total wealth held in housing is constant over time for each household, and it is possible to use a subset of contemporaneous observations of housing value and total wealth to estimate unobserved total wealth from annual observations of housing wealth.

My imputation procedure based on housing values has a number of advantages. First, as discussing the Data Section, I observe portfolios of real estate for a significant fraction of Forbes 400 households, and thus the method is broadly applicable across the population of interest without need for individual- or family-specific adjustments. Second, the wealth estimates are timely and likely reflect household's current level of wealth. There is significant turnover in real estate portfolios, as Forbes 400 households buy and sell properties often. I observe transaction values for these properties, and am able to exclude transactions between related parties using both the buyer and seller names, linking to the family structure data

[9] In general, households in my sample employ little leverage in their home purchases. Among potential explanations, I am most sympathetic to the idea that these households self-finance so as to avoid paying spreads to financial intermediaries.

I collected, as well as transaction-level identifiers for intra-family transfers provided by the LexisNexis data set. Thus, my estimates of housing value are market-transaction based, addressing concerns that my estimates of wealth growth are reflecting passive capital gains on a static housing portfolio. Finally, the wealth estimates are based on changes in housing portfolios, rather than levels. I do not require that all households have the same preference parameters ψ_i, ϕ_i, and ρ_i. My identifying assumption is that the proportion of housing to total wealth at the household level remain constant over time. In Appendix 1.10, I discuss another advantage of this specification: robustness to potential household-level heterogeneity in the use of shell corporations to obfuscate home ownership.

A weakness of my method of imputing wealth from the observed real estate panel is that I am assuming a time-invariant relationship between wealth and real estate. While this is a quantitatively reasonable assumption in aggregated data, it abstracts from the underlying portfolio problem faced by the household. In particular, my results cannot speak to rate at which households adjust their real estate holdings in response to changing net worth. A hypothetical process in which a random fraction of households adjust their holdings each year, analogous to a Calvo model of prices, would produce identical results in aggregate real estate holdings. In my panel, this assumption manifests in that my estimates of the wealth of Forbes 400 dropoffs is typically too high at the one year horizon when compared to the available estimates published in Forbes, as dropoff households do not all adjust their real estate values immediately upon falling off the Forbes 400 list. This is one reason that I focus on relatively large cohorts of households and compute growth rates at long horizons.

1.10 Appendix: Robustness

Identifying Wealthy Households My estimated growth rates are based on a panel of ex-ante wealthy individuals. For the purposes of decomposing the growth of the top wealth share into the contribution of incumbents and entrants, it is not essential that incumbents are defined as the 400 richest households. The empirical strategy is to identify a population of ex-ante wealthy households and estimate the dynamics of their wealth. The advantage

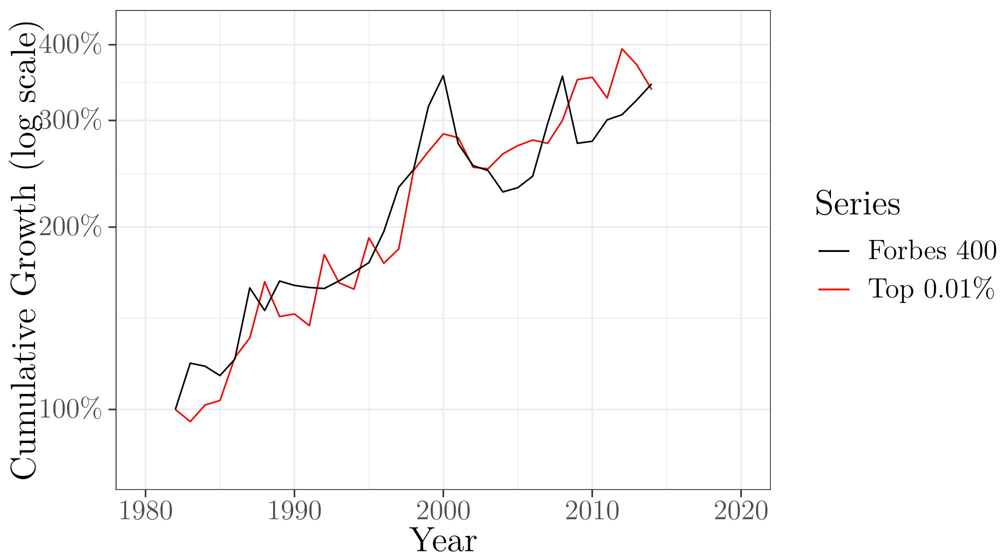


Figure 1.15: Two measures of Top Wealth Inequality, 1982–2014. I plot the cumulative growth rate of the Forbes 400 (Black) and Top 0.01% (Red) Wealth Shares. The Forbes 400 Wealth Share is calculated from the Forbes 400 lists published by Forbes Magazine and is available since 1982. The Top Wealth Share data is from Piketty, Saez, and Zucman (2017) and is available through 2014.

of using Forbes is that they are considered to be the *wealthiest* households, and the relative wealth of biographical information about these families enables me to match Forbes 400 households to real estate holdings via the LexisNexis data set. The specific choice to focus on increases in the Forbes 400 wealth share as opposed to other measures of top wealth inequality is innocuous. Figure 1.15 plots my series for the cumulative increase in the Forbes 400 wealth share against the estimates of the Top 0.01% wealth share from Piketty, Saez, and Zucman (2017). The two measures are very similar and have virtually identical implications for the long term increase in wealth inequality.

Validating Wealth Estimates Both the construction of my panel and the bulk of my empirical results rely heavily upon the estimates of wealth published by Forbes Magazine. Over the sample period, the rise in total Forbes 400 wealth has been consistent with the rise in the wealth share of the top 0.01% of households. This serves as validation for the implications drawn from Forbes estimates regarding relative wealth shares and wealth inequality. However, this does not address the potential for individual-level measurement error in Forbes

Magazine's wealth estimates. While I cannot validate historical individual wealth estimates published by Forbes, I am able to compare contemporaneous wealth estimates published by Forbes Magazine and Wealth-X. As seen in Figure 1.14, there is a high level of agreement between the estimates produced by Forbes Magazine and those produced by Wealth-X. Regressing one source of wealth estimates on the other produces both a high R-squared of 0.64, corresponding to a pairwise correlation of 0.8, and an unbiased coefficient close to one.

Intra-year Wealth Estimates Forbes Magazines publishes the Forbes 400 list and Forbes Billionaire lists each year, but releases these lists at different points in the year. The Forbes 400 list is typically published in the fall, while the Forbes Billionaire list is published in the spring. The wealth estimates from those data sets are current as of publication, and the discrepancy in publication timings can potentially introduce issues when joining together the data sets into a single, larger panel. The Forbes Dropoff lists, available post-2011, are published alongside the Forbes 400 list in the fall. Wherever possible, I defer to Forbes Dropoff list wealth estimates over Forbes Billionaire list estimates in the same calendar year. The real estate value estimates from LexisNexis are not tied to a given month, and likely correspond to the transaction or assessment date, depending on the exact source of the valuation.

Given my focus on long-run growth rates, these small intra-year timing differences are not instrumental to my results, and so I largely ignore timing discrepancies when joining the various data sets. In an effort to make the market-residualized wealth estimates as accurate as possible, I use July through June market factor returns in my empirical analysis. This is another motivation for using year fixed effects, rather than directly including the market factor, in several of my empirical specifications.

Imputation of Household Wealth In the construction of my panel, I use households' real estate holdings to impute wealth observations. In my primary specification, I assume a unit elasticity between housing wealth and total wealth. This is equivalent to a constant portfolio share of residential housing. In Table [INCOMPLETE], I present regression evidence that is consistent with the assumption of unit elasticity. For each cohort, I regress total real estate holdings of surviving cohort members against total wealth of surviving cohort members, where surviving members are defined as those who still appear on the Forbes 400 list. Because housing portfolios are persistent, I run the regression in first differences. The coefficient of the regression is economically close to one at the one, two, five, and ten year horizons. Furthermore, I find that the explanatory power of the regressions increases with horizon. The increased explanatory power at longer horizons can be explained by short-run adjustment costs in household portfolios. Results are quantitatively similar when I re-estimate my growth rates using the empirical elasticity, rather than my assumption of a unit elasticity.

I also investigate the sensitivity of my results to different ways of measuring real estate value. In my primary specification, I use the most recent purchase or sale price associated with the property. Where no deed transfer data is available, I rely on annual property value assessments. In the latter case, for years in which no property assessment is reported, I use the most recent property value assessment. Results are quantitatively unchanged when I inflate / deflate real estate values using five-digit zip code specific House Price Indices.

Measurement of Aggregate Wealth The estimates presented compute growth of aggregate wealth using the net worth of U.S. households. This corresponds to item 35 in Table B.1 of the Financial Accounts of the United States. This series differs from U.S. net wealth presented in line 1 of Table B.1 despite capturing the same conceptual quantity, aggregate wealth. As discussed in Holmquist and McIntosh (2015), the discrepancy between the series arises due to differences in the treatment of government non-financial assets, such as defined benefit pension plan entitlements. Because U.S. net wealth ignores these non-financial assets, it produces a downwards biased estimate of aggregate wealth. As of the fourth quarter of 2018, Household net wealth is roughly 12 percentage points greater than U.S. net wealth. However, the discrepancy between the two series was less than 1 percentage point at the start of the sample period, reflecting the increasing importance of non-financial assets in the calculation of aggregate wealth. As the total wealth of the Forbes 400 is measurement independently of the Federal Reserves' estimates of aggregate wealth, using an increasingly downwards biased estimate of aggregate wealth would overstate the rise in wealth inequality over the sample period. For both these reasons, I present results calculated using Household net worth as the measure of aggregate wealth. At long horizons, this results in an estimate of the annual growth rate of aggregate wealth of 5.7 percentage points compared to a growth rate of 5.3 percentage points when using U.S. net wealth. The choice of measure for aggregate wealth does not drive my results, either qualitatively or quantitatively. Taking the estimate of 5.3 percentage points as the estimate of the annual growth rate of average wealth leads to me to attribute 75 percent of the increase in top wealth inequality to displacement compared to my preferred estimate of 82 percent.

Properties held in Tax Shelters It is certainly true that wealthy households do not hold all their real estate under their own name. The most convincing evidence for this is the fact that I do *not* observe property ownership for every Forbes 400 family. At the same time, I can reasonably assume that virtually every Forbes 400 family owns at least one home. One source of these omissions is that these homes may be owned by limited liability corporations. A case in which I can verify this is Mark Zuckerberg, the founder of Facebook. His primary address is reported in numerous articles online, and I am able to link him to this primary address in LexisNexis' data set. What is missing from LexisNexis, and from my data set, is proof that he owns this property. The deeds for this property are linked to an limited liability corporation which cannot be linked back to Mr. Zuckerberg, and thus I do not observe his housing portfolio. This can be modeled as the following

$$H_{it} = \kappa_i H^*_{it}.$$

For a given household i, I observe fraction κ of their total house value H^* in LexisNexis. For a small proportion of households, such as that of Mr. Zuckerberg, $\kappa = 0$, and thus I cannot estimate his total wealth using my methodology. However, given that I do observe some housing, corresponding to the case that $\kappa > 0$, my methodology is unbiased so long as κ remains constant over time at the household level. In imputing these observations, I am assuming that households do not engage in increased usage of obfuscatory methods as a function of wealth, cohort age, or time.

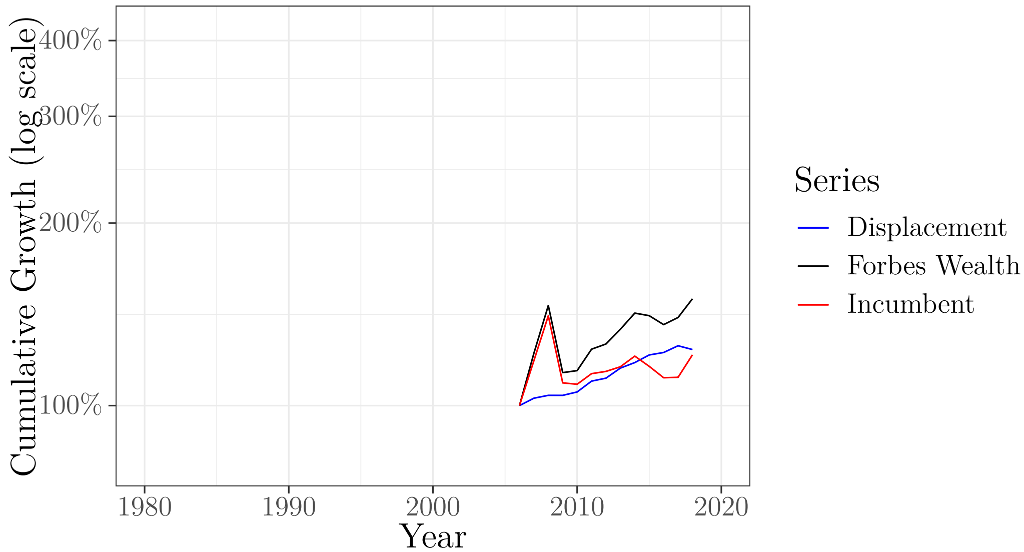


Figure 1.16: Decomposition of Wealth Inequality, 2006–2018. I plot the cumulative wealth growth of the 2006 Incumbent Cohort (Red), the Forbes 400 (Black), and the implied contribution of Displacement (Blue). Growth rates of incumbent cohort wealth and top wealth are deflated by the growth of aggregate wealth and should be interpreted as growth rates of incumbent cohort and top wealth shares.

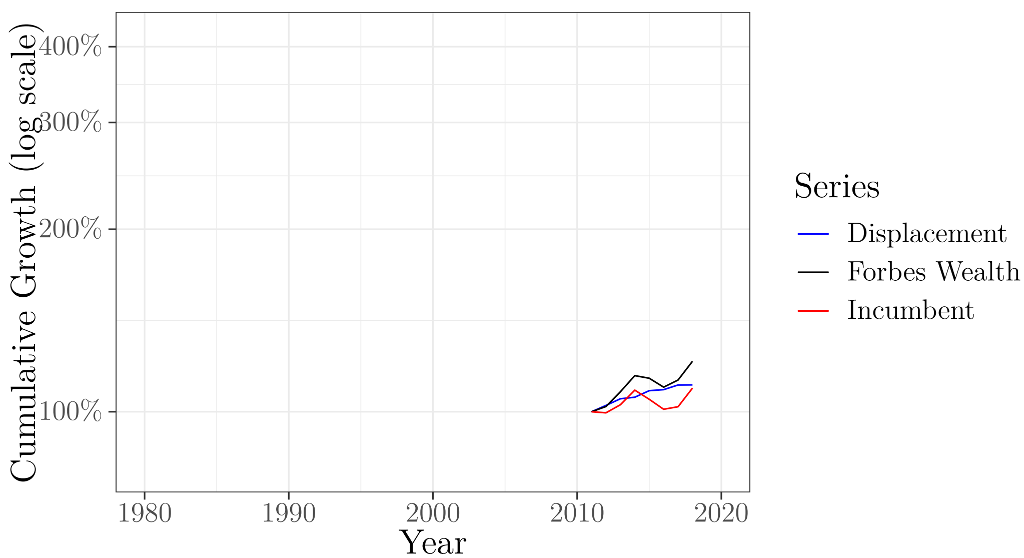


Figure 1.17: Decomposition of Wealth Inequality, 2011–2018. I plot the cumulative wealth growth of the 2011 Incumbent Cohort (Red), the Forbes 400 (Black), and the implied contribution of Displacement (Blue). Growth rates of incumbent cohort wealth and top wealth are deflated by the growth of aggregate wealth and should be interpreted as growth rates of incumbent cohort and top wealth shares.

1.11 Bibliography

Anderson, Ronald C. and David M. Reeb (2003). "Founding-Family Ownership and Firm Performance: Evidence from the S&P 500". *The Journal of Finance* 58.3, pp. 1301–1328. DOI: 10.1111/1540-6261.00567.

Autor, David et al. (May 2017). *The Fall of the Labor Share and the Rise of Superstar Firms*. Working Paper 23396. National Bureau of Economic Research. DOI: 10.3386/w23396.

Barone, Guglielmo and Sauro Mocetti (2016). "Intergenerational mobility in the very long run: Florence 1427-2011". *Bank of Italy Temi di Discussione (Working Paper) No* 1060.

Benhabib, Jess, Alberto Bisin, and Mi Luo (2015). *Wealth distribution and social mobility in the US: A quantitative approach*. Tech. rep. National Bureau of Economic Research.

Bennedsen, Morten et al. (May 2007). "Inside the Family Firm: The Role of Families in Succession Decisions and Performance*". *The Quarterly Journal of Economics* 122.2, pp. 647–691. ISSN: 0033-5533. DOI: 10.1162/qjec.122.2.647.

Bernstein, P.W. and A. Swan (2008). *All the Money in the World*. Knopf Doubleday Publishing Group. ISBN: 9780307267702.

Bertrand, Marianne and Antoinette Schoar (June 2006). "The Role of Family in Family Firms". *Journal of Economic Perspectives* 20.2, pp. 73–96. DOI: 10.1257/jep.20.2.73.

Blanchard, Olivier J. (1985). "Debt, Deficits, and Finite Horizons". *Journal of Political Economy* 93.2, pp. 223–247. ISSN: 00223808, 1537534X.

Calvet, Laurent E, Laurent Bach, and Paolo Sodini (Dec. 2015). *Rich Pickings? Risk, Return, and Skill in the Portfolios of the Wealthy*. HEC Research Papers Series 1126. HEC Paris.

Campbell, John Y, Tarun Ramadorai, and Benjamin Ranish (2018). *Do the rich get richer in the stock market? Evidence from India*. Tech. rep. National Bureau of Economic Research.

Civale, Simone, Luis Diez-Catalan, and Serio Salgado (2017). *Wealth Dynamics of the Super-Rich*. Working Paper. University of Minnesota.

Clark, Gregory and Neil Cummins (2013). *Surnames and social mobility: England 1230-2012*. Economic History Working Papers. London School of Economics and Political Science, Department of Economic History.

Constantinides, George M and Darrell Duffie (1996). “Asset pricing with heterogeneous consumers”. *Journal of Political economy* 104.2, pp. 219–240.

Cronqvist, Henrik, Anil K Makhija, and Scott E. Yonker (2012). “Behavioral consistency in corporate finance: CEO personal and corporate leverage”. *Journal of financial economics* 103.1, pp. 20–40.

Davidson, J., B. Davidson, and L. Vanderkam (2007). *Genius Denied: How to Stop Wasting Our Brightest Young Minds*. Simon & Schuster. ISBN: 9781416595687.

Di Tella, Sebastian (2019). “Optimal regulation of financial intermediaries”. *American Economic Review* 109.1, pp. 271–313.

Fagereng, Andreas, Luigi Guiso, et al. (2016). *Heterogeneity and persistence in returns to wealth*. Tech. rep. National Bureau of Economic Research.

Fagereng, Andreas, Martin Blomhoff Holm, et al. (2019). “Saving Behavior Across the Wealth Distribution: The Importance of Capital Gains”. *Working Paper*.

Gabaix, Xavier et al. (2016). “The Dynamics of Inequality”. *Econometrica* 84.6, pp. 2071–2111. DOI: 10.3982/ECTA13569.

Garbinti, Bertrand, Jonathan Goupille-Lebret, and Thomas Piketty (2017). “Accounting for wealth inequality dynamics: Methods, estimates and simulations for France (1800-2014)”.

Gârleanu, Nicolae, Leonid Kogan, and Stavros Panageas (2012). “Displacement risk and asset returns”. *Journal of Financial Economics* 105.3, pp. 491–510.

Gârleanu, Nicolae and Stavros Panageas (2015). “Young, Old, Conservative, and Bold: The Implications of Heterogeneity and Finite Lives for Asset Pricing”. *Journal of Political Economy* 123.3, pp. 670–685. DOI: 10.1086/680996.

— (2017). *Finance in a time of disruptive growth*. Tech. rep. Working paper.

Gomez, Matthieu (2018). “Displacement and the Rise in Top Wealth Inequality”. *Working Paper*.

Gomez, Matthieu et al. (2016). “Asset prices and wealth inequality”. *Unpublished paper: Princeton*.

Guvenen, Fatih, Serdar Ozkan, and Jae Song (2014). “The nature of countercyclical income risk”. *Journal of Political Economy* 122.3, pp. 621–660.

Haddad, Valentin (2012). "Concentrated ownership and equilibrium asset prices".

Hartman-Glaser, Barney, Hanno Lustig, and Mindy Z Xiaolan (2017). "Capital share dynamics when firms insure workers". *The Journal of Finance*.

Herskovic, Bernard et al. (2016). "The common factor in idiosyncratic volatility: Quantitative asset pricing implications". *Journal of Financial Economics* 119.2, pp. 249–283.

Holmquist, Elizabeth Ball and Susan Hume McIntosh (Oct. 2015). "U.S. Net Wealth in the Financial Accounts of the United States". *FEDS Notes* 2015.1642. DOI: 10.17016/2380-7172.1642.

Hubmer, Joachim, Per Krusell, and Jr. Smith Anthony A (Dec. 2016). *The Historical Evolution of the Wealth Distribution: A Quantitative-Theoretic Investigation*. Working Paper 23011. National Bureau of Economic Research. DOI: 10.3386/w23011.

Kobliner, B. (2017). *Make Your Kid A Money Genius (Even If You're Not): A Parents' Guide for Kids 3 to 23*. Simon & Schuster. ISBN: 9781476766829.

Kogan, Leonid, Dimitris Papanikolaou, and Noah Stoffman (2013). *Winners and losers: Creative destruction and the stock market*. Tech. rep. National Bureau of Economic Research.

Kopczuk, Wojciech and Emmanuel Saez (2004). *Top wealth shares in the united states: 1916-2000: Evidence from estate tax returns*. Tech. rep. National Bureau of Economic Research.

Koudijs, Peter and Laura Salisbury (2016). "For Richer, For Poorer: Banker's Skin-in-the-game and Risk Taking in New England, 1867-1880".

Kuhn, Moritz, Moritz Schularick, and Ulrike I Steins (2017). *Income and Wealth Inequality in America, 1949-2013*. Tech. rep. CESifo Working Paper.

Luttmer, Erzo GJ (2007). "Selection, growth, and the size distribution of firms". *The Quarterly Journal of Economics* 122.3, pp. 1103–1144.

Malloy, Christopher J, Tobias J Moskowitz, and Annette Vissing-Jørgensen (2009). "Long-run stockholder consumption risk and asset returns". *The Journal of Finance* 64.6, pp. 2427–2479.

Morck, Randall, David Stangeland, and Bernard Yeung (Jan. 2000). "Inherited Wealth, Corporate Control, and Economic Growth The Canadian Disease?" *Concentrated Corporate Ownership*. University of Chicago Press, pp. 319–372.

Pérez-González, Francisco (Dec. 2006). "Inherited Control and Firm Performance". *American Economic Review* 96.5, pp. 1559–1588. DOI: 10.1257/aer.96.5.1559.

Peter, Alessandra (2019). *Owning Up: Closely Held Firms and Wealth Inequality*. Tech. rep. Stanford University.

Piketty, Thomas and Arthur Goldhammer (2014). *Capital in the Twenty-First Century*. Harvard University Press. ISBN: 9780674430006.

Piketty, Thomas, Emmanuel Saez, and Gabriel Zucman (Oct. 2017). "Distributional National Accounts: Methods and Estimates for the United States*". *The Quarterly Journal of Economics* 133.2, pp. 553–609. ISSN: 0033-5533. DOI: 10.1093/qje/qjx043.

Song, Jae et al. (2018). "Firming up inequality". *The Quarterly journal of economics* 134.1, pp. 1–50.

Villalonga, Belen and Raphael Amit (2006). "How do family ownership, control and management affect firm value?" *Journal of Financial Economics* 80.2, pp. 385–417. ISSN: 0304-405X. DOI: https://doi.org/10.1016/j.jfineco.2004.12.005.

Yonker, Scott E. (2017). "Geography and the Market for CEOs". *Management Science* 63.3, pp. 609–630. DOI: 10.1287/mnsc.2015.2319.

CHAPTER 2

Growth Options, Incentives, and Pay-for-Performance

with Sebastian Gryglewicz and Barney Hartman-Glaser

2.1 Introduction

A fundamental insight of agency theory is that managers need incentives to maximize shareholder value (Jensen and Meckling, 1976). At a basic level, such incentives require that a manager's expected pay be sensitive to her actions. In practice, these actions are often unobservable, and, as a result, compensation contracts implement incentives by making managers' pay sensitive to performance. As such, a substantial literature has developed that estimates managerial incentives by measuring pay-performance sensitivity (Kevin J Murphy, 1985; Jensen and Kevin J Murphy, 1990; Baker and Hall, 2004). However, as a measure of incentives, pay-performance sensitivity is confounded by the sensitivity of performance to managers' actions. We show that growth options cause the sensitivity of performance to managers' efforts to vary both across firms and within a firm over time. This variation means that pay-performance sensitivity is not a sufficient statistic for incentives. Intuitively, growth options should increase the optimal amount of incentives receive. We show that pay-performance sensitivity can be decreasing in growth option intensity and provide empirical evidence that supports this relation.

We first present the basic intuition behind why pay-performance sensitivity is not a sufficient statistic for incentives in the context of a simple principal-agent problem. In this problem, a manager takes a hidden action, i.e., effort, that affects firm value. Shareholders provide the manager with incentives by making her pay a function of firm value. Three distinct quantities emerge as related to the manager's incentives: expected-pay-effort sensitivity, performance-effort sensitivity, and pay-performance sensitivity. Expected-pay-effort sensitivity is defined by how sensitive the manager's expected pay is to her choice of effort and directly determines her incentives. When expected-pay-effort sensitivity is higher, the manager expects to receive a greater reward for any additional effort she applies, and she will thus respond by applying more effort. Performance-effort sensitivity is the marginal value of managerial effort to the firm. Finally, pay-performance sensitivity is the sensitivity of the manager's pay to the value of the firm. In this simple framework, pay-performance sensitivity is the ratio of expected-pay-effort sensitivity and performance-effort sensitivity so that there is a wedge between pay-performance sensitivity and incentives. As a result, a change in an underlying characteristic of the firm that leads to changes both incentives and performance-effort sensitivity can have an ambiguous effect on pay-performance sensitivity.

Although the arguments we make in our simple principal-agent framework apply to any firm characteristic that affects the wedge between pay-performance sensitivity and incentives, our focus is on growth options. Intuitively, an increase in growth options leads to an increase in the sensitivity of firm value to effort, and at the same time, increases the optimal amount of incentives that the shareholders choose to implement. Herein lies the difficulty of measuring incentives with pay-performance sensitivity. When the elasticity of incentives to growth options is less than that of performance-effort sensitivity, an increase in growth options increases incentives and decreases pay-performance sensitivity.

While our simple principal-agent framework illustrates the core intuition of results, it lacks sufficient richness to address why performance-effort sensitivity might be more or less elastic than expected-pay-effort sensitivity with respect to growth options. To address this question, we present a continuous-time moral hazard model in which the presence of a growth option interacts with the provision of incentives, and characterize the circumstances under which pay-performance sensitivity differs from expected-pay-effort sensitivity.

In the model, an investor hires a manager to run a firm. The investor also possesses a growth option to increase the firm's capital base. The manager can exert unobservable effort to increase productivity growth. The investor provides the manager with incentives by exposing her to fluctuations in productivity. The investor is risk neutral, and the manager is risk averse, so it is costly to provide the manager with incentives. Thus, from the investor's perspective, there are two components of the total cost of effort: effort costs paid by the manager and the incentive costs of forgone risk sharing. Our main result is that expected-pay-effort sensitivity increases with the size of the growth option, whereas pay-performance sensitivity decreases if incentive costs are more convex than effort costs.

The intuition for the result follows from the relation between expected-pay-effort sensitivity and performance-effort sensitivity under the optimal contract. An increase in the size of the growth option increases the marginal benefit of effort and thus increases the optimal level of effort. The contract must increase expected-pay-effort sensitivity to implement such an increase in effort. The manager's incentive compatibility constraint implies that expected-pay-effort sensitivity is equal to her marginal effort cost. When incentive costs are more convex than the manager's effort costs, marginal incentive costs increase more than marginal effort costs. At the optimum, the first-order condition equates the marginal benefit of effort to the sum of marginal effort and incentive costs. As performance-effort sensitivity is equal to the marginal benefit of effort, the first-order condition implies that performance-effort sensitivity increases by more than expected-pay-effort sensitivity whenever incentive costs are more convex that effort cost. As a result, pay-performance sensitivity decreases with the size of the growth option, whereas expected-pay-effort sensitivity increases.

We go on to present new evidence for the relationship between pay-performance sensitivity and growth options. Using data on pay-performance sensitivity calculated by Coles, Daniel, and Naveen (2013), as well as executive and firm characteristics from the Execucomp and Compustat databases, we find that pay-performance sensitivity is negatively related to proxies for growth options. Specifically, we regress dollar changes in manager wealth to dollar changes in firm value, a measure of pay-performance sensitivity suggested by Jensen and Kevin J Murphy (1990) that we call PPS, on market-to-book ratio and other proxies. We find that, for a given firm, a one standard deviation increase in the market-to-book ratio is associated with a 5.7% decrease in PPS. As a standalone fact, this relationship seems inconsistent with the intuition that growth options make manager effort more valuable, which should necessitate stronger incentives when growth options are present. Viewed through the lens of our model, we find this intuition to be compatible with the empirical relationship, as more growth options make manager effort more valuable to the firm. If the value of manager effort increases faster than manager incentives, we would expect PPS to decrease, despite stronger managerial incentives. Thus our model also has implications for the provision of incentives. We investigate our model's predictions regarding compensation design and find them consistent with observed compensation structures among high-growth pre-IPO firms.

Although our main result is about the wedge between expected-pay-effort sensitivity and pay-performance sensitivity in the presence of growth options, the intuition behind our results holds in a much broader setting. We illustrate the generality of our results by extending our model to consider the case of an abandonment option. Rather than an opportunity to invest in additional capital, the investor instead has the opportunity to shut down the firm and sell its assets for a fixed value. In this setting, the liquidation value of the asset is a measure of the assets' redeployability. We find that redeployability has an effect on incentives that is symmetric to the effect of growth options: an increase in redeployability decreases expected-pay-effort sensitivity, but increases pay-performance sensitivity.

Our work is related to the large literature on executive compensation. Frydman and Jenter (2010) and Kevin J. Murphy (2013) provide comprehensive reviews of the theoretical and empirical findings in this literature. In emphasizing one important aspect of the empirical evaluation of PPS, we are indebted to the research that extensively documents the importance of managerial incentives in firm decision making. Coles, Daniel, and Naveen (2006), Hirshleifer and Suh (1992), and Rajgopal and Shevlin (2002) are among the many papers that document the effect of managerial incentives on operational decisions. There has also been work on the effect of incentives on other financing decisions, as studied by Babenko (2009) and Chava and Purnanandam (2010), among others. Zhiguo He, Li, et al. (2014) analyzes the impact of uncertainty on managerial incentives and finds that the desire for faster learning leads the investor to offer stronger incentives to the manager.

Theoretical studies have characterized the optimal compensation contract in a variety of settings.[1] An important observation made by Baker and Hall, 2004 is that the optimal structure of compensation depends on the model's assumptions about how managerial effort affects firm value. In the models of Lambert (1983), Rogerson (1985), Edmans, Gabaix, Sadzik, et al. (2012), and Zhiguo He, Li, et al. (2014), a feature of the optimal contract is the effect of present performance on both current and future compensation. Those models are in discrete time, whereas our work, like that of DeMarzo and Sannikov (2006) and Zhiguo He, Wei, et al., 2017, uses a continuous-time setting. A continuous-time model is desirable because it permits characterization of the optimal contract and the firm's value function using ordinary differential equations.

In our model, real options are a source of convexity in firm value and create the wedge between incentives and pay-performance sensitivity. First introduced in Brennan and Schwartz (1985), there is substantial literature analyzing the presence and implications of investment opportunities as options. Berk, Green, and Naik (1999) finds that the optimal exercise of investment opportunities can simultaneously reproduce a multitude of cross-sectional asset pricing features. Carlson, Fisher, and Giammarino (2004) builds on this analysis by introducing operating leverage and reversible investment. In a similar spirit, by analyzing real options in the context of managerial incentives, we work to understand the rich interdependence between managerial decision making and investment opportunities.

By studying the effect of real options on incentives, our paper contributes to the literature on manager incentives. The seminal paper in this area is Holmstrom and Milgrom (1987), which studies the contract between a risk-averse manager and a risk-neutral firm. Our model is similar to that of Z. He (2011) in that it features a risk-averse manager who can exert effort to increase expected cash flows. Unlike that model, our model gives the firm a growth option. Similar to earlier models, there are two kinds of costs in implementing effort, as described first in Holmstrom and Milgrom (1987): the direct monetary cost and the risk-compensation term to encourage the risk-averse agent to bear incentives.

Empirical studies on measuring PPS were pioneered by the competing measures of Jensen

[1] See, for example, Gabaix and Landier (2008), Chaigneau, Edmans, and Gottlieb (2014), and Edmans and Gabaix (2011).

and Kevin J Murphy (1990) and Hall and Liebman (1998). An important contribution was made by Core and Guay (2002), who provided a methodology for estimating the sensitivity of option-based compensation. Our work both relies upon and contributes to the measurement of PPS by identifying growth options as an important source of variation in PPS. In this way, our work contributes to the literature on the determinants of executive compensation.[2]

Finally, our paper is related to the literature on option exercise in the presence of agency problems and asymmetric information. Grenadier and Wang, 2005 analyze how agency conflicts such as moral hazard and hidden information can affect the timing of real option exercise. Grenadier and A. Malenko, 2011 study a setting in which informed agents signal their private information by exercising real options. Grenadier, A. Malenko, and N. Malenko, 2016 analyze how timing decisions interact with communication. Cong, 2017 studies the relation between auctions of real options and investment timing decisions. The setup of our model follows that of Gryglewicz and Hartman-Glaser, 2016, which looks at the timing of investment decisions in the presence of agency conflicts. Rather than focusing on the investment decision, we focus on how growth options can affect manager incentives.

2.2 Pay-Performance Sensitivity and Incentives

In this section, we present a simple principal-agent problem that illustrates our main point: pay-performance-sensitivity is not a direct measure of incentives. To that end, consider an investor who hires a manager to operate a firm. The gross value of this firm, V, is an increasing function of a state $X = a + Z$, where a is the manager's hidden action a, and Z is mean zero noise. The shape of V is determined by an exogenous parameter λ. For example, λ could represent the firm's size, its level of productivity, or, as we focus on in this paper, the firm's endowment of growth options.

A contract specifies a compensation rule c that determines the amount to be paid to the manager by the investor. As the investor cannot observe the actions of the manager, he

[2] See also Baker, Jensen, and Kevin J Murphy (1988), Deckop (1988), Yermack (1995), and Becker (2006).

can only condition the manager's compensation on the realization of firm value or X. For simplicity, we restrict attention to contracts that are affine in firm value:

$$c(V) = W + \phi V, \tag{2.1}$$

so that ϕ corresponds to the dollar increase in the manager's pay per dollar increase in firm value., i.e., the standard definition of pay-performance sensitivity as in Jensen and Kevin J Murphy (1990). While this restriction is not without loss of generality, it substantially simplifies the intuition we present below. One advantage of the dynamic model we present in Section 2.3 is that it will allow a characterization of the optimal contract over an unrestricted contract space.

Taking the compensation rule c as given, the manager chooses the action a^* that maximizes her expected compensation net of a convex effort cost $g(a)$. The firm chooses the optimal contract to maximize the value V net of manager compensation c, taking into account the manager's choice of action a^*. The optimal contract then solves the following problem

$$\max_{c} \left\{\mathbb{E}\left[V\left(X\right) - c\left(V\left(X\right)\right) \mid a^*\right]\right\} \tag{2.2}$$

such that

$$a^* \in \arg\max_{a} \left\{\mathbb{E}\left[c\left(V\left(X\right)\right) \mid a\right] - g\left(a\right)\right\}, \tag{2.3}$$

and

$$\mathbb{E}\left[c\left(V\left(X\right)\right) \mid a^*\right] - g\left(a^*\right) \geq u_0, \tag{2.4}$$

where u_0 is the value of the manager's outside option.

In this setting, the manager's incentives are determined by her expected-pay-effort sensitivity, i.e., how much expected compensation increases in response to an increase in effort.

This quantity, denoted β, is given by

$$\beta = \frac{\mathrm{d}\mathbb{E}\left[c\left(V\left(X\right)\right) \mid a\right]}{\mathrm{d}a}. \tag{2.5}$$

To see why β captures the strength of the manager's incentives, examine the incentive compatibility condition in Equation (2.3). It implies that the manager will choose an effort level that equates the marginal cost of effort with the marginal benefit, i.e., her expected-pay-effort sensitivity. Since the cost of effort is convex, the higher is her expected-pay-effort sensitivity, the higher is her optimal effort level.

We are interested in determining the conditions under which inferences about manager incentives, i.e., expected-pay-effort sensitivity, can be drawn by observing pay-performance sensitivity. For example, suppose empirical evidence shows that pay-performance sensitivity is decreasing in some firm characteristic λ, for example, growth option intensity. Can we conclude that incentives are also decreasing in this characteristic? To answer this question, we can examine the comparative statics of both ϕ and β with respect to λ. If these two comparative statics have the same sign, then the two measures are aligned, and we can conclude that incentives are also decreasing in λ. However, as we now show, β can be increasing in λ even though ϕ is decreasing. This relation implies that evidence that pay-performance sensitivity is decreasing in some firm characteristic is not sufficient to conclude that incentives are also decreasing.

Given our restriction to affine contracts, expected-pay-effort sensitivity is the product of pay-performance sensitivity and the marginal value of manager effort

$$\beta = \phi \frac{\mathrm{d}\mathbb{E}\left[V(X) \mid a\right]}{\mathrm{d}a} = \phi \mathbb{E}\left[V'(X) \mid a\right]. \tag{2.6}$$

Taking a derivative of Equation (2.6) with respect to λ gives

$$\frac{1}{\beta}\frac{\partial \beta}{\partial \lambda} = \frac{1}{\phi}\frac{\partial \phi}{\partial \lambda} + \left(\frac{1}{\mathbb{E}\left[V'(X) \mid a\right]}\right)\frac{\partial \mathbb{E}\left[V'(X) \mid a\right]}{\partial \lambda}. \tag{2.7}$$

In words, Equation (2.7) just states that the elasticity of β with respect to λ is the sum of the elasticities of ϕ and the marginal value of manager effort. Thus, ϕ can be decreasing in λ, while β is increasing in λ if the elasticity of the marginal value of manager effort is sufficiently positive.

In the specific case in which λ represents growth opportunity intensity, an increase in λ increases $V'(X)$. An increase in λ thus increases the marginal value of manager effort, and the amount of incentives provided per unit of pay-performance sensitivity increases. As a result, actual incentives can increase even as pay-performance sensitivity decreases.

The above analysis also has implications for the design of incentives. A typical feature of many contracting models is that executive pay that is convex in performance provides strong incentives. This feature would seem to imply that if a firm's owners seek to provide powerful incentives, then executive pay should include option-like compensation. Equation (2.6) shows that some convexity in incentives is present just because the firm value is itself a convex function the underlying effort of the manager. When a firm has growth opportunities, firm value is an option-like function of productivity. As a result, paying executives with stocks or deep-in-the-money options still provides convex incentives.

Equation (2.6) also indicates that the problem of drawing inferences about incentives using data on pay-performance sensitivity is akin to using average q to draw inference about marginal q. As is well understood, marginal and average q are not necessarily equivalent if the marginal value of investment is not constant. In the same vein, pay-performance sensitivity is not necessarily equivalent to incentives if the marginal value of manager effort is not constant. However, just as average q is useful in the empirical investigation of real investment because it is readily observable and measurable, so is pay-performance sensitivity in the empirical investigation of incentives. Our point is that one must take care to control for the marginal value of manager effort when using pay-performance sensitivity as a proxy for incentives in the same way that one must take care to properly control for the marginal value of investment when using average q as a proxy for marginal q.

In this simple principal-agent problem, not only is pay-performance sensitivity distinct from incentives, but observed changes in the former are uninformative about the latter. In the next section, we formalize the simple intuition we have presented in the context of a fully specified model in which the relation between growth option intensity and pay-performance sensitivity arises endogenously from a dynamic principal-agent problem.

2.3 A Dynamic Model of Real Options and Manager Moral Hazard

We now present a dynamic model that builds on Holmstrom and Milgrom (1987) and Z. He (2011), which explicitly solves for the optimal contract and relates the manager's incentives to the primitives of the model. The dynamic model has the advantage of allowing us to characterize firm value and the optimal contract in closed form and analyze comparative statics of pay-performance sensitivity versus pay-effort sensitivity. Building on the intuition of the previous section, we find that the two are distinct and respond differently to changes in the value of managerial effort.

In the model, time is continuous and indexed by t. An infinitely lived firm generates a continuous cash flow given by $X_t K_t$, where K_t is the capital base and X_t is firm productivity. Capital K_t takes the initial value $K_0 = 1$, and the firm has a real option to pay P and increase capital to k. Let τ denote the time of investment.

A risk-neutral investor hires a risk-averse manager to run the firm. The common discount rate is denoted by r. Both X_t and K_t are observable to the investor. A moral hazard problem arises because the manager affects the firm's productivity. Specifically, prior to investment, productivity X_t depends on that manager's effort $a_t \in [0, a_{\max}]$ and follows the process

$$dX_t = a_t X_t \, dt + \sigma X_t dZ_t, \tag{2.8}$$

where σ is a positive constant and dZ_t is a Brownian motion that is unobservable to the investor. We assume that $r > a_{\max}$, so that firm value is finite. The manager's effort is unobservable to the investor.

In our model, the value of the manager is due to her ability to grow the firm's productivity

X_t. This view of a manager is consistent with characterizations of CEOs as focused on growth and future performance. As our interest lies in the interaction of agency conflicts and growth opportunities, we simplify the analysis and assume that after an investment at time τ, firm productivity stays at X_τ forever and there are no agency conflicts. Thus, the post-investment value of the firm's cash flow is just $(X_\tau k)/r$. In what follows, we examine the optimal contracting and valuation of the firm before investment.

The investor receives the cash flows from the firm and pays the manager compensation c_t so that her net cash flow D_t follows dynamics given by

$$dD_t = X_t K_t \, dt - c_t \, dt - P \, dJ_t, \tag{2.9}$$

where $J_t = \mathbb{I}(t \geq \tau)$. We note that this specification for cash flows links current cash flows and operations to the payoff to the growth option. This feature is not essential. An alternative formulation of our model is to let X_t only affect the productivity of new capital, not current cash flows and would yield the same economic mechanism we discuss below. The key ingredient for the results we present below is that managerial effort affects the value of the growth option through productivity growth.

The manager has constant absolute risk aversion (CARA) preferences over consumption and effort with instantaneous utility

$$u(c_t, a_t) = -\frac{1}{\gamma} e^{-\gamma (c_t - g(a_t) X_t)}. \tag{2.10}$$

The manager's private cost of effort, $g(a_t)X_t$, is measured in units of consumption. We assume the cost function $g(a)$ is continuous, increasing, and convex in effort a: $g(a) \in \mathcal{C}^1([0, a_{\max}])$, $g'(a) > 0$, $g''(a) > 0$, $g(0) = g'(0) = 0$, and $g'(a_{\max}) = \infty$. This specification for effort costs ensure that any optimal contract will specify interior effort in $(0, a_{\max})$. The cost of effort increases with the firm's current level of productivity, and therefore with firm size. This captures the intuition that it is more difficult and costly for a manager to improve the productivity of an already productive firm.

The manager has the ability to engage in unobserved savings and borrowing at the rate r. This assumption restricts the type of incentives that the investor can impose on the manager. If the manager did not have access to private savings, the investor could implement more powerful incentives by distorting her intertemporal margin and ratcheting up incentives over time. When the manager has access to private savings, the investor can only expose the manager to long-term risk via deferred compensation; otherwise, the manager could use precautionary savings to undo incentives. Without loss of generality, we assume that the manager starts with zero savings. It is also important to note that the manager cannot gain exposure to the firm's equity except through the contract, as this would also allow her to undo incentives. Furthermore, the manager has a outside option which she values at w_0.

A contract consists of a compensation rule, a recommended effort level, and an investment policy, denoted $\Pi = (\{c, a\}, \tau)$, where $\{c\} = \{c_t\}_{t\geq 0}$ and $\{a\} = \{a_t\}_{t\geq 0}$ are stochastic processes adapted to the filtration of public information $\mathcal{F}_t$, and τ is an $\mathcal{F}_t$-stopping time.

Given contract Π, the manager chooses the stochastic processes $\{\tilde{c}, \tilde{a}\}$ (which can differ from those recommended by the contract, $\{c, a\}$) to maximize her utility from the contract as follows:

$$W_0(\Pi) = \max_{\{\tilde{c},\tilde{a}\}} \mathbb{E}\left[\int_0^\infty -\frac{1}{\gamma} e^{-\gamma(\tilde{c}_t + g(\tilde{a}_t, X_t)) - rt},\right] \tag{2.11}$$

such that X_t, K_t, and S_t follow the dynamics induced by the consumption and effort plan $\{\tilde{c}, \tilde{a}\}$. The investor's value given a contract Π is

$$v_0(\Pi) = \mathbb{E}^{\{\tilde{a}\}}\left[\int_0^\infty e^{-rt} dD_t\right], \tag{2.12}$$

such that X_t, K_t, and S_t follow the dynamics induced by the consumption and effort plan $\{\tilde{c}, \tilde{a}\}$ and where $\{\tilde{c}, \tilde{a}\}$ solves (2.11). The expectation operator $\mathbb{E}^{\tilde{a}}$ denotes dependence of expectations on the dynamics under effort $\{\tilde{a}\}$. Therefore, the investor chooses the optimal contract to maximize $v(\Pi)$ subject to providing the manager at least her outside option w_0.

A contract Π is termed *incentive-compatible* and *zero-savings* if the manager's choice of

$\{\tilde{c}, \tilde{a}\}$ is equal to the payment rule and recommended effort plan $\{c, a\}$ given in Π. We restrict our attention to incentive-compatible and zero-savings contracts by virtue of the following version of the revelation principle.

Lemma 1. *Let $\tilde{\Pi}$ be an arbitrary contract. There exists an incentive-compatible and zero-savings contract Π that satisfies $v(\Pi) \geq v(\tilde{\Pi})$ and $W(\Pi) \geq W(\tilde{\Pi})$.*

2.3.1 No-Savings and Incentive-Compatibility Conditions

The manager is compensated in current pay and promised deferred pay. The zero-savings property of the optimal contract has implications for current pay. As the manager is risk averse, she values smooth consumption. Thus, if current compensation is high relative to her wealth, she will only consume a part of the compensation and save the rest. Conversely, if current compensation is low relative to her wealth, she will borrow to increase current consumption. With CARA preferences, the manager will not save or borrow if her current utility from consuming exactly her compensation equals the risk-free yield on her continuation utility from the contract.

Lemma 2. *A contract implements zero savings if and only if the manager's instantaneous utility is equal to the yield on her continuation utility:*

$$u\left(c_t, a_t\right) = rW_t. \tag{2.13}$$

Given the manager's continuation utility W_t and effort a_t, the no-savings property pins down an exact level of current pay. Next, we analyze deferred pay and its role in providing incentives. To do so, we characterize the dynamics of W_t under the recommended consumption and effort plan. Utility from current pay and a change of the continuation utility from deferred pay must in expectation equal the required return on the continuation utility, that is, it holds that

$$\mathbb{E}_t\left[u(c_t, a_t)dt + dW_t\right] = rW_t\, dt. \tag{2.14}$$

The no-saving condition (2.13) implies that $\mathbb{E}_t [dW_t] = 0$. Using the martingale representation theorem as in Sannikov (2008), we can write the following dynamics for the manager's continuation utility:

$$dW_t = \beta_t (-\gamma r W_t) (dX_t - a_t X_t \, dt) \tag{2.15}$$

for some progressively measurable process β_t. The term $\beta_t(-\gamma r W_t)$ is the sensitivity of the manager's continuation utility to unexpected shocks to the firms' productivity. The term $-\gamma r W_t$ is a scaling factor that equals the marginal utility of consumption. As a result, β_t measures the sensitivity of the manager's continuation value to unexpected shocks to productivity in monetary terms. If the manager deviates from the recommended effort policy, she expects the investor to perceive an unexpected shock to productivity, and her continuation value to adjust by β_t. Thus, β_t measures the manager's incentive to deviate from the contract's recommended effort policy.

We can now characterize the incentive compatibility constraint for the manager. Consider the manager's choice of effort $\tilde{a}_t$. As the manager chooses $\tilde{a}_t$ to maximize the sum of her instantaneous utility $u(c_t, \tilde{a}_t) \, dt$ and the expected change in her continuation utility W_t, her expected change in continuation utility achieved by a deviation from the recommended effort level a_t to $\tilde{a}$ is

$$\mathbb{E} [dW_t | \tilde{a}] = \beta_t (-\gamma r W_t) (\tilde{a} - a_t) X_t \, dt. \tag{2.16}$$

For the recommended effort level a_t to be incentive-compatible, it must be the case that

$$a_t \in \arg\max_{\tilde{a}} \{ u(c_t, \tilde{a}) + \beta_t (-\gamma r W_t) (\tilde{a} - a_t) X_t \} . \tag{2.17}$$

According to our assumptions about the cost function $g(a)$, the optimal choice of effort will take on an interior solution in the interval $(0, a_{\max})$. Taking the first-order condition yields

$$u_a (c_t, a_t) + \beta_t (-\gamma r W_t) X_t = 0. \tag{2.18}$$

Substituting $u_a(c_t, a_t) = -u_c(c_t, a_t)\, g'(a_t) X_t$ and the no-savings condition (2.13), we can rearrange the first-order condition above as follows:

$$\beta_t = g'(a_t). \tag{2.19}$$

Intuitively, incentive-compatibility requires that the sensitivity, β_t, of the manager's continuation utility to unexpected output shocks is equal to her marginal cost of effort $g'(a_t)X_t$, scaled by the marginal effect of effort on output, X_t.

Lemma 3. *A contract is incentive-compatible and implements zero savings if and only if the solution W_t to the manager's problem has dynamics given by* (2.15)*, where β_t is defined by* (2.19)*.*

The agent's continuation utility W_t can be used as a state variable to solve for the optimal contract. It is useful to further transform W_t into its certainty equivalent

$$Y_t = -\frac{1}{\gamma r} \ln(-\gamma r W_t), \tag{2.20}$$

so that we can take Y_t to be a state variable for the investor's problem in place of W_t. Applying Ito's Lemma shows that the dynamics of Y_t under an incentive-compatible, zero-savings contract are given by the following equation:

$$dY_t = \frac{1}{2}\gamma r \left(\sigma \beta_t X_t\right)^2 \, dt + \sigma \beta_t X_t \, dZ_t^a, \tag{2.21}$$

where Z_t^a is a Brownian motion under the probability measure induced by effort a. Although W_t is a martingale, the difference in risk aversion between the investor and the manager implies that the certainty equivalent Y_t must have additional drift for each additional unit of volatility. This positive drift will appear in the investor's Hamilton-Jacobi-Bellman (HJB) equation as the cost of providing incentives.

2.3.2 Solving for the Optimal Contract

We now present a heuristic derivation of the optimal contract. First, we characterize the payment rule to the manager. Recall that the zero-savings condition links the manager's instantaneous utility $u(c_t, a_t)$ and her continuation utility W_t. This link allows us to express the manager's compensation as a function of the state of the firm X_t, the recommended effort level a_t, and the certainty equivalent Y_t:

$$c_t = rY_t + g(a_t)X_t. \tag{2.22}$$

We see that the manager's compensation is the yield on her continuation utility, plus her cost of effort.

Note that Equation (2.22) also specifies the manager's compensation after investment. As there is no more effort implemented after investment for $t \geq \tau$, the manger's continuation utility stays constant, and the manager's compensation is simply the yield on her continuation utility, $c_t = rY_\tau$. The present dollar value of such compensation is Y_τ.

We take the dynamic programming approach to determine the optimal effort and the investment timing. The investor's value function $v(X,Y)$ depends on both the firm's productivity X and the certainty equivalent of the manager's continuation utility Y. Over any interval of time in which there is no investment, the investor receives the flow equal to X minus compensation c. An application of Ito's Lemma to the dynamics of X and Y gives the following Hamilton-Jacobi-Bellman equation for $v(X,Y)$:

$$\begin{aligned} rv(X,Y) = \max_a \Big\{ & X - (rY + g(a)X) + aXv_X(X,Y) + \frac{1}{2}\sigma^2 X^2 v_{XX}(X,Y) \\ & + \frac{1}{2}\gamma r \left(\sigma g'(a)X\right)^2 v_Y(X,Y) + \frac{1}{2}(\sigma g'(a)X)^2 v_{YY}(X,Y) \Big\}. \end{aligned} \tag{2.23}$$

As firm value is monotonically increasing in X, the optimal investment time τ follows a threshold rule given by $\tau = \inf\{t | X_t \geq \bar{X}\}$. Following standard solution methods, we find this threshold using value-matching and smooth-pasting conditions:

$$v(\overline{X}, Y) = \frac{\overline{X}k}{r} - P - Y \tag{2.24}$$

$$v_X(\overline{X}, Y) = \frac{k}{r}. \tag{2.25}$$

We can simplify the problem by noting that due to the absence of wealth effects implied by the manager's CARA preferences, the total firm value is independent of the manger's continuation utility. In other words, the investor's value depends on the manager's continuation utility only by the certainty equivalent cost of the future obligation to the manager. It thus holds that $v(X, Y) = V(X) - Y$, where $V(X)$ represents total firm value. Using this relation, we can rewrite Equation (2.23) as

$$rV(X) = \max_a \left\{ X - (g(a) + \rho(a))X + aXV'(X) + \frac{1}{2}\sigma^2 X^2 V''(X) \right\}, \tag{2.26}$$

where

$$\rho(a) = \frac{1}{2}\gamma r \left(\sigma g'(a)\right)^2 X \tag{2.27}$$

represents the incentive cost of effort. Boundary conditions (2.24) and (2.25) can be rewritten as

$$V(\overline{X}) = \frac{\overline{X}k}{r} - P, \tag{2.28}$$

$$V'(\overline{X}) = \frac{k}{r}. \tag{2.29}$$

In summary, we obtain the following result.

Proposition 1. *The optimal contract is given by the solution to* (2.26), (2.28), *and* (2.29).

2.4 Growth Options and Optimal Incentives

In this section, we consider the implications of real options for managerial incentives. Our question is how optimal incentives and common measures of pay-performance sensitivity respond to an increase in the size of the growth option, as measured by k. Keeping the cost of investment P constant, increased k means that the growth option is larger and more valuable. Although there have been many empirical investigations, reviewed by Kevin J. Murphy (1999) and Frydman and Jenter (2010), into the relation between pay-performance sensitivity and firm size, there has been less attention paid to the relation between investment and pay-performance sensitivity. Our results guide the empirical analysis presented in the following section.

2.4.1 Measuring Incentives in the Presence of Growth Options

The manager's compensation and incentives depend on the level of effort stipulated by the optimal contract. Therefore, we begin our inquiry with a discussion of managerial effort. Given our assumptions on the manager's effort cost function $g(a)$, the optimal effort is interior and satisfies the first-order condition:

$$V'(X) = g'(a^*(X)) + \rho'(a^*(X)). \tag{2.30}$$

The marginal benefit of effort is the value of increasing the growth rate of productivity, or $V'(X)$. The marginal cost of effort includes two terms. The first is the marginal increase in compensation the manager required to cover her effort. The second is the marginal increase in incentive costs the investor must pay to increase incentives. In the following analysis, we restrict our attention to parameter values such that the maximum $a^*(X)$ satisfies the second-order condition.[3] Optimal effort, which varies with productivity X_t, depends on the fundamental parameters of the model and the presence of growth opportunities.

A direct measure of the manager's incentives in our model is the sensitivity of her dollar (certainty-equivalent) continuation utility to productivity shocks.[4] Prior to investment, the

[3] If the second-order derivative of the objective function with respect to a is zero (a knife-edge case given its dependence of X), then the implicit function theorem is not applicable.

[4] Given a performance metric, a standard way of measuring incentives is ΔManager's Wealth/ΔPerformance. When performance is a diffusion process Q, the continuous-time analog to this measure is dY/dQ since Y measures the dollar value of the manager's wealth. Since

optimal contract sets the quantity to

$$\beta^*(X) = g'(a^*(X)). \tag{2.31}$$

This expression follows directly from substituting the optimal effort policy $a^*(X)$ into the incentive compatibility condition given by Equation (2.19). Note that $\beta^*(X)$ is also the expected dollar increase in the manager's dollar wealth resulting from an additional unit of effort, as an additional unit of effort is expected to raise productivity by one unit. In other words, $\beta^*(X)$ is the manager's expected-pay-effort sensitivity and is equivalent to pay-performance sensitivity as long as changes in performance are measured by changes in current productivity. Unfortunately, changes in productivity are difficult to measure empirically.

A standard approach for the measurement of incentives is to compute the sensitivity of the manager's wealth to changes in firm value, i.e., the manager's value-based pay-performance sensitivity, as first proposed by Jensen and Kevin J Murphy (1990).[5] This approach is particularly convenient from an empirical point of view, as it is based on firm value changes, which are easy to measure. In our model, as in Z. He, 2011, the manager's dollar value-based pay-performance sensitivity is equal to the sensitivity of the manager's dollar continuation value to changes in firm value, $V(X)$. Under the optimal contract, this quantity is given by

$$\phi^*(X) = \frac{\beta^*(X)}{V'(X)} = \frac{g'(a^*(X))}{V'(X)}. \tag{2.32}$$

Note that although $\phi^*(X)$ is closely related to $\beta^*(X)$, it is scaled by the slope of the value function in output $V'(X)$. Thus, the presence of growth options affects $\phi^*(X)$ by changing both $\beta^*(X)$ and $V'(X)$. To relate to the simple analysis that we conduct in Section 2.2,

---

$dZ \cdot dt = 0$ and $dZ^2 = dt$, we have

$$\frac{dY}{dQ} = \frac{dY}{dQ}\frac{dZ}{dZ} = \frac{\sigma g'(a^*(X))X}{\sigma_Q},$$

where the numerator is the volatility of Y given in Equation (2.21) and σ_Q is the volatility of Q. Performance metric Q is X in Equation (2.31) and is V in Equation (2.32).

[5] See also Yermack (1996) and Bergstresser and Philippon (2006).

the size of the growth option affects both the optimal expected-pay-effort sensitivity and the sensitivity of the performance measure, in this case, firm value, to effort. As we show in the next proposition, the wedge between β^* and ϕ^* induced by $V'(X)$ can lead the two quantities to respond in opposite ways to changes in growth option size.

Proposition 2. *As the size of the growth option* increases:

1. *optimal effort $a^*(X)$ and expected-pay-effort sensitivity $\beta^*(X)$* increase*,*
2. *pay-performance sensitivity $\phi^*(X)$* decreases *if and only if the incentive cost of effort is more convex than the direct cost of effort, that is if and only if*

$$\frac{\rho''(a)}{\rho'(a)} > \frac{g''(a)}{g'(a)}. \tag{2.33}$$

The intuition behind Proposition 2 is as follows. A larger growth opportunity increases the benefits that the investor derives from managerial effort and hence increases optimal effort. To induce this increased effort, expected-pay-effort sensitivity increases. The intuition for the second part of the proposition relies on the relation between expected-pay-effort and value-effort sensitivity. First note that as a unit of effort leads to a unit of expected increase in X, we can interpret $V'(X)$ as value-effort sensitivity. Thus, the first-order condition in Equation (2.30) states that value-effort sensitivity is equal to expected-pay-effort sensitivity, $\beta^*(X)$, plus the marginal incentives costs evaluated at the optimal level of effort, $\rho'(a^*)$. When the incentive cost of effort is more convex than the direct cost of effort, an increase in optimal effort results in marginal incentive costs comprising a greater proportion of the total marginal effort costs. The first-order condition then implies that expected-pay-effort sensitivity does not increase by as much as value-effort sensitivity in response to an increase in growth option size. As a result, value-based pay-performance sensitivity decreases.

We note that a wide range of effort cost functions satisfy the condition given in Equation (2.33). First note that given the definition of $\rho(a)$, an equivalent way to state the condition is that the marginal cost of effort is convex, i.e.,

$$g'''(a) > 0, \tag{2.34}$$

which is common in the contract theory literature (Cheng, Hong, and Scheinkman, 2015; Bolton, Santos, and Scheinkman, 2016). For example, the condition is satisfied if effort costs are given by a power function, $g(a) = a^\eta$ where $\eta > 2$, by a log-linear function $g(a) = (e^{\eta a} - 1)$ with $\eta > 0$, or by an increasing convex function $g(a) = \frac{a^\eta}{a_{\max} - a}$ with $\eta \geq 1$ (which ensures interior effort).

One way to interpret the shape of the marginal cost of effort is as a measure of the degree of complexity of the task. Some tasks are relatively simple no matter the scale of effort and therefore have an increasing, but concave, marginal cost of effort. Other tasks get more and more complex as the scale of effort increases and thus have a convex marginal cost of effort. For example, implementing process systems that increase the productivity of capital likely gets more and more complex as the scale of these systems increases. This latter case applies to our model, and we expect that the condition in Equation (2.33) should hold in the data.

Another implication of our model is that different definitions of pay-performance sensitivity can have different comparative statics with respect to the same underlying parameter. For example, if we measure pay-performance sensitivity using the sensitivity of the manager's dollar continuation value to percent changes in firm value, denoted $\varphi^*(X) = V(X)\phi^*(X)$, we can write the comparative static as

$$\frac{\partial \varphi^*(X)}{\partial k} = V(X)\frac{\partial \phi^*(X)}{\partial k} + \frac{\partial V(X)}{\partial k}\phi^*(X). \tag{2.35}$$

Proposition (2) gives a condition for ϕ^* to be decreasing in the size of the growth option and thus for the first term on the right-hand side of (2.35) to be negative. At the same time, the second term on the right-hand side of (2.35) is positive as it is always the case that the value of the firm is increasing in the size of the growth option, k, and that the optimal sensitivity ϕ^* is positive. Thus, even if ϕ^* is decreasing in k, φ^* need not be decreasing. As a result, our model provides guidance as to why different conclusions regarding managerial incentives can arise when using seemingly similar measures of pay-performance sensitivity.

2.4.2 Implications for the Implementation of Incentives

In this section, we discuss the implications that the presence of growth opportunities have for the practical implementation of incentives. A common question in the literature on incentives, for example, as Kevin J. Murphy, 1999 summarizes, is what is the shape of incentive structures, either under an optimal contracting model or in the data? To shed light on this question, we first consider a simple implementation of our optimal contract.

The optimal contract can be implemented using a combination of wages and a managed equity account to provide incentives. The manager's wages ensure that the manager has compensation net of effort costs equal to rY, the riskless yield on her certainty equivalent. The managed equity account ensures that the manager's wealth is sensitive to changes in firm value. The share units in the equity account adjust in response to changes in firm value to maintain the managers pay-performance sensitivity ϕ. Alternatively, the equity account can implement the same pay-performance sensitivity using a managed portfolio of options with appropriate delta sensitivity Δ. Varying levels of incentives can be achieved by performance vesting of stock and option grants or by non-linearity of option holdings.

The typical approach for the analysis of incentives in the context of an implementation like the one above is to determine the convexity or concavity of the manager's managed incentive account with respect to the firms share price. In our model, this exercise corresponds to determining the slope of pay-performance sensitivity in firm value V. For example, a pay-performance sensitivity increasing in firm value V indicates an incentive scheme convex in share price. However, it is crucial to account for the fact illustrated in Section 2.2 and Proposition 2 that pay-performance sensitivity is not equivalent to incentives.

Suppose that managerial incentives measured by pay-effort sensitivity β are increasing in firm value V. Recall that pay-performance sensitivity ϕ equals β divided by $V'(X)$, value-effort sensitivity. In the presence of growth options, as productivity X increases and the firm gets closer to the investment threshold, $V'(X)$ increases. As a result, ϕ must be increasing less steeply in V than β is and in fact can even be decreasing in V. The intuition is that growth options increase the impact of the manager on firm value by generating convexity in firm value, $V''(X) > 0$, so that the manager's pay can be less sensitive to performance and still provide sufficient incentives. To illustrate the distinction between pay-effort sensitivity (incentives) and pay-performance sensitivity, we plot pay-effort sensitivity β and pay-performance sensitivity ϕ against firm value V in Figure 2.1. We see that pay-effort sensitivity is increasing in firm value, implying that optimal effort is also increasing in firm value, yet pay-performance sensitivity is decreasing in firm value.

Figure 2.1 illustrates that in firms with significant growth opportunities, it is not necessary for the optimal contract to prescribe an incentive scheme that is convex in firm value because firm value itself is convex. This pattern is broadly consistent with the common practice of granting employees in-the-money stock options in technology companies.[6] In January 2014, GoPro (GPRO) issued employees stock options as part of their compensation which had a strike price of \$16.22 per share. This price was significantly lower than GPRO's June 2014 IPO price of \$24.00 per share. Other firms which did this include Snap (SNAP), which in 2014 offered options at a strike price of \$1.00 per share when their latest valuation put them at \$3.40; as well as Veritone (VERI), which offered options at a strike of \$8.24 per share against an IPO price of \$15.00 per share. Given the institutional constraints that necessitate the use of options in employee compensation schemes, granting in-the-money options reduces the convexity of the incentive package compared to an incentive package featuring at-the-money or out-of-the-money options, and thus more closely matches the desired convexity of the optimal contract. By highlighting the fact that convexity of pay-performance sensitivity can come from either convexity in the compensation structure or convexity of the underlying firm value, our model helps explain the popular choice to use in-the-money options in compensation packages.

[6]Data on firm-specific employee stock options are from SEC Filings around the respective firms' IPO's.

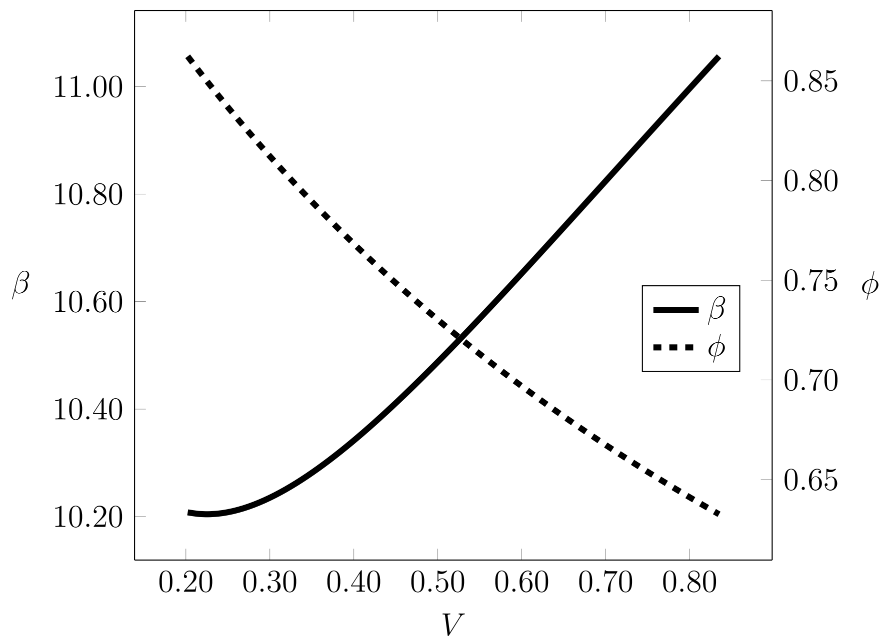


Figure 2.1: The shape of incentives in firm value. Optimal pay-effort β and pay-performance sensitivity ϕ are presented over a range of firm value V. Parameters used for this plot are given by $r = 10\%$, $a_{\max} = 5\%$, $\sigma = 20\%$, $\theta = 1000$, $\gamma = 1$, $k = 1.75$, $p = 0.25$. The effort function $g(a) = \theta \frac{a^3}{a_{\max} - a}$ is chosen to satisfy the conditions laid out in Proposition 2.

2.5 Empirical Findings

In this section, we provide evidence that value-based pay-performance sensitivity decreases with the size of growth opportunities.

2.5.1 Data

We merge data from three main sources. We use data on pay-performance sensitivity for the 1992-2014 period at the manager-firm level from the website of Lalitha Naveen.[7] An empirical equivalent of our model's value-based pay-performance sensitivity is Jensen and Kevin J Murphy, 1990's measure of pay-performance sensitivity, that is, dollar changes in manager wealth divided by dollar changes in firm value. We call this variable Jensen and

[7] Available at http://astro.temple.edu/~lnaveen/data.html. Core and Guay, 2002 and Coles, Daniel, and Naveen, 2006 are the first papers to use these data; see Coles, Daniel, and Naveen, 2013 for an explanation of their construction.

Murphy's PPS, and we use the logarithm of it as the dependent variable in the regressions in this section. We merge the PPS data with data on manager characteristics from Execucomp and data on firm characteristics from Compustat for the same period.

We use several proxies for growth opportunities. As there is no consensus in the literature on the measurement of growth opportunities, our approach is to use a broad set of several proxies suggested in previous studies and to show that our findings are robust across these proxies. Our first proxy for growth opportunities is the market-to-book ratio. Market value is defined as the market value of equity plus the book value of debt, divided by total assets. A number of studies, including Gompers, 1995, Collins and Kothari, 1989, Korteweg and Polson, 2009, and Zhiguo He, Li, et al. (2014), have used the market-to-book ratio as a proxy for growth options, and previous theoretical work by Berk, Green, and Naik, 1999 and Carlson, Fisher, and Giammarino, 2004 establishes the link between growth options and market-to-book ratios. The use of price data in our proxies is both a blessing and a curse. It is grounded in the assumption that the market incorporates a firm's future investment opportunities into its stock price, thus elevating the market value of a firm's assets beyond the book value of those assets. However, as discussed in Berk, 1995, the potential for mispricing means that it is unsatisfactory to rely solely on this measure. Equally worrying, a relation based on price-based measures can be unrelated to the operating characteristics of the firms, and can instead reflect changes in market risk premiums. Despite these well-founded concerns, previous research by Adam and Goyal, 2008 and Kallapur and Trombley, 1999 has found that the market-to-book ratio performs well as a proxy for growth options and investment opportunities. Nevertheless, we also include several non-price-based growth option proxies.

Our second proxy is the value-to-book ratio, as introduced in Rhodes–Kropf, Robinson, and Viswanathan (2005). This measure attempts to preserve the intuition behind the market-to-book ratio while correcting for potential mispricing by estimating firm value using a regression. Rhodes–Kropf, Robinson, and Viswanathan, 2005 decomposes the market-to-book ratio into three terms: (i) firm-specific mispricing, (ii) industry mispricing, and (iii) value-to-book. However, we use the two-term decomposition found in Lyandres and Zhdanov, 2013: (i) firm-specific, within-industry mispricing, and (ii) value-to-book. We estimate the value of firm i in industry j at time t by performing a within-industry j regression with logarithms of market value M on book value B:

$$\log M_{ijt} = \alpha_{jt} + \beta_{jt} \log B_{ijt} + \varepsilon_{jt}. \tag{2.36}$$

Subtracting the log book value from the fitted value from the regression $\hat{M}_{ijt}$ yields an estimate of log value-to-book. As discussed in Rhodes–Kropf, Robinson, and Viswanathan, 2005, the link between firm value, corrected for mispricings, and book value rests on two assumptions: the first links future returns on equity to future discount rates within industries, and the other assumes that book equity grows at a constant rate. To the extent that these assumptions are unsatisfactory, the value-to-book ratio we use will be an imperfect proxy.

In addition to market-based proxies, we include research and development (R&D) expenditures, an investment-based measure used in Kallapur and Trombley, 1999 and Lyandres and Zhdanov, 2013. We scale R&D expenditures by the book value of assets. In our main analysis, we omit firms with missing R&D expenditures and, as a robustness check, repeat this analysis using all firms, setting R&D expenditures to zero if they are missing. These measures are independent of a firm's price data and are thus uncontaminated by mispricing. The downside is that industry-specific accounting practices restrict the classification of R&D expense, exposing this measure to concerns of a systematic bias that varies by industry. A firm's growth opportunities may include acquisition opportunities or investments in subsidiaries, which are not included in R&D expenses. Kallapur and Trombley, 1999 finds that R&D spending is inconsistently correlated with realized measures of realized growth in a three to five-year horizon, making R&D-based measures a weaker proxy for short-term investment opportunities than the market-to-book ratio, which they find to be a more relevant proxy.

Another set of investment-based measures are based on capital expenditures and following Purnanandam and Rajan, 2017. This measure assumes that capital expenditures correspond to the exercise of growth options and their conversion into physical assets. Like R&D, capital expenditure based measures are independent of a firm's stock price. To account for the fact that a firm's capital expenditure includes maintenance costs for an existing capital base, we calculate our first measure as the residual of a regression of firm CapEx scaled by assets, including a firm fixed effect to capture the predictable investment level of the firm. In terms of regression coefficients, this produces identical estimates to a regression in which capital expenditures is directly used as a regressor. The second measure is the residual from a one-lag auto-regressive model of expected scaled capital expenditures and is thus a better measure of unanticipated capital expenditures.

These proxies are motivated by the fact that a firm's reported capital expenditures might reflect preexisting projects or other ongoing commitments, making the level of capital expenditures a noisy measurement that misrepresents a firm's growth opportunities. By taking the residual, we better capture the discretionary or uncommitted portion of a firm's capital expenditures, which better captures the exercise (and thus reduction) of growth options at the firm. A potential downside of capital expenditure-based measures is that the price of capital is affected by economy-wide demand, and thus the firm's level of capital expenditures is exposed to mispricing at a market- or industry-wide level, albeit in a more indirect way than a measure based on the firm's stock price.

Standard measures of Tobin's q fail to account for intangible capital, which, per accounting rules, is usually expensed rather than capitalized, and thus not found on a firm's balance sheet. The augmented Tobin's q measure of Peters and Taylor, 2017 accounts for firms' intangible assets using an accruals-based accounting approach. In doing so, their measure better captures the market value of firms' assets, and predicts investment better than standard estimates of firm-level q.

Each of our previous proxies captures the presence of growth options but also contains measurement error. We use principal component analysis to extract a statistical measure of growth options and reduce the impact of measurement error. By taking the first principal component, we extract the common variation in these proxies, which we call Hybrid Growth Opportunities. Under the assumption that the other determinants of our proxies are uncorrelated with the true measure of growth options, Hybrid Growth Opportunities better

	Obs.	Mean	Std. Dev.	Min	Max	Median
Jensen & Murphy PPS	182,395	1.070	2.646	0.002	18.858	0.285
$ to % PPS (PPS2)	182,447	197.307	494.850	0.193	3,573.206	45.971
Wealth Performance Sensitivity (PPS3)	35,725	31.252	104.043	0.000	888.708	6.636
Market-to-Book	182,391	1.917	1.293	0.771	8.529	1.473
Value-to-Book	182,432	1.735	0.566	0.956	4.023	1.636
R&D	95,546	0.054	0.069	0.000	0.366	0.027
Total q	154,342	1.336	1.426	0.044	7.899	0.851
Capital Expenditure	175,638	0.054	0.054	0.000	0.294	0.038
Firm Size	182,432	9,897.879	27,705.211	50.598	202,475.000	1,643.600
Cash Flow Volatility	182,447	0.032	0.035	0.002	0.231	0.023
Firm Age	182,447	21.735	13.878	0.000	56.000	19.000
Tangibility	180,163	0.270	0.237	0.003	0.880	0.197
Profitability	180,986	0.126	0.099	-0.242	0.423	0.124
Advertisement	182,447	0.011	0.029	0.000	0.176	0.000
Leverage	181,668	0.223	0.183	0.000	0.820	0.205
Dividend Paying	182,097	0.556	0.497	0.000	1.000	1.000
CEO Chair	115,893	0.589	0.492	0.000	1.000	1.000
Fraction of Inside Directors	115,893	0.284	0.163	0.000	1.000	0.250
CEO	182,447	0.186	0.389	0.000	1.000	0.000
Female	182,447	0.059	0.236	0.000	1.000	0.000

Table 2.1: Summary Statistics. The sample covers all executives and firms in Execucomp from 1992 to 2015 and is merged with Compustat data. Jensen & Murphy PPS is dollar-to-dollar pay-performance sensitivity. Control variables are defined in Appendix B.

capture firms' underlying growth options. The drawback of this approach is that, due to sample limitations, we are limited to firm-year observations for which we have observations of all our growth proxies, thus limiting the sample and our statistical power.

Our sample then includes all firm-executive combinations from ExecuComp from 1992 to 2015. The Execucomp database focuses on largest 1,500 publicly traded companies and has similar industry coverage to the Compustat database. We employ a broad set of standard firm- and manager-level control variables; Appendix B provides their exact definitions. Additionally, we include year and industry dummies (the latter based on the 48 Fama-French industries) to control for time and industrial fixed effect in managerial incentives. We winsorize the continuous variables at the 1st and 99th percentiles. In all the regressions presented below, we lag independent variables by one year (as in, e.g., Zhiguo He, Li, et al., 2014).

2.5.2 Results

We regress Jensen and Murphy's PPS variable on the Market-to-Book variable and various controls for manager and firm characteristics. The results for these regressions are presented in Table 2.2, alongside results for regressions using our other market-based measure of growth options, Value-to-Book. The dependent variable is the logarithm of the dollar-to-dollar pay-performance sensitivity (Jensen and Murphy's PPS). We construct the fixed effects for industry fixed effect using the Fama and French, 1997 48 sectors. The fixed effects in the model in Column (3) are in firm-executive pairs. All of the standard errors are robust and clustered at the firm level. The main effect of interest can be seen in the coefficient on Market-to-Book in Column (3). This coefficient states that a one-standard-deviation change in Market-to-Book is associated with a roughly 5.7% decrease in Jensen and Murphy's PPS. Although the magnitude of the effect on PPS is smaller than that of firm size, this effect is still economically significant.

For our Value-to-Book results, other than the alternative measure of growth options, all of the other controls are identical to those in Columns (1-3). The coefficient in column (6) states that a one standard deviation increase in Value-to-Book is associated with a 1.1% decrease in Jensen and Murphy's PPS. We note that the effect of the value-to-book ratio is statistically significant and of a larger magnitude than our other specifications, and we also find that the quantitative effect associated with a one-standard-deviation change in the value-to-book ratio is significantly stronger when we focus on subsamples of our panel. For example, when we restrict our sample to the 2006-2014 period and still include firm-manager and year fixed effects, we find that a one standard deviation change is associated with a decline of 4.5% in PPS, which is in line with our estimates using the market-to-book ratio as a proxy.

Next, we regress Jensen and Murphy's PPS on R&D and various controls for manager and firm characteristics. The results for these regressions are given in Table 2.3. The dependent variable is the logarithm of the dollar-to-dollar pay-performance sensitivity. All of the other controls are identical to those in Table 2.2. Again the main effect of interest is the coefficient

	Market-to-Book			Value-to-Book		
	(1) log(PPS1)	(2) log(PPS1)	(3) log(PPS1)	(4) log(PPS1)	(5) log(PPS1)	(6) log(PPS1)
Market-to-Book	-0.066^{***}	-0.063^{***}	-0.041^{***}			
	(-8.75)	(-5.85)	(-6.63)			
Value-to-Book				-0.071^{***}	-0.068^{**}	-0.019
				(-3.06)	(-2.41)	(-1.29)
Firm Size	-0.408^{***}	-0.378^{***}	-0.373^{***}	-0.404^{***}	-0.383^{***}	-0.361^{***}
	(-48.50)	(-35.50)	(-18.70)	(-47.40)	(-35.75)	(-18.19)
Cash Flow Volatility		-1.028^{***}	-0.858^{***}		-1.359^{***}	-0.966^{***}
		(-3.50)	(-3.92)		(-4.65)	(-4.38)
Firm Age		-0.087^{***}	-0.317^{***}		-0.077^{***}	-0.302^{***}
		(-4.32)	(-6.17)		(-3.84)	(-5.89)
Tangibility		-0.333^{***}	-0.145		-0.282^{***}	-0.124
		(-3.55)	(-1.28)		(-2.96)	(-1.09)
Profitability		-0.339^{**}	-0.093		-0.742^{***}	-0.257^{***}
		(-2.46)	(-1.03)		(-5.48)	(-2.73)
Advertisement		-0.368	-0.748		-0.473	-0.803
		(-0.69)	(-1.34)		(-0.86)	(-1.43)
Advertisement Missing		0.033	0.011		0.035	0.007
		(1.11)	(0.51)		(1.16)	(0.33)
Leverage		0.496^{***}	0.393^{***}		0.540^{***}	0.420^{***}
		(6.28)	(6.35)		(6.82)	(6.80)
Dividend Paying		-0.170^{***}	-0.137^{***}		-0.169^{***}	-0.139^{***}
		(-5.99)	(-5.28)		(-5.90)	(-5.37)
CEO Chair		0.164^{***}	0.023^{*}		0.165^{***}	0.021^{*}
		(7.95)	(1.86)		(7.93)	(1.76)
Fraction of Inside Directors		0.684^{***}	-0.079		0.681^{***}	-0.081^{*}
		(8.63)	(-1.61)		(8.56)	(-1.66)
CEO		1.745^{***}	0.400^{***}		1.745^{***}	0.400^{***}
		(93.88)	(23.27)		(93.81)	(23.32)
Female		-0.267^{***}			-0.264^{***}	
		(-8.95)			(-8.85)	
Industry Dummies	Yes	Yes	No	Yes	Yes	No
Firm-Manager Dummies	No	No	Yes	No	No	Yes
Year Dummies	Yes	Yes	Yes	Yes	Yes	Yes
Observations	158,278	92,715	92,715	158,309	92,730	92,730
R^2	0.278	0.503	0.121	0.276	0.502	0.119

Table 2.2: Market-based Proxies and Pay-Performance Sensitivity. The sample covers all executives and firms in Execucomp from 1992 to 2015 and is merged with Compustat data. The dependent variable is the logarithm of the dollar-to-dollar pay-performance sensitivity. Market value is defined as the market value of equity plus the book value of debt, divided by total assets. Value-to-book is calculated as the fitted value from a within-industry regression of log market value on log book value, less log book value. Control variables are defined in Appendix B. t statistics based on heteroskedasticity-consistent, and firm-level-clustered standard errors are provided in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

on R&D in column (3). A one standard deviation increase in R&D expenses is associated with a 5.0% decrease in Jensen and Murphy's PPS, which is on the same order of magnitude as our previous regressions. We note that reported R&D expenses, while directly measuring growth opportunities, suffer from relatively low coverage in the Compustat database. We obtain the same results if we take an alternative approach and substitute missing R&D expenses for zero.

As another robustness check, we regress Jensen and Murphy's PPS on the Capital Expenditures variable, along with the same set of controls for manager and firm characteristics. The results of these regressions are in Table 2.4. The dependent variable is again the logarithm of dollar-to-dollar pay sensitivity. The coefficient of 0.391 on Capital Expenditures in column (3) shows that a one standard deviation increase in Capital Expenditures is associated with a 2.3% increase in PPS. Significantly, because capital expenditures represent the exercise of growth options, the expected sign of our estimate is reversed. An increase in growth options leads to a decrease in PPS, and so the exercise of growth options leads to an increase in PPS. We get an estimate of similar magnitude when we replace Capital Expenditures with Capital Expenditure Innovations as a dependent variable. We obtain the innovations from fitting an AR(1) model to a firm's capital expenditures and capturing the unanticipated or discretionary portion of a firm's investments. In situations where a large portion of a firm's investments are recurring or reflect ongoing commitments, it is the incidence of new projects that is informative about the exercise of growth options.

We also regress Jensen and Murphy's PPS on augmented Tobin's q and our hybrid measure of growth opportunities. The results for these regressions are given in Table 2.5. All other controls are identical to those in Table 2.2. The main effect of interest is the coefficient on Tobin's q in column (3) and the coefficient on Hybrid Growth Opportunities in column (6). This coefficient shows that a one standard deviation increase in Tobin's q is associated with a 2.7% decrease in PPS and a one standard deviation increase in our Hybrid measure is associated with a 4.2% decrease in PPS, both of which are consistent with our previous specifications.

Finally, we present results of regressions of Hybrid Growth Opportunities on alterna-

	R&D			R&D (0 if missing)		
	(1) log(PPS1)	(2) log(PPS1)	(3) log(PPS1)	(4) log(PPS1)	(5) log(PPS1)	(6) log(PPS1)
R&D	−0.592***	−0.469	−0.649**			
	(−2.63)	(−1.41)	(−2.40)			
R&D (0 if missing)				−0.643***	−0.437	−0.541*
				(−2.91)	(−1.45)	(−1.93)
Firm Size	−0.430***	−0.414***	−0.361***	−0.401***	−0.380***	−0.366***
	(−42.47)	(−32.85)	(−13.82)	(−46.82)	(−35.36)	(−17.94)
Cash Flow Volatility		−1.483***	−0.792***		−1.297***	−0.947***
		(−4.11)	(−3.07)		(−4.37)	(−4.29)
Firm Age		−0.078***	−0.338***		−0.078***	−0.298***
		(−3.44)	(−5.42)		(−3.87)	(−5.83)
Tangibility		−0.139	0.016		−0.289***	−0.106
		(−1.14)	(0.10)		(−3.03)	(−0.93)
Profitability		−0.604***	−0.310**		−0.794***	−0.278***
		(−3.77)	(−2.57)		(−5.84)	(−2.95)
Advertisement		−0.012	−0.802		−0.528	−0.791
		(−0.02)	(−1.18)		(−0.97)	(−1.41)
Advertisement Missing		0.053	−0.019		0.030	0.008
		(1.47)	(−0.60)		(1.00)	(0.34)
Leverage		0.617***	0.360***		0.534***	0.421***
		(7.16)	(4.38)		(6.63)	(6.84)
Dividend Paying		−0.186***	−0.176***		−0.172***	−0.138***
		(−5.13)	(−5.91)		(−6.02)	(−5.32)
CEO Chair		0.165***	0.010		0.163***	0.021*
		(6.58)	(0.64)		(7.85)	(1.73)
Fraction of Inside Directors		0.515***	−0.099*		0.675***	−0.081*
		(5.25)	(−1.67)		(8.50)	(−1.66)
CEO		1.745***	0.402***		1.745***	0.399***
		(74.98)	(18.25)		(93.81)	(23.29)
Female		−0.256***			−0.265***	
		(−7.56)			(−8.86)	
Industry Dummies	Yes	Yes	No	Yes	Yes	No
Firm-Manager Dummies	No	No	Yes	No	No	Yes
Year Dummies	Yes	Yes	Yes	Yes	Yes	Yes
Observations	82,431	50,088	50,088	158,309	92,730	92,730
R^2	0.280	0.530	0.119	0.276	0.502	0.119

Table 2.3: R&D-based Proxies and Pay-Performance Sensitivity. The sample covers all executives and firms in Execucomp from 1992 to 2015 and is merged with Compustat data. The dependent variable is the logarithm of the dollar-to-dollar pay-performance sensitivity. Control variables are defined in Appendix B. t statistics based on heteroskedasticity-consistent, and firm-level-clustered standard errors are provided in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

	Capex			Capex Innovations		
	(1) log(PPS1)	(2) log(PPS1)	(3) log(PPS1)	(4) log(PPS1)	(5) log(PPS1)	(6) log(PPS1)
Capital Expenditure	0.361	1.659***	0.431***			
	(1.62)	(5.45)	(3.12)			
Capex Innovations				0.459	1.429***	0.155
				(1.58)	(4.65)	(1.34)
Firm Size	−0.397***	−0.378***	−0.360***	−0.387***	−0.371***	−0.340***
	(−46.15)	(−34.88)	(−17.61)	(−38.12)	(−30.65)	(−13.95)
Cash Flow Volatility		−1.524***	−1.009***		−1.225***	−1.008***
		(−5.23)	(−4.57)		(−3.73)	(−3.98)
Firm Age		−0.065***	−0.295***		−0.074***	−0.337***
		(−3.23)	(−5.62)		(−3.29)	(−5.34)
Tangibility		−0.520***	−0.198*		−0.386***	0.012
		(−4.72)	(−1.66)		(−3.66)	(0.09)
Profitability		−0.912***	−0.287***		−0.821***	−0.180*
		(−6.60)	(−3.04)		(−5.53)	(−1.71)
Advertisement		−0.553	−0.859		−0.470	−0.893
		(−1.01)	(−1.51)		(−0.81)	(−1.57)
Advertisement Missing		0.032	−0.000		0.036	0.005
		(1.04)	(−0.00)		(1.06)	(0.16)
Leverage		0.576***	0.437***		0.519***	0.449***
		(7.23)	(6.95)		(6.09)	(6.35)
Dividend Paying		−0.159***	−0.140***		−0.166***	−0.139***
		(−5.54)	(−5.41)		(−5.39)	(−4.90)
CEO Chair		0.163***	0.021*		0.171***	0.037**
		(7.87)	(1.67)		(7.46)	(2.47)
Fraction of Inside Directors		0.680***	−0.075		0.790***	−0.072
		(8.39)	(−1.50)		(8.68)	(−1.28)
CEO		1.749***	0.407***		1.670***	0.415***
		(92.82)	(23.19)		(82.87)	(19.73)
Female		−0.267***			−0.235***	
		(−9.00)			(−6.85)	
Industry Dummies	Yes	Yes	No	Yes	Yes	No
Firm-Manager Dummies	No	No	Yes	No	No	Yes
Year Dummies	Yes	Yes	Yes	Yes	Yes	Yes
Observations	151,830	89,646	89,646	92,500	64,828	64,828
R^2	0.278	0.505	0.119	0.274	0.516	0.114

Table 2.4: Capex-based Proxies and Pay-Performance Sensitivity. The sample covers all executives and firms in Execucomp from 1992 to 2015 and is merged with Compustat data. The dependent variable is the logarithm of the dollar-to-dollar pay-performance sensitivity. Capital Expenditure Innovation is calculated as the residual from a one-lag firm-specific auto-regressive model of expected scaled capital expenditures. Control variables are defined in Appendix B. t statistics based on heteroskedasticity-consistent, and firm-level-clustered standard errors are provided in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

	Total Q			Hybrid		
	(1) log(PPS1)	(2) log(PPS1)	(3) log(PPS1)	(4) log(PPS1)	(5) log(PPS1)	(6) log(PPS1)
Total q	−0.018**	−0.041***	−0.022***			
	(−2.53)	(−3.94)	(−3.45)			
Hybrid Growth Opportunities				−0.053***	−0.061***	−0.033***
				(−6.41)	(−5.55)	(−4.96)
Firm Size	−0.419***	−0.390***	−0.361***	−0.428***	−0.394***	−0.371***
	(−49.07)	(−35.46)	(−17.15)	(−49.13)	(−36.10)	(−17.24)
Cash Flow Volatility		−1.235***	−0.921***		−1.029***	−0.909***
		(−4.04)	(−4.04)		(−3.35)	(−3.97)
Firm Age		−0.075***	−0.324***		−0.073***	−0.324***
		(−3.63)	(−6.25)		(−3.57)	(−6.25)
Tangibility		−0.277***	−0.175		−0.263***	−0.163
		(−2.94)	(−1.57)		(−2.80)	(−1.46)
Profitability		−0.393***	−0.168*		−0.340**	−0.139
		(−2.78)	(−1.75)		(−2.46)	(−1.45)
Advertisement		−0.497	−1.096*		−0.328	−1.056*
		(−0.90)	(−1.81)		(−0.59)	(−1.73)
Advertisement Missing		0.046	−0.013		0.043	−0.007
		(1.47)	(−0.46)		(1.37)	(−0.27)
Leverage		0.560***	0.415***		0.529***	0.406***
		(7.54)	(6.41)		(7.08)	(6.27)
Dividend Paying		−0.184***	−0.154***		−0.186***	−0.154***
		(−6.31)	(−5.95)		(−6.39)	(−5.89)
CEO Chair		0.149***	0.020		0.147***	0.020
		(6.88)	(1.54)		(6.78)	(1.54)
Fraction of Inside Directors		0.695***	−0.067		0.676***	−0.073
		(8.38)	(−1.32)		(8.10)	(−1.44)
CEO		1.745***	0.405***		1.744***	0.406***
		(89.25)	(21.93)		(89.27)	(21.85)
Female		−0.264***			−0.268***	
		(−8.51)			(−8.60)	
Industry Dummies	Yes	Yes	No	Yes	Yes	No
Firm-Manager Dummies	No	No	Yes	No	No	Yes
Year Dummies	Yes	Yes	Yes	Yes	Yes	Yes
Observations	134,062	80,482	80,482	133,092	79,985	79,985
R^2	0.287	0.511	0.123	0.289	0.512	0.124

Table 2.5: Additional Proxies and Pay-Performance Sensitivity. The sample covers all executives and firms in Execucomp from 1992 to 2015 and is merged with Compustat data. The dependent variable is the logarithm of the dollar-to-dollar pay-performance sensitivity. Tobin's q is taken from WRDS based on the methodology of Peters and Taylor (2017). Hybrid Growth Opportunities is calculated as the first principal component of Market-to-Book, Value-to-Book, scaled R&D and scaled Capex. Control variables are defined in Appendix B. t statistics based on heteroskedasticity-consistent, and firm-level-clustered standard errors are provided in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

tive measures of Pay-Performance Sensitivity. These results are given in Table 2.6. Given the differing scales of these measures of PPS, it is more informative to consider the scaled interpretation of the coefficients in Columns (3) and (6), corresponding to regressions on Dollar-to-Percent PPS and Wealth Performance Sensitivity, respectively (Edmans, Gabaix, and Landier, 2008). The economic interpretation of these coefficients is that a one standard deviation increase in Hybrid Growth Opportunities increases the PPS measures by 5.1% and 3.5%, respectively. Consistent with the predictions of our model, the effect of Growth Opportunities on Jensen & Murphy PPS can be negative whereas its effect on percent-based PPS measures is positive.

The presented regression coefficients are from OLS and fixed-effects models. Similar results are obtained from a random-effects model. In Appendix C, we present the results of an analysis in which we address potential biases introduced to the Execucomp database by the inclusion of backfilled data. We find that our results are qualitatively identical and quantitatively larger in magnitude.

2.6 Redeployability and Optimal Incentives

Although our main focus is on investment options, many real options within firms pertain to the optimal time to abandon an ongoing project. In this section, we investigate the implications of abandonment options for the measurement of incentives. Specifically, we consider the redeployability of capital by assuming that at any point the firm can liquidate its existing capital for a price P. For simplicity, we abstract from the growth option and assume that the firm has a fixed capital stock until liquidation. Given this assumption, the problem of providing the manager with incentives is essentially the same as the case we consider in Section 2.3. The optimal contract and firm value are given by the solution to the following Hamilton-Jacobi-Bellman equation for $V(X)$:

$$rV(X) = \max_{a \in [0, a_{\max}]} \left\{ X - (g(a) + \rho(a))X + aXV'(X) + \frac{1}{2}\sigma^2 X^2 V''(X) \right\}, \tag{2.37}$$

	Dollar-to-Percent PPS			Wealth Performance Sensitivity		
	(1) log(PPS2)	(2) log(PPS2)	(3) log(PPS2)	(4) log(PPS3)	(5) log(PPS3)	(6) log(PPS3)
Hybrid Growth Opportunities	0.369***	0.284***	0.149***	0.246***	0.180***	0.037***
	(36.16)	(22.59)	(15.45)	(15.53)	(9.56)	(2.65)
Firm Size	0.513***	0.570***	0.144***	0.053***	0.078***	−0.021
	(60.03)	(53.00)	(5.14)	(3.99)	(4.60)	(−0.55)
Cash Flow Volatility		−1.120***	−0.641**		−1.589***	−0.192
		(−3.23)	(−1.97)		(−3.26)	(−0.36)
Firm Age		−0.085***	−0.303***		−0.085***	−0.255***
		(−3.73)	(−4.61)		(−2.70)	(−2.78)
Tangibility		−0.249***	−0.260*		0.201	−0.155
		(−2.66)	(−1.70)		(1.42)	(−0.83)
Profitability		2.065***	1.006***		1.313***	0.606***
		(12.37)	(7.90)		(5.83)	(3.17)
Advertisement		0.097	−1.988***		−0.553	−2.259**
		(0.16)	(−2.59)		(−0.49)	(−2.02)
Advertisement Missing		0.027	−0.014		−0.058	−0.042
		(0.79)	(−0.39)		(−0.98)	(−0.79)
Leverage		−0.535***	−0.362***		−0.607***	−0.170
		(−6.42)	(−4.50)		(−4.84)	(−1.54)
Dividend Paying		−0.139***	−0.190***		0.006	−0.099**
		(−4.29)	(−5.71)		(0.13)	(−2.09)
CEO Chair		0.163***	0.029*		0.458***	−0.003
		(6.92)	(1.76)		(12.39)	(−0.14)
Fraction of Inside Directors		0.751***	−0.079		1.702***	0.042
		(7.97)	(−1.21)		(11.37)	(0.40)
CEO		1.724***	0.390***		0.713***	0.144***
		(91.95)	(19.32)		(21.96)	(5.55)
Female		−0.288***			−0.413***	
		(−8.65)			(−3.76)	
Industry Dummies	Yes	Yes	No	Yes	Yes	No
Firm-Manager Dummies	No	No	Yes	No	No	Yes
Year Dummies	Yes	Yes	Yes	Yes	Yes	Yes
Observations	133,130	79,997	79,997	27,855	16,630	16,630
R^2	0.301	0.523	0.257	0.165	0.251	0.0781

Table 2.6: Alternative measures of Pay-Performance Sensitivity. The sample covers all executives and firms in Execucomp from 1992 to 2015 and is merged with Compustat data. The dependent variable for Columns (1) through (3) is the logarithm of the dollar-to-percent pay-performance sensitivity. The dependent variable for Columns (4) through (6) is the logarithm of Wealth Performance Sensitivity. Hybrid Growth Opportunities is calculated as the first principal component of Market-to-Book, Value-to-Book, scaled R&D and scaled Capex. Control variables are defined in Appendix B. t statistics based on heteroskedasticity-consistent, and firm-level-clustered standard errors are provided in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

where $\rho(a)$ is the cost of incentives as derived above. Again, as firm value monotonically increases with manager effort a, the optimal abandonment policy will be to liquidate the firm when X crosses some lower boundary $\underline{X}$, pinned down by the following value matching and smooth pasting conditions:

$$V(\underline{X}) = P, \tag{2.38}$$

$$V'(\underline{X}) = 0. \tag{2.39}$$

As X tends to infinity, the probability of abandonment becomes zero. Moreover, the incentive cost of effort grows faster than the increase in cash flow due to effort, and therefore, the optimal effort will tend to zero. Thus, the value function must approach a linear function consistent with zero effort and no growth as X goes to infinity:

$$\lim_{X \to \infty} V'(X) = \frac{1}{r}. \tag{2.40}$$

As in the growth options case, expected-pay-effort sensitivity and pay-performance sensitivity diverge as redeployability increases, so long as incentive costs are more convex than effort costs.

Proposition 3. *As redeployability* increases:

1. *optimal effort $a^*(X)$ and expected-pay-effort sensitivity $\beta^*(X)$* decrease*,*
2. *pay-performance sensitivity $\phi^*(X)$* increases *if the incentive cost of effort is more convex than the direct cost of effort, that is if*

$$\frac{\rho''(a)}{\rho'(a)} > \frac{g''(a)}{g'(a)}.$$

The intuition behind this result is symmetric to that of Proposition 2. As redeployability increases, the marginal benefit of effort decreases and incentives optimally decrease. When incentive costs are more convex than direct effort costs, marginal incentives costs decrease proportionally more than marginal effort costs. As the sum of marginal effort and incentive costs are equal to the marginal benefit of effort, the marginal benefit of effort will decrease by less than the marginal cost of effort. Which, in turn, implies that pay-performance sensitivity increases even though incentives decrease.

Although we do not test our results on redeployability and pay-performance sensitivity to the data, we note that they provide a further dimension along which to determine the empirical validity of our theory. For example, one could examine the relation between changes in industry-level redeployability and pay-performance sensitivity.

2.7 Conclusion

We analyze a model in which an investor needs a manager to operate a firm. In our setting, the investor would like the manager to exert costly effort and grow the firm but is unable to directly observe whether she exerts the recommended effort. To incentivize the recommended effort level, the investor provides the manager with exposure to firm cash flows as part of the manager's compensation package. The investor also has an option to increase the firm's capital level, increasing the effect of the manager's effort on firm value. We characterize the optimal contract between the investor and the manager and analyze the manager's incentives in this setting.

An optimal contract provides the manager with sensitivity to the firm's performance through exposure to unexpected output shocks. Due to the growth option, the manager's expected-pay-effort sensitivity differs from her pay-performance sensitivity. We develop conditions under which decreasing pay-performance sensitivity occurs alongside increasing expected-pay-effort sensitivity, i.e., incentives. We go on to document evidence consistent with our model. Pay-performance sensitivity is strongly and negatively related to proxies for growth options.

Although our model provides clean results on managerial incentives, we acknowledge that a variety of other factors may interact with and complicate real-world manager compensation. In particular, the origin and size of growth options at a firm, which we take as exogenously given, are themselves decisions made by firms and are affected by moral hazard. Further research could explore the multifaceted role of managerial effort in simultaneously creating growth options, increasing growth option size, and increasing firm productivity.

2.8 Appendix: Proofs

Proof. Proof of Lemma 2 Suppose that $\{\tilde{c}, \tilde{a}\}$ solves the manager's problem for a given contract Π and results in zero savings. Further suppose that the manager is endowed with savings $\mathcal{S} > 0$ at time $t \geq 0$. As the manager has CARA preferences, the optimal consumption plan for $s \geq t$ will be $\tilde{c}_s + r\mathcal{S}$, and her effort provision $\tilde{a}_s$ will be unchanged. Thus, an increase in savings from 0 to $\mathcal{S}$ increases the manager's instantaneous utility by a factor of $e^{-\gamma r \mathcal{S}}$ for $s \geq t$. Therefore, we can write the manager's utility for contracts Π and savings $\mathcal{S}$ as follows:

$$W_t\left(\Pi; \mathcal{S}\right) = e^{-\gamma r \mathcal{S}} W_t\left(\Pi; 0\right). \tag{2.41}$$

For the zero-savings condition to hold, it must be the case that

$$u_c\left(\tilde{c}_t, \tilde{a}_t\right) = \frac{\partial}{\partial \mathcal{S}} W_t\left(\Pi; 0\right), \tag{2.42}$$

which implies that $-\gamma u\left(\tilde{c}_t, \tilde{a}_t\right) = -\gamma r\, W_t\left(\Pi; 0\right)$ or $u\left(\tilde{c}_t, \tilde{a}_t\right) = r\, W_t\left(\Pi; 0\right)$. □

Proof. Proof of Lemma 3 and Verification of Incentive Compatibility

We restrict the manager's consumption plan to satisfy the following integrability and transversality conditions:

$$\mathbb{E}\left[\int_0^\infty -e^{-rs} u\left(\tilde{c}_s, \tilde{a}_s\right) ds\right] < \infty \tag{2.43}$$

$$\lim_{t\to\infty} S_t \overset{a.s}{=} 0. \tag{2.44}$$

Consider an arbitrary contract, comprised of the tuple (β_t, a_t, τ), and note that, if W_t solves Equation (2.15), then W_t is equal, by construction, to the manager's continuation utility from choosing savings $S_t = 0$ and effort a_t. Now suppose β_t and a_t satisfy Equation (2.17) and consider an arbitrary consumption and effort policy $(\tilde{c}_t, \tilde{a}_t)$. Let

$$G_t = \int_0^t e^{-rs} u\left(\tilde{c}_s, \tilde{a}_s\right) ds + e^{-rt} e^{-\gamma r S_t} W_t, \tag{2.45}$$

where $S_t = \int_0^t e^r (t-s)(c_s - \tilde{c}_s)\, ds$ is the manager's accumulated savings at the point he chooses the alternative consumption plan. An application of Ito's Lemma gives

$$e^{rt+\gamma r S_t} dG_t = \left(-\gamma r W_t (c_t - \tilde{c}_t) - \gamma r W_t \beta_t (\tilde{a}_t - a_t) X_t + e^{\gamma r S_t} u(\tilde{c}_t, \tilde{a}_t)\right) dt - \gamma r W_t \beta_t dZ_t. \tag{2.46}$$

The $\tilde{c}_t$ and $\tilde{a}_t$ that maximize the drift term above must satisfy the following first-order conditions:

$$\gamma r W_t = -e^{\gamma r S_t} u_c(\tilde{c}_t, \tilde{a}_t), \text{ and} \tag{2.47}$$

$$\gamma r W_t \beta_t X_t = -X_t K_t g'(a) e^{\gamma r S_t} u_c(\tilde{c}_t, \tilde{a}_t) \tag{2.48}$$

as $u_a = -u_c X_t K_t g'(a)$. These first-order conditions are solved for $\tilde{c}_t = c_t + rS_t$ and $\tilde{a}_t = a_t$, as $rW_t = u(c_t, a_t)$. Moreover, for $\tilde{c}_t = c_t + rS_t$ and $\tilde{a}_t = a_t$, the drift term is zero. Thus, for all other choices of consumption and effort, the drift term is weakly negative and G_t is a super-martingale.

Now consider the manager's value from choosing the policy $(\tilde{c}_t, \tilde{a}_t)$.

$$\mathbb{E}\left[\int_0^\infty e^{-rs} u(\tilde{c}_s, \tilde{a}_s)\, ds\right] = \mathbb{E}[G_t] + \mathbb{E}\left[\int_t^\infty e^{-rs} u(\tilde{c}_s, \tilde{a}_s)\, ds - e^{-r(t+\gamma S_t)} W_t\right] \tag{2.49}$$

$$\leq G_0 + \mathbb{E}\left[\int_t^\infty e^{-rs}\left(u(\tilde{c}_s, \tilde{a}_s) - e^{\gamma r S_t} u(c_s, a_s)\right) ds\right]. \tag{2.50}$$

Now note that $\lim_{t\to\infty} S_t \overset{a.s}{=} 0$, so that $\lim_{t\to\infty} |\tilde{c}_t - c_t| \overset{a.s}{=} 0$, which in turn implies that

$$\lim_{t\to\infty} \int_t^\infty e^{-rs}\left(u(\tilde{c}_s, \tilde{a}_s) - e^{\gamma r S_t} u(c_s, a_s)\right) ds \overset{a.s}{=} 0. \tag{2.51}$$

Finally, by the condition given in Equation (2.43) and Fubini's Theorem, we can take the limit as $t \to \infty$ of both sides of Equation (2.50) to get

$$\mathbb{E}\left[\int_0^\infty e^{-rs} u\left(\tilde{c}_s, \tilde{a}_s\right) ds\right] \leq G_0 + \lim_{t\to\infty} \mathbb{E}\left[\int_t^\infty e^{-rs}\left(u\left(\tilde{c}_s, \tilde{a}_s\right) - e^{\gamma r S_t} u\left(c_s, a_s\right)\right) ds\right] \tag{2.52}$$

$$= G_0 = W_0. \tag{2.53}$$

Therefore, all other consumption and effort plans $(\tilde{c}_t, \tilde{a}_t)$ yield no more utility than (c_t, a_t) to the manager, and the contract is an incentive compatible, no savings contract.

The conditions given are necessary for a contract to be no savings by Lemma 2. To see that the conditions are also necessary for incentive compatibility, consider any contract (β_t, a_t, τ) such that β_t does not satisfy the condition given in Equation (2.17), then the same argument given above shows that the optimal response to such a contract would be to choose $\tilde{a}_t \neq a_t$. □

Proof. Proof of Proposition 1

We verify the optimality of the proposed contract with the following steps. In Step 1, we show that we can replace the investor's maximization problem with one in which we maximize a function independent of Y_t. We then assume that the optimal investment policy must be a threshold rule that satisfies the boundary conditions given in Equations (2.28) and (2.29). In Step 2, we consider a fixed investment threshold and verify that the solution to the HJB equations solves the investor's problem for this investment threshold. Finally, we note that we have already verified that the proposed contract is incentive compatible and satisfies the no-savings condition in the proof of Lemma 3. Although the model as presented in the paper assumes a $k = 1$ pre-exercise of the option, we prove the proposition for a general k_s pre-exercise and k_b post-exercise, where $k_b > k_s$.

Before we complete these steps, we make the following technical assumption on β_t:

$$\mathbb{E}\left[\int_0^\infty \beta_t^2 X_t^2 dt\right] < \infty, \tag{2.54}$$

where the expectation is taken with respect to the measure induced by the incentive compatible dynamics of X_t, given β_t. This restriction does not rule out contracts under which the manager has incentives to exert maximal effort forever. However, such contracts would be infinitely costly to implement, so this assumption can be made without loss of generality.

Step 1: Let $v(x,y)$ be the value to the investor under a given incentive-compatible, no-savings contract (c,a,τ) with $X_0 = X$ and $Y_0 = X$, where $Y_0 = -\frac{1}{2}\ln(-\gamma r W_0)$. Note that Lemmas 2 and 3 imply that the compensation process c_t must be given by Equation (2.22). The investor's value is simply the present value of the cash flows of the firm, net of compensation to the manager, and so we have

$$
\begin{aligned}
v(X,Y) &= \mathbb{E}\left[\int_0^\infty e^{-rt}(X_tK_t - c_t)\,dt - e^{-r\tau}P \mid X_0 = X, Y_0 = Y\right] && (2.55)\\
&= \mathbb{E}\left[\int_0^\infty e^{-rt}(X_tK_t(1-g(a_t)) - rY_t)\,dt - e^{-r\tau}P \mid X_0 = X, Y_0 = Y\right] && (2.56)\\
&= \mathbb{E}\left[\int_0^\infty e^{-rt}X_tK_t(1-g(a_t))\,dt - e^{-r\tau}P \mid X_0 = X, Y_0 = Y\right] && (2.57)\\
&\quad + \mathbb{E}\left[re^{-rt}\left(Y_0 + \int_0^t \frac{1}{2}\gamma r\sigma^2\beta_s^2X_s^2ds + \int_0^t \sigma X_t\beta_t dZ_t^u\right)dt \mid X_0 = X, Y_0 = Y\right],
\end{aligned}
$$

where the last line follows from the dynamics of Y_t given in Equation (2.21). Evaluating separately the three terms of the last expectation above, we have

$$
\begin{aligned}
\mathbb{E}\left[\int_0^\infty re^{-rt}Y_0dt\right] &= Y_0,\\
\mathbb{E}\left[\int_0^\infty re^{-rt}\int_0^t \frac{1}{2}\gamma r\sigma^2\beta_s^2X_s^2dsdt\right] &= \mathbb{E}\left[\int_0^\infty\int_s^\infty re^{-rt}\frac{1}{2}\gamma r\sigma^2\beta_s^2X_s^2dtds\right]\\
&= \mathbb{E}\left[\int_0^\infty e^{-rs}\frac{1}{2}\gamma r\sigma^2\beta_s^2X_s^2ds\right],\\
\mathbb{E}\left[\int_0^\infty re^{-rt}\int_0^t \sigma X_t\beta_s dZ_t^u dt\right] &= \int_0^\infty re^{-rt}\sigma\mathbb{E}\left[\int_0^t X_t\beta_s t\,dZ_t^u\right]dt\\
&= 0,
\end{aligned}
$$

where we exchange the order of integration according to Fubini's Theorem and the assumption given in Equation (2.54). Collecting terms gives

$$v(X,Y)=\mathbb{E}\left[\int_0^\infty e^{-rt}\left(X_t\left(K_t-g(a)\right)-\frac{1}{2}\gamma r\sigma^2\beta_t^2X_t^2\right)dt-e^{-r\tau}P\mid X_0=x\right]-Y. \quad (2.58)$$

Thus, the investor's problem is equivalent to the following problem:

$$V(X_0)=\max_{\beta,a,\tau}\mathbb{E}\left[\int_0^\infty e^{-rt}\left(X_t\left(K_t-g(a_t)\right)-\frac{1}{2}\gamma r\sigma^2\beta_t^2\right)dt-e^{-r\tau}P\right], \quad (2.59)$$

such that

$$dX_t=a_tX_tdt+\sigma X_tdZ_t, \quad (2.60)$$

$$K_t=k_s+(k_b-k_s)\,\mathbb{I}(t\geq\tau)\,,\text{ and} \quad (2.61)$$

$$\beta_t=g'(a_t)\,. \quad (2.62)$$

Step 2: Fix an arbitrary investment rule $\hat{\tau}$. Let $\hat{V}$ and $\hat{\beta}_t$ solve

$$r\hat{V}=\max_\beta\left\{\mathcal{L}(X,k,\hat{V};\beta,a)\right\}, \quad (2.63)$$

where

$$\mathcal{L}(X,k,V;\beta,a)=X(k-g(a))-\frac{1}{2}\gamma r\beta^2X^2+aX\frac{dV}{dX}+\frac{1}{2}\sigma^2X^2\frac{d^2V}{dX^2} \quad (2.64)$$

such that

$$\beta=g'(a)\,, \quad (2.65)$$

$$V(X_\tau;K=k_s)\overset{a.s}{=}V(X_\tau;K=k_b)-P, \quad (2.66)$$

and let $\hat{c}_t$ be the compensation given by Equation (2.22) that makes $\hat{a}_t$ incentive compatible. In other words, $\left(\hat{\beta},\hat{a}\right)$ is the optimal contract given investment time $\hat{\tau}$. Now, consider an arbitrary incentive compatible, no-savings contract $\left(\tilde{\beta}_t, \tilde{a}_t\right)$ and let

$$G_t = \int_0^t e^{-rs} \left(\tilde{X}_s \left(\tilde{K}_s - g\left(\tilde{a}_s\right) \right) - \frac{1}{2} \gamma r \sigma \tilde{\beta}_s^2 \tilde{X}_s^2 \right) ds + e^{-rt} \hat{V} \left(\tilde{X}_t, \tilde{K}_t \right) - \mathbb{I}\left(\hat{\tau} \leq t\right) e^{-r\hat{\tau}} P, \tag{2.67}$$

where G_t measures the gains in present value at time $t = 0$, derived from using $\left(\tilde{\beta}_t, \tilde{a}_t, \tau\right)$ up to time t, and $\tilde{X}_t$ and $\tilde{K}_t$ are the productivity and capital induced by the contract $\left(\left\{\tilde{\beta}_t, \tilde{a}_t\right\}, \hat{\tau}\right)$. Using Ito's Lemma gives

$$\begin{aligned} e^{rt} dG_t = & \left(\mathcal{L}\left(\tilde{X}_t, \tilde{K}_t; \tilde{\beta}_t, \tilde{a}_t \right) - r\hat{V} \right) dt + \sigma \tilde{X}_t \frac{d\hat{V}}{dx} dZ_t \\ & + \left(\hat{V}\left(X_t, k_b\right) - \hat{V}\left(X_t, k_s\right) - P \right) d\hat{N}_t, \end{aligned} \tag{2.68}$$

where $d\hat{N}_t = \mathbb{I}\left(t = \hat{\tau}\right)$ is a counting process that measures the arrival of the investment time $\hat{\tau}$. Note that the drift term given in (2.68) is always weakly negative by Equation (2.63), and that the last term of (2.68) is always zero. Therefore, G_t is a super-martingale.

Now, consider the value from choosing the contract $\left(\tilde{\beta}_t, \tilde{a}_t\right)$. We have

$$\mathbb{E}\left[\int_0^\infty \left(\tilde{X}_s \left(\tilde{K}_s - g\left(\tilde{a}_s\right) \right) - \frac{1}{2} \gamma r \sigma^2 \tilde{\beta}_s^2 \tilde{X}_s^2 \right) ds - e^{-r\hat{\tau}} P \right] \tag{2.69}$$

$$= \mathbb{E}\left[G_t\right] + e^{-rt} \mathbb{E}\left[\int_t^\infty e^{-r(s-t)} \left(\tilde{X}_s \left(\tilde{K}_s - g\left(\tilde{a}_s\right) \right) - \frac{1}{2} \gamma r \sigma^2 \tilde{\beta}_s^2 \tilde{X}_s^2 \right) ds - \hat{V}\left(\tilde{X}_t, \tilde{K}_t \right) \right] \tag{2.70}$$

$$\leq G_0 + e^{-rt} \mathbb{E}\left[\int_t^\infty e^{-r(s-t)} \left(\tilde{X}_s \left(\tilde{K}_s - g\left(\tilde{a}_s\right) \right) - \frac{1}{2} \gamma r \sigma^2 \tilde{\beta}_s^2 \tilde{X}_s^2 \right) ds - \hat{V}\left(\tilde{X}_t, \tilde{K}_t \right) \right]. \tag{2.71}$$

Now note that, as $g(\tilde{a}_s) \geq 0$ and $\tilde{\beta}_s^2 \tilde{X}_s^2 > 0$, we have

$$\mathbb{E}\left[\int_t^\infty e^{-r(s-t)} \left(\tilde{X}_s \left(\tilde{K}_s - g(\tilde{a}_s)\right) - \frac{1}{2}\gamma r \sigma^2 \tilde{\beta}_s^2 \tilde{X}_s^2\right) ds\right] \leq \mathbb{E}\left[\int_t^\infty e^{-r(s-t)} \tilde{X}_s \tilde{K}_s ds\right] \tag{2.72}$$

$$\leq \mathbb{E}\left[\int_t^\infty e^{-r(s-t)} \tilde{X}_s k_b ds\right] \tag{2.73}$$

$$\leq \frac{\tilde{X}_t k_b}{r - a_{\max}}, \tag{2.74}$$

where the last inequality states that the firm value is bounded above by the expected present value of the gross (of effort and incentive costs) cash flow $\tilde{X}_t \tilde{K}_t$ achieved when $\tilde{a}_t = a_{\max}$ and $K_t = k_b$ for all t. Next note that

$$\hat{V}(X, k) \geq \frac{Xk}{r} > 0 \tag{2.75}$$

by Equation (2.63). Therefore,

$$\mathbb{E}\left[\int_0^\infty e^{-rs} \left(\tilde{X}_s \left(\tilde{K}_s - g(\tilde{a}_s)\right) - \frac{1}{2}\gamma r \sigma^2 \tilde{\beta}_s^2 \tilde{X}_s^2\right) ds - e^{-r\hat{\tau}} P\right] \leq G_0 + e^{-rt} \mathbb{E}\left[\frac{\tilde{X}_t k}{r - a_{\max}}\right] \tag{2.76}$$

$$\leq G_0 + e^{-(r-1)t} \frac{X_0 k}{r - a_{\max}}, \tag{2.77}$$

where we bound $\mathbb{E}\left[\tilde{X}_t\right]$ above by evaluating the expectation under the assumption of perpetual maximum effort, so that $\tilde{X}_t$ is a geometric Brownian motion. Taking limits of both sides as $t \to \infty$ gives

$$\mathbb{E}\left[\int_0^\infty e^{-rs} \left(\tilde{X}_s \left(\tilde{K}_s - g(\tilde{a}_s)\right) - \frac{1}{2}\gamma r \sigma^2 \tilde{\beta}_s^2 \tilde{X}_s^2\right) ds - e^{-r\hat{\tau}} P\right] \leq G_0 = \hat{V}(X_0, K_0), \tag{2.78}$$

and thus we conclude that any contract $\left(\tilde{\beta}, \tilde{a}, \hat{\tau}\right)$ yields a weakly lower value than the contract $\left(\hat{\beta}, \hat{a}, \hat{\tau}\right)$.

□

Proof. Proof of Proposition 2

We first note that the manager's performance-effort sensitivity is measured by β:

$$\beta^*(X) = g'(a^*(X)), \tag{2.79}$$

and the manager's pay-performance sensitivity is given by

$$\phi^*(X) = \frac{\beta^*(X)}{V'(X)} = \frac{g'(a^*(X))}{V'(X)}. \tag{2.80}$$

Under the optimal contract, the optimal effort policy $a^*(X)$ is given by the first-order condition:

$$-g'(a^*(X)) - \gamma r\sigma^2 g'(a^*(X))g''(a^*(X))X + V'(X) = 0. \tag{2.81}$$

Differentiating the first-order condition with respect to k and rearranging it gives the expression

$$\frac{da^*}{dk} = -\frac{V_{Xk}(X)}{-g''(a^*) - \gamma r\sigma^2 X\left(g''(a^*)^2 + g'(a^*)g'''(a^*)\right)}. \tag{2.82}$$

In the following analysis, we restrict our attention to parameter values such that the optimal $a^*(X)$ satisfies the second-order condition. As the denominator of Equation (2.82) is simply the second derivative of the value function with respect to effort, we find that optimal effort is increasing with the size of the growth option k. We address each measure separately below.

Expected-pay-effort sensitivity We first show that expected-pay-effort sensitivity increases with growth options k. Differentiating the expression for output-based incentives, we have

$$\frac{d\beta^*}{dk} = \sigma X g''(a^*)\frac{da^*}{dk}, \tag{2.83}$$

where, using (2.82), we can see that

$$\text{sign}\left(\frac{d\beta^*}{dk}\right) = \text{sign}\left(\frac{da^*}{dk}\right). \tag{2.84}$$

Recall that the denominator of (2.82) is negative according to our assumption that the second-order condition for the optimality of a^* holds.

Furthermore, we demonstrate that $V_{Xk} > 0$. Beginning with the Hamilton-Jacobi-Bellman equation

$$rV = X - g\left(a^*\left(X\right)\right)X - \frac{1}{2}\gamma r\left(\sigma g'\left(a^*\left(X\right)\right)X\right)^2 + a^*\left(X\right)X\frac{\partial V}{\partial X} + \frac{1}{2}\sigma^2 X^2\frac{\partial^2 V}{\partial X^2}, \tag{2.85}$$

we differentiate with respect to both size, X, and growth option intensity, k, to get

$$\left(r - a^*\left(X\right) - a^*_X\left(X\right)X\right)V_{Xk} = \left(a^*\left(X\right) + \sigma^2\right)XV_{XXk} + \frac{1}{2}\sigma^2 X^2 V_{XXXk}, \tag{2.86}$$

where the Envelope Theorem tells us that the effect of varying k on the optimal effort level $a^*\left(X\right)$ can be ignored when taking the derivative. This result is due to the optimality of a^* and the first-order condition of the Hamilton-Jacobi-Bellman equation.

We invoke a generalized version of the Feynman-Kac formula, provided as Lemma 4 below, to write the function V_{Xk} as the following expectation:

$$\begin{aligned} V_{Xk}\left(X\right) &= E\left[e^{-\int_0^\tau\left(r-a^*(X_t)-a^*_X(X_t)X_t\right)dt}V_{Xk}\left(\overline{X}\right) \mid X_0 = X\right] \\ &= E\left[e^{-\int_0^\tau\left(r-a^*(X_t)-a^*_X(X_t)X_t\right)dt}\frac{\partial^2}{\partial X\partial k}\left.\frac{Xk}{r}\right|_{X=\overline{X}} \mid X_0 = X\right] \\ &= E\left[e^{-\int_0^\tau\left(r-a^*(X_t)-a^*_X(X_t)X_t\right)dt}\frac{1}{r} \mid X_0 = X\right] > 0. \end{aligned} \tag{2.87}$$

With this, we have the result that $V_{Xk} > 0$, and therefore growth options increase expected-pay-effort sensitivity, $\frac{d\beta^*}{dk} > 0$.

Pay-performance sensitivity We can write pay-performance sensitivity as

$$\phi^*(X) = 1 - \frac{\rho'\left(a^*\right)}{V'\left(X\right)}. \tag{2.88}$$

Differentiating Equation (2.88) with respect to k we have

$$\frac{\partial \phi^*(X)}{\partial k} = -\frac{1}{V'(X)^2}\left[\rho''(a^*)\frac{\partial a^*}{\partial k}V'(X) - \rho'(a^*)\frac{\partial V'(X)}{\partial k}\right].$$

Since $V'(X)^2 > 0$, we can ignore the denominator and write

$$\operatorname{sign}\left(\frac{\partial \phi^*}{\partial k}\right) = -\operatorname{sign}\left(\rho''(a^*)\frac{\partial a}{\partial k}V'(X) - \rho'(a^*)\frac{\partial V'(X)}{\partial k}\right). \tag{2.89}$$

Note that differentiating the first-order condition in Equation (2.30) with respect to the size of the growth option k gives

$$g''(a^*)\frac{\partial a^*}{\partial k} + \rho''(a^*)\frac{\partial a^*}{\partial k} = \frac{\partial V'(X)}{\partial k}. \tag{2.90}$$

Thus,

$$\operatorname{sign}\left(\frac{\partial \phi^*}{\partial k}\right) = -\operatorname{sign}\left(\rho''(a^*)\frac{\partial a}{\partial k}V'(X) - \rho'(a^*)\left(g''(a)\frac{\partial a^*}{\partial k} + \rho''(a^*)\frac{\partial a^*}{\partial k}\right)\right). \tag{2.91}$$

where we have substituted (2.90) for the derivative of marginal firm value with respect to the size of the growth option. Canceling and combining like terms, we have

$$\operatorname{sign}\left(\frac{\partial \phi^*}{\partial k}\right) = -\operatorname{sign}\left(\rho''(a^*)\left(V'(X) - \rho'(a^*)\right) - \rho'(a^*)g''(a^*)\right). \tag{2.92}$$

Substituting the first order condition (2.30) allows us to write this condition in terms of the ratio of marginal incentive costs $\rho'(a^*)$ to marginal effort costs $g'(a^*)$

$$\operatorname{sign}\left(\frac{\partial \phi^*}{\partial k}\right) = -\operatorname{sign}\left(g'(a^*)\rho''(a^*) - g''(a^*)\rho'(a^*)\right). \tag{2.93}$$

Thus, $\frac{\partial \phi^*}{\partial k} < 0$ if and only if

$$\frac{\rho''(a)}{\rho'(a)} > \frac{g''(a)}{g'(a)}.$$

Lemma 4. *Suppose that X_t evolves according to $dX_t = \mu(X_t)\,dt + \sigma(X_t)\,dZ_t$. Then, for bounded functions $f : (0, Y] \to \mathbb{R}$, $r : (0, Y] \to \mathbb{R}^+$, and $\Omega : \mathbb{R} \to \mathbb{R}$, a function*

$F : (0, Y] \to \mathbb{R}$ *solves both:*

$$r(X) F(X) = f(X) + \mu(X) F_X(X) + \frac{1}{2}\sigma(X)^2 F_{XX}(X), \tag{2.94}$$

with a boundary condition $F(Y) = \Omega(Y)$ *and*

$$F(X) = E\left[\int_0^{\tau} e^{-\int_0^t r(X_s)ds} f(X_t)\, dt + e^{-\int_0^{\tau} r(X_s)ds}\Omega(Y) \mid X_0 = X\right], \tag{2.95}$$

where $\tau = \inf\{t \geq 0 \mid X_t \geq Y\}$.

Proof. Proof of Lemma 4 The proof essentially follows the proof of Lemma 4 in DeMarzo and Sannikov, 2006. Suppose that V solves equation (2.94) and define a process H_t by:

$$H_t = \int_0^t e^{-\int_0^s r(X_u)du} f(X_s)\, ds + e^{-\int_0^t r(X_s)ds} V(X_s).$$

An application of Ito's formula gives the dynamics for H_t as:

$$\begin{aligned} e^{\int_0^t r(X_s)ds} dH_t = &\left(f(X_t) + \mu(X_t) V_X(X_t) + \frac{1}{2}\sigma(X_t)^2 V_{XX}(X_t) - r(X_t) V(X_t)\right) dt \\ &+ \sigma(X_t) V(X_t)\, dZ_t. \end{aligned}$$

By Equation (2.94), the drift of H_t is zero, and H_t is a martingale. As $V(X)$ is bounded on the interval $\left[0, \overline{X}\right]$, H_τ is a martingale and V satisfies

$$\begin{aligned} V(X_0) = H_0 = E[X_\tau \mid X_0] &= E\left[\int_0^{\tau} e^{-\int_0^t r(X_s)ds} f(X_t)\, dt + e^{-\int_0^{\tau} r(X_s)ds} V(X_\tau) \mid X_0\right] \\ &= E\left[\int_0^{\tau} e^{-\int_0^t r(X_s)ds} f(X_t)\, dt + e^{-\int_0^{\tau} r(X_s)ds} \Omega(Y) \mid X_0\right], \end{aligned}$$

where the last equality follows from the definition of τ as a stopping time, and the boundary condition $V(Y) = \Omega(Y)$. □

□

Proof. Proof of Proposition 3

We prove the proposition by first showing that the cross-derivative of firm value V with respect to productivity X and redeployability P is negative, so that the marginal value of effort is decreasing in redeployability. From there, the proof follows the results of Proposition 2 to show that expected-pay-effort sensitivity decreases in redeployability, whereas pay-performance sensitivity increases.

Let p_1 and p_2 denote two levels of redeployability, with $p_2 > p_1$. The value of a firm with the option to sell its capital for p_2 will always exceed the value of a firm with the inferior option to sell for p_1, so that $V_2 > V_1$. From the lemma below, we can focus our analysis simply on the gap G between V_2 and V_1 for a fixed value of productivity X.

Lemma 5. *If the difference in value between the high value firm and the low value firm is decreasing in X, then the marginal value of effort is decreasing in redeployability.*

Proof. The difference in firm values is given by

$$G(X) \triangleq V_2(X) - V_1(X),$$

so that $G'(X) < 0$ implies $V_1'(X) > V_2'(X)$. As p_1 and p_2 are arbitrary subject to $p_2 > p_1$, $\frac{\partial^2 V}{\partial X \partial P} < 0$. □

We first establish some properties of the function G. V_i is the solution to an ODE, so we know that $G \in \mathcal{C}(2)$. Furthermore, $G(0) = p_2 - p_1$, and $\lim_{X\to\infty} G(X) = 0$. At zero, the value of the firm is given by the redeployability of the firm's capital. As productivity increases and the probability of exercising the option decreases, effort also becomes too expensive, and firm value is simply the perpetuity value of its period cash flows $\frac{X}{r}$, which does not depend on the option to redeploy capital.

We proceed by proof by contradiction. Assume that there is some interval (x_0, x_1) on which G is weakly increasing. As $\lim_{X\to\infty} G(X) = 0$, there must then exist some $x_2 \in [x2, \infty)$ and some positive ε such that $G'(x_2) = 0$ and $G'(x) < 0$ for all $x \in (x_2, x_2 + \varepsilon)$. This means that $G''(x_2) \leq 0$. This is equivalent to $V_1''(x_2) \geq V_2''(x_2)$.

Recall that the HJB equation for firm value was

$$rV(X) = \max_{a \in [0, a_{\max}]} \left\{ X - (g(a) + \rho(a))X + aXV'(X) + \frac{1}{2}\sigma^2 X^2 V''(X) \right\}.$$

The second derivative of firm value V'' does not depend upon a, and so the optimal level of effort a^* is a function of only V' and X. Therefore, if $V_1'(x_2) = V_2'(x_2)$, then $a_1^*(x_2) = a_2^*(x_2)$ and both firms will recommend the same level of effort for the manager at $X = x_2$. Given that both firms choose the same level of effort and $V_1'' > V_2''$, the HJB equation implies that $V_1(x_2) \geq V_2(x_2)$. However, this is a contradiction of the fact that firm value is increasing in redeployability $\frac{\partial V}{\partial P} > 0$, so G must be strictly decreasing.

From here, the proof is identical to the proof of Proposition 2, in that we use $\frac{\partial^2 V}{\partial X \partial P}$ to sign the derivative of optimal effort a^* with respect to redeployability P. Using Equation (2.82), we have that effort is decreasing in redeployability. Then, by Equation (2.84), the derivative of expected-pay-effort sensitivity β^* has the same sign as $\frac{\partial a^*}{\partial P}$, $\frac{\partial \beta^*}{\partial P} < 0$ and expected-pay-effort sensitivity is decreasing in redeployability.

The sign of the derivative of pay-performance sensitivity is given by

$$\text{sign}\left(\frac{d\phi^*}{dk}\right) = \text{sign}\left(-g'''(a^*)\frac{da^*}{dk}\right).$$

When $g'''(a^*) > 0$, then $\frac{\partial \phi^*}{\partial P} > 0$. This corresponds to the case in which incentive costs are more convex than effort costs, so that pay-performance sensitivity is increasing in redeployability, completing the proof. □

2.9 Appendix: Definitions of Variables

Advertisement. This variable is advertising expense/total assets = XAD/AT. *Advertisement Missing* is an indicator variable for whether this measure was missing data.

Capital Expenditures. This variable is capital expenditures/total assets = CAPX/AT.

CEO. This variable is an indicator variable for whether the manager in question is the CEO of the firm.

CEO Chair. This variable is an indicator variable for whether the CEO is also chairman of the board.

Dividend Paying. This variable is an indicator variable for whether dividends on common stock (DVC) is strictly positive.

Female. This variable is an indicator for whether the manager is female.

Firm Age. This variable equals the year of the data entry less the first year the firm appeared in the CRSP database.

Firm Size. This variable is the natural log of total assets = log(AT).

Fraction of Inside Directors. This variable is the number of inside board directors divided by board size. Inside directors are those who personally or had a family member serve as a current or former firm manager.

Leverage. This variable is (long term debt + short term debt)/total assets = (DLTT + DLC)/AT.

Market-to-Book. This variable equals (market value of equity + book value of debt)/book value of assets = (CSHO × PRCC_F + AT - CEQ)/AT.

Profitability. This variable is operating income before D&A/total assets = OIVDP/AT.

R&D. This variable equals R&D expense/book value of assets = XRD/AT.

Tangibility. This variable equals net PP&E/total assets = PPENT/AT.

Tobin's q. This variable is the Peters-Taylor measure of total Tobin's q found on WRDS = Q_TOT.

Value-to-Book. First, we regress log(market value of equity plus book value of debt) = log(CSHO × PRCC_F + AT - CEQ) on log(book value) of assets (log(AT)), including an industry fixed effect, where industry is determined by four-digit SIC codes. Second, we subtract log book value of assets (log(AT)) from the fitted values from the regression.

2.10 Appendix: Accounting for Biases in the Execucomp Dataset

In this appendix, we address concerns of selection and bias in our dataset. Our dataset consists of a merge between Compustat, which covers all public firms, and Execucomp, which primarily covers larger public firms. In Figure 2.2, we plot the distribution of Fama-French 48 industries for both Compustat as a whole and our merged dataset. We see that the distribution of industry coverage does not differ significantly with the exception of Pharmaceuticals and Trading. These firms tend to be smaller than other public firms, and thus are systematically underrepresented in Execucomp relative to the universe of public firms.

Another potential source of bias stems from the practice of backfillings data in Execucomp. As discussed in Gillan et al. (2017), the habit of including backfilled data means that ex-post successful firms are overrepresented in the data, as they are added onto indices if clients of S&P request the data be added. This practice of backfilling ceased after 2006 due to changes in the regulatory environment.

A natural test would be to perform our regressions on our entire dataset, excluding those observations that were included due to backfilling. However, based on the dataset provided by Gillan et al. (2017)[8], virtually all of compensation data from the 1994-2005 period is backfilled. Therefore, we instead restrict our sample to the post-backfilling period from 2006 onwards, and find that our qualitative results are unchanged. Summary statistics for the restricted sample are reported in Table 2.7. The results are reported in Tables 2.8-2.12. We find that in the latter part of the sample, a one standard deviation increase in the market-to-book ratio is associated with a 6.75% decrease in Jensen and Murphy's PPS. This is slightly larger than the 5.7% decrease we estimate over the full sample, but still of a similar magnitude. Our estimates using only post-2005 data are of similar magnitude to those corresponding to the full sample. Importantly, the sign of our coefficient estimates do not change, and remain consistent with the direction predicted by theory.

[8]Available at Andrew Koch's website: http://www.pitt.edu/~awkoch/

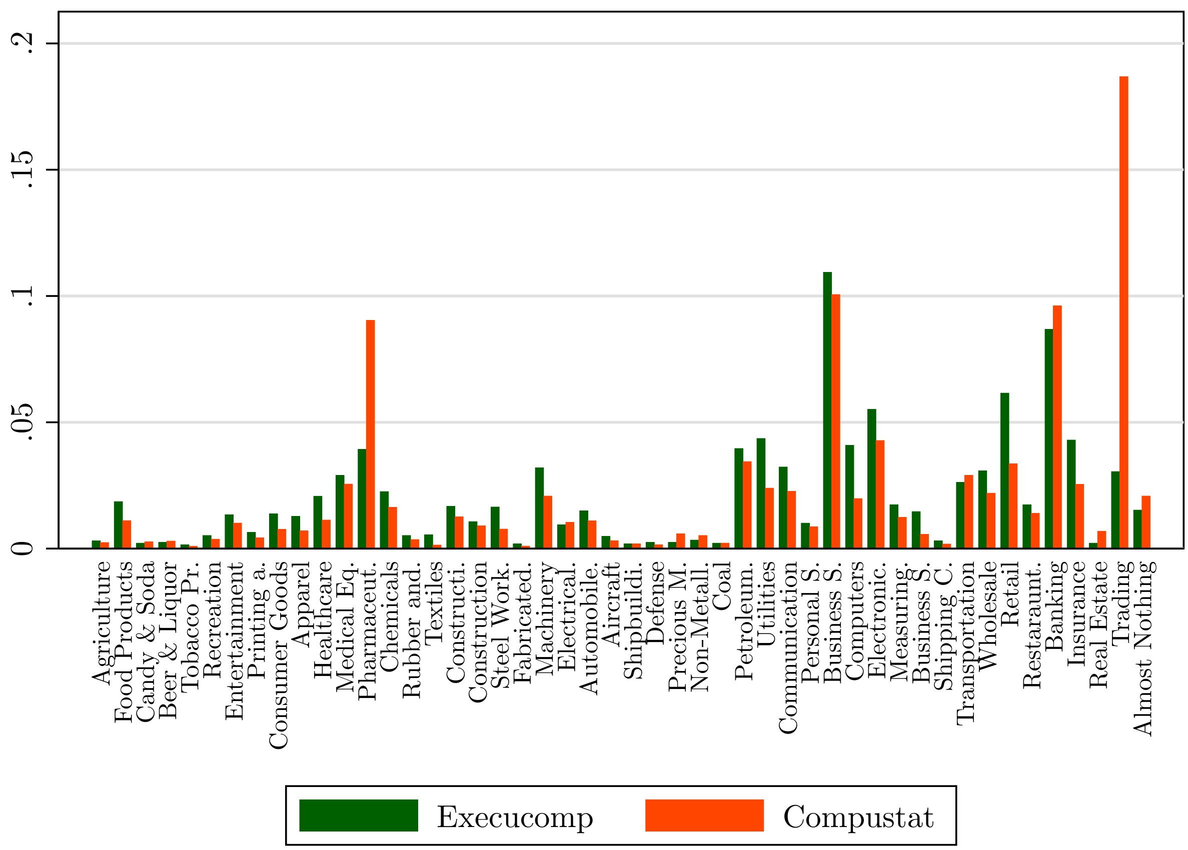


Figure 2.2: Fama-French 48 Industry Coverage. The representation of each of the 48 Fama-French Industries is presented for both the Execucomp and Compustat databases. While Trading and Pharmaceutical firms represent a larger proportion of Compustat than of Execucomp, we attribute this to Execucomp's focus on larger firms.

	Obs.	Mean	Std. Dev.	Min	Max	Median
Jensen & Murphy PPS	75,829	0.749	2.046	0.002	18.858	0.208
$ to % PPS (PPS2)	75,844	176.999	452.758	0.193	3,573.206	42.791
Wealth Performance Sensitivity (PPS3)	14,547	17.667	65.274	0.000	888.708	4.988
Market-to-Book	75,826	1.810	1.125	0.771	8.529	1.443
Value-to-Book	75,840	1.646	0.454	0.956	4.023	1.599
R&D	40,141	0.050	0.065	0.000	0.366	0.024
Total q	64,149	1.214	1.267	0.044	7.899	0.824
Capital Expenditure	75,721	0.044	0.050	0.000	0.294	0.029
Firm Size	75,840	12,141	31,095	50.598	202,475	2,306
Cash Flow Volatility	75,844	0.035	0.040	0.002	0.231	0.023
Firm Age	75,844	24.463	14.989	0.000	56.000	21.000
Tangibility	74,844	0.239	0.233	0.003	0.880	0.155
Profitability	75,488	0.122	0.096	-0.242	0.423	0.119
Advertisement	75,844	0.011	0.029	0.000	0.176	0.000
Leverage	75,498	0.216	0.185	0.000	0.820	0.192
Dividend Paying	75,724	0.543	0.498	0.000	1.000	1.000
CEO Chair	56,518	0.504	0.500	0.000	1.000	1.000
Fraction of Inside Directors	56,518	0.215	0.114	0.000	1.000	0.200
CEO	75,844	0.192	0.394	0.000	1.000	0.000
Female	75,844	0.081	0.273	0.000	1.000	0.000

Table 2.7: Summary Statistics. The sample covers all executives and firms in Execucomp from 2006 to 2015 and is merged with Compustat data. Jensen & Murphy PPS is the dollar-to-dollar pay-performance sensitivity. Control variables are defined in Appendix B.

	Market-to-Book			Value-to-Book		
	(1) log(PPS1)	(2) log(PPS1)	(3) log(PPS1)	(4) log(PPS1)	(5) log(PPS1)	(6) log(PPS1)
Market-to-Book	-0.103^{***}	-0.084^{***}	-0.059^{***}			
	(-7.94)	(-4.53)	(-5.27)			
Value-to-Book				-0.211^{***}	-0.090	-0.090^{**}
				(-3.82)	(-1.41)	(-2.48)
Firm Size	-0.407^{***}	-0.406^{***}	-0.419^{***}	-0.409^{***}	-0.407^{***}	-0.408^{***}
	(-38.19)	(-29.86)	(-13.69)	(-37.22)	(-29.57)	(-13.43)
Cash Flow Volatility		-0.836^{**}	-0.410		-1.204^{***}	-0.535^{**}
		(-2.00)	(-1.54)		(-2.88)	(-2.02)
Firm Age		-0.046^{*}	-0.200^{***}		-0.037	-0.184^{***}
		(-1.72)	(-2.84)		(-1.40)	(-2.66)
Tangibility		-0.199^{*}	0.133		-0.136	0.156
		(-1.72)	(0.87)		(-1.16)	(1.02)
Profitability		-0.424^{**}	-0.034		-0.977^{***}	-0.194
		(-1.99)	(-0.29)		(-5.22)	(-1.65)
Advertisement		0.248	-0.975		0.211	-0.997
		(0.39)	(-0.94)		(0.33)	(-0.96)
Advertisement Missing		0.040	0.011		0.044	0.009
		(1.03)	(0.29)		(1.15)	(0.24)
Leverage		0.565^{***}	0.312^{***}		0.608^{***}	0.337^{***}
		(5.70)	(3.67)		(6.14)	(3.98)
Dividend Paying		-0.104^{***}	-0.122^{***}		-0.107^{***}	-0.127^{***}
		(-2.83)	(-3.31)		(-2.90)	(-3.45)
CEO Chair		0.229^{***}	0.028^{*}		0.229^{***}	0.029^{*}
		(8.32)	(1.69)		(8.27)	(1.73)
Fraction of Inside Directors		0.910^{***}	-0.081		0.901^{***}	-0.076
		(6.78)	(-1.05)		(6.64)	(-0.97)
CEO		1.736^{***}	0.365^{***}		1.737^{***}	0.365^{***}
		(81.00)	(15.67)		(80.86)	(15.67)
Female		-0.269^{***}			-0.263^{***}	
		(-8.31)			(-8.09)	
Industry Dummies	Yes	Yes	No	Yes	Yes	No
Firm-Manager Dummies	No	No	Yes	No	No	Yes
Year Dummies	Yes	Yes	Yes	Yes	Yes	Yes
Observations	63,307	45,759	45,759	63,317	45,764	45,764
R^2	0.244	0.495	0.0811	0.240	0.493	0.0789

Table 2.8: Market-based Proxies and Pay-Performance Sensitivity, Backfill-bias free sample. The sample covers all executives and firms in Execucomp from 2006 to 2015 and is merged with Compustat data. The dependent variable is the logarithm of the dollar-to-dollar pay-performance sensitivity. Market value is defined as the market value of equity plus the book value of debt, divided by total assets. Value-to-book is calculated as the fitted value from a within-industry regression of log market value on log book value, less log book value. Control variables are defined in Appendix B. t statistics based on heteroskedasticity-consistent and firm-level-clustered standard errors are provided in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

	R&D			R&D (0 if missing)		
	(1) log(PPS1)	(2) log(PPS1)	(3) log(PPS1)	(4) log(PPS1)	(5) log(PPS1)	(6) log(PPS1)
R&D	-0.676^{*}	-0.157	-0.723			
	(-1.72)	(-0.30)	(-1.62)			
R&D (0 if missing)				-0.609	-0.170	-0.495
				(-1.61)	(-0.36)	(-1.06)
Firm Size	-0.418^{***}	-0.435^{***}	-0.405^{***}	-0.397^{***}	-0.403^{***}	-0.406^{***}
	(-29.07)	(-25.64)	(-9.47)	(-36.27)	(-29.33)	(-12.78)
Cash Flow Volatility		-1.379^{**}	-0.228		-1.205^{***}	-0.551^{**}
		(-2.49)	(-0.71)		(-2.86)	(-2.07)
Firm Age		-0.028	-0.294^{***}		-0.038	-0.177^{**}
		(-0.84)	(-3.15)		(-1.41)	(-2.56)
Tangibility		-0.002	0.116		-0.135	0.189
		(-0.01)	(0.53)		(-1.15)	(1.23)
Profitability		-0.884^{***}	-0.486^{***}		-1.015^{***}	-0.221^{*}
		(-3.66)	(-2.96)		(-5.41)	(-1.87)
Advertisement		0.271	-1.689		0.156	-0.907
		(0.35)	(-1.20)		(0.24)	(-0.89)
Advertisement Missing		0.058	0.024		0.043	0.009
		(1.23)	(0.44)		(1.11)	(0.24)
Leverage		0.790^{***}	0.209^{*}		0.605^{***}	0.338^{***}
		(6.55)	(1.94)		(6.06)	(4.01)
Dividend Paying		-0.148^{***}	-0.194^{***}		-0.108^{***}	-0.125^{***}
		(-2.95)	(-4.42)		(-2.93)	(-3.40)
CEO Chair		0.235^{***}	-0.009		0.229^{***}	0.028^{*}
		(6.50)	(-0.40)		(8.26)	(1.69)
Fraction of Inside Directors		0.652^{***}	-0.175		0.903^{***}	-0.073
		(3.59)	(-1.62)		(6.67)	(-0.94)
CEO		1.746^{***}	0.400^{***}		1.737^{***}	0.365^{***}
		(62.73)	(13.84)		(80.87)	(15.67)
Female		-0.220^{***}			-0.262^{***}	
		(-5.16)			(-8.07)	
Industry Dummies	Yes	Yes	No	Yes	Yes	No
Firm-Manager Dummies	No	No	Yes	No	No	Yes
Year Dummies	Yes	Yes	Yes	Yes	Yes	Yes
Observations	33,195	24,648	24,648	63,317	45,764	45,764
R^2	0.253	0.520	0.0962	0.240	0.493	0.0784

Table 2.9: R&D-based Proxies and Pay-Performance Sensitivity, Backfill-bias free sample. The sample covers all executives and firms in Execucomp from 2006 to 2015 and is merged with Compustat data. The dependent variable is the logarithm of the dollar-to-dollar pay-performance sensitivity. Control variables are defined in Appendix B. t statistics based on heteroskedasticity-consistent and firm-level-clustered standard errors are provided in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

	Capex			Capex Innovations		
	(1) log(PPS1)	(2) log(PPS1)	(3) log(PPS1)	(4) log(PPS1)	(5) log(PPS1)	(6) log(PPS1)
Capital Expenditure	0.047	1.640***	0.262			
	(0.15)	(4.03)	(1.25)			
Capex Innovations				0.213	1.665***	0.054
				(0.49)	(3.20)	(0.22)
Firm Size	−0.394***	−0.400***	−0.400***	−0.407***	−0.406***	−0.482***
	(−36.69)	(−29.34)	(−13.01)	(−30.40)	(−25.06)	(−11.58)
Cash Flow Volatility		−1.315***	−0.538**		−0.980**	−0.136
		(−3.16)	(−2.02)		(−2.10)	(−0.37)
Firm Age		−0.032	−0.174**		−0.036	−0.195**
		(−1.21)	(−2.52)		(−1.11)	(−2.01)
Tangibility		−0.384***	0.123		−0.231	−0.170
		(−2.78)	(0.74)		(−1.57)	(−0.81)
Profitability		−1.146***	−0.251**		−1.327***	−0.286**
		(−5.98)	(−2.10)		(−6.25)	(−2.00)
Advertisement		0.120	−0.942		0.434	−1.749**
		(0.19)	(−0.92)		(0.58)	(−2.50)
Advertisement Missing		0.045	0.010		0.031	−0.033
		(1.17)	(0.27)		(0.70)	(−0.93)
Leverage		0.644***	0.340***		0.621***	0.475***
		(6.49)	(4.02)		(5.45)	(4.54)
Dividend Paying		−0.095**	−0.127***		−0.101**	−0.154***
		(−2.57)	(−3.44)		(−2.42)	(−3.53)
CEO Chair		0.229***	0.027		0.256***	0.018
		(8.33)	(1.64)		(7.74)	(0.81)
Fraction of Inside Directors		0.907***	−0.070		1.064***	−0.184*
		(6.74)	(−0.89)		(6.00)	(−1.65)
CEO		1.738***	0.367***		1.697***	0.305***
		(80.84)	(15.75)		(67.86)	(10.31)
Female		−0.259***			−0.239***	
		(−7.93)			(−6.00)	
Industry Dummies	Yes	Yes	No	Yes	Yes	No
Firm-Manager Dummies	No	No	Yes	No	No	Yes
Year Dummies	Yes	Yes	Yes	Yes	Yes	Yes
Observations	63,220	45,680	45,680	33,845	26,494	26,494
R^2	0.239	0.494	0.0780	0.248	0.505	0.0665

Table 2.10: Capex-based Proxies and Pay-Performance Sensitivity, Backfill-bias free sample. The sample covers all executives and firms in Execucomp from 2006 to 2015 and is merged with Compustat data. The dependent variable is the logarithm of the dollar-to-dollar pay-performance sensitivity. Capital Expenditure Innovation is calculated as the residual from a one-lag firm-specific auto-regressive model of expected scaled capital expenditures. Control variables are defined in Appendix B. t statistics based on heteroskedasticity-consistent and firm-level-clustered standard errors are provided in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

	Total Q			Hybrid		
	(1) log(PPS1)	(2) log(PPS1)	(3) log(PPS1)	(4) log(PPS1)	(5) log(PPS1)	(6) log(PPS1)
Total q	−0.061***	−0.065***	−0.022**			
	(−4.96)	(−3.79)	(−2.13)			
Hybrid Growth Opportunities				−0.102***	−0.083***	−0.045***
				(−6.98)	(−4.05)	(−3.71)
Firm Size	−0.414***	−0.414***	−0.405***	−0.428***	−0.420***	−0.420***
	(−36.90)	(−28.96)	(−12.63)	(−37.57)	(−29.36)	(−13.10)
Cash Flow Volatility		−0.906**	−0.423		−0.700	−0.360
		(−2.08)	(−1.48)		(−1.59)	(−1.26)
Firm Age		−0.040	−0.139*		−0.037	−0.139*
		(−1.45)	(−1.89)		(−1.35)	(−1.89)
Tangibility		−0.198*	0.099		−0.192*	0.086
		(−1.70)	(0.65)		(−1.65)	(0.57)
Profitability		−0.409*	−0.195		−0.377*	−0.137
		(−1.89)	(−1.52)		(−1.73)	(−1.09)
Advertisement		0.307	−1.286		0.493	−1.291
		(0.50)	(−1.15)		(0.79)	(−1.15)
Advertisement Missing		0.053	−0.003		0.047	−0.002
		(1.33)	(−0.07)		(1.17)	(−0.04)
Leverage		0.627***	0.313***		0.607***	0.306***
		(6.24)	(3.56)		(6.00)	(3.48)
Dividend Paying		−0.126***	−0.153***		−0.131***	−0.152***
		(−3.30)	(−4.38)		(−3.43)	(−4.35)
CEO Chair		0.205***	0.026		0.201***	0.026
		(7.11)	(1.48)		(6.95)	(1.48)
Fraction of Inside Directors		0.847***	−0.083		0.813***	−0.086
		(5.96)	(−1.00)		(5.72)	(−1.03)
CEO		1.746***	0.364***		1.746***	0.363***
		(75.72)	(14.46)		(75.74)	(14.44)
Female		−0.264***			−0.265***	
		(−7.51)			(−7.58)	
Industry Dummies	Yes	Yes	No	Yes	Yes	No
Firm-Manager Dummies	No	No	Yes	No	No	Yes
Year Dummies	Yes	Yes	Yes	Yes	Yes	Yes
Observations	53,591	39,645	39,645	53,575	39,636	39,636
R^2	0.253	0.504	0.0801	0.255	0.504	0.0813

Table 2.11: Additional Proxies and Pay-Performance Sensitivity, Backfill-bias free sample. The sample covers all executives and firms in Execucomp from 2006 to 2015 and is merged with Compustat data. The dependent variable is the logarithm of the dollar-to-dollar pay-performance sensitivity. Tobin's q is taken from WRDS based on the methodology of Peters and Taylor (2017). Hybrid Growth Opportunities is calculated as the first principal component of Market-to-Book, Value-to-Book, scaled R&D, and scaled Capex. Control variables are defined in Appendix B. t statistics based on heteroskedasticity-consistent and firm-level-clustered standard errors are provided in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

	Dollar-to-Percent PPS			Wealth Performance Sensitivity		
	(1) log(PPS2)	(2) log(PPS2)	(3) log(PPS2)	(4) log(PPS3)	(5) log(PPS3)	(6) log(PPS3)
Hybrid Growth Opportunities	0.385***	0.309***	0.141***	0.282***	0.226***	0.082***
	(21.48)	(12.99)	(7.45)	(11.90)	(7.83)	(3.36)
Firm Size	0.530***	0.548***	−0.017	0.066***	0.067***	−0.171***
	(44.60)	(37.28)	(−0.43)	(3.86)	(3.23)	(−2.93)
Cash Flow Volatility		−1.327***	−0.690		−1.669**	−0.734
		(−2.69)	(−1.49)		(−2.48)	(−0.98)
Firm Age		−0.035	−0.101		−0.041	−0.187
		(−1.18)	(−1.17)		(−1.09)	(−1.35)
Tangibility		−0.202*	−0.394*		0.171	−0.683**
		(−1.70)	(−1.76)		(0.97)	(−2.25)
Profitability		1.769***	0.481***		1.249***	0.013
		(6.82)	(2.93)		(3.81)	(0.05)
Advertisement		0.690	−2.446		−1.521	−2.160
		(0.94)	(−1.61)		(−1.22)	(−0.92)
Advertisement Missing		0.034	−0.032		−0.129*	−0.166*
		(0.77)	(−0.57)		(−1.89)	(−1.76)
Leverage		−0.329***	−0.206**		−0.377**	−0.160
		(−3.05)	(−1.97)		(−2.44)	(−0.99)
Dividend Paying		−0.098**	−0.151***		0.036	−0.104*
		(−2.28)	(−3.51)		(0.64)	(−1.84)
CEO Chair		0.221***	0.024		0.510***	−0.012
		(7.05)	(1.01)		(10.57)	(−0.31)
Fraction of Inside Directors		0.814***	−0.141		1.758***	0.059
		(5.20)	(−1.32)		(7.25)	(0.37)
CEO		1.735***	0.355***		0.791***	0.170***
		(77.49)	(12.43)		(16.97)	(4.30)
Female		−0.281***			−0.361***	
		(−7.42)			(−3.04)	
Industry Dummies	Yes	Yes	No	Yes	Yes	No
Firm-Manager Dummies	No	No	Yes	No	No	Yes
Year Dummies	Yes	Yes	Yes	Yes	Yes	Yes
Observations	53,583	39,640	39,640	11,604	8,524	8,524
R^2	0.293	0.510	0.242	0.113	0.220	0.105

Table 2.12: Alternative Measures of Pay-Performance Sensitivity, Backfill-bias free sample. The sample covers all executives and firms in Execucomp from 2006 to 2015 and is merged with Compustat data. The dependent variable for Columns (1) through (3) is the logarithm of the dollar-to-percent pay-performance sensitivity. The dependent variable for Columns (4) through (6) is the logarithm of Wealth Performance Sensitivity. Hybrid Growth Opportunities is calculated as the first principal component of Market-to-Book, Value-to-Book, scaled R&D, and scaled Capex. Control variables are defined in Appendix B. t statistics based on heteroskedasticity-consistent and firm-level-clustered standard errors are provided in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

2.11 Bibliography

Adam, Tim and Vidhan K Goyal (2008). "The investment opportunity set and its proxy variables". *Journal of Financial Research* 31.1, pp. 41–63.

Babenko, Ilona (2009). "Share repurchases and pay-performance sensitivity of employee compensation contracts". *The Journal of Finance* 64.1, pp. 117–150.

Baker, George P and Brian J Hall (2004). "CEO incentives and firm size". *Journal of Labor Economics* 22.4, pp. 767–798.

Baker, George P, Michael C Jensen, and Kevin J Murphy (1988). "Compensation and incentives: Practice vs. theory". *The journal of Finance* 43.3, pp. 593–616.

Becker, Bo (2006). "Wealth and executive compensation". *The Journal of Finance* 61.1, pp. 379–397.

Bergstresser, Daniel and Thomas Philippon (2006). "CEO incentives and earnings management". *Journal of Financial Economics* 80.3, pp. 511–529.

Berk, Jonathan B (1995). "A critique of size-related anomalies". *Review of Financial Studies* 8.2, pp. 275–286.

Berk, Jonathan B, Richard C Green, and Vasant Naik (1999). "Optimal investment, growth options, and security returns". *The Journal of Finance* 54.5, pp. 1553–1607.

Bolton, Patrick, Tano Santos, and Jose A Scheinkman (2016). "Cream-Skimming in Financial Markets". *The Journal of Finance* 71.2, pp. 709–736.

Brennan, Michael J and Eduardo S Schwartz (1985). "Evaluating Natural Resource". *The Journal of Business* 58.2, pp. 135–157.

Carlson, Murray, Adlai Fisher, and Ron Giammarino (2004). "Corporate investment and asset price dynamics: implications for the cross-section of returns". *The Journal of Finance* 59.6, pp. 2577–2603.

Chaigneau, Pierre, Alex Edmans, and Daniel Gottlieb (2014). *The value of informativeness for contracting*. Tech. rep. National Bureau of Economic Research.

Chava, Sudheer and Amiyatosh Purnanandam (2010). "CEOs versus CFOs: Incentives and corporate policies". *Journal of Financial Economics* 97.2, pp. 263–278.

Cheng, Ing-Haw, Harrison Hong, and Jose A Scheinkman (2015). "Yesterday's heroes: compensation and risk at financial firms". *The Journal of Finance* 70.2, pp. 839–879.

Coles, Jeffrey L, Naveen D Daniel, and Lalitha Naveen (2006). "Managerial incentives and risk-taking". *Journal of Financial Economics* 79.2, pp. 431–468.

— (2013). *Calculation of compensation incentives and firm-related wealth using Execucomp: Data, program, and explanation.* Working Paper.

Collins, Daniel W and SP Kothari (1989). "An analysis of intertemporal and cross-sectional determinants of earnings response coefficients". *Journal of Accounting and Economics* 11.2-3, pp. 143–181.

Cong, Lin William (2017). *Auctions of Real Options.* Working Paper, University of Chicago.

Core, John and Wayne Guay (2002). "Estimating the value of employee stock option portfolios and their sensitivities to price and volatility". *Journal of Accounting Research* 40.3, pp. 613–630.

Deckop, John R (1988). "Determinants of chief executive officer compensation". *ILR Review* 41.2, pp. 215–226.

DeMarzo, Peter M and Yuliy Sannikov (2006). "Optimal Security Design and Dynamic Capital Structure in a Continuous-Time Agency Model". *The Journal of Finance* 61.6, pp. 2681–2724.

Edmans, Alex and Xavier Gabaix (2011). "The effect of risk on the CEO market". *Review of Financial Studies* 24.8, pp. 2822–2863.

Edmans, Alex, Xavier Gabaix, and Augustin Landier (2008). "A multiplicative model of optimal CEO incentives in market equilibrium". *The Review of Financial Studies* 22.12, pp. 4881–4917.

Edmans, Alex, Xavier Gabaix, Tomasz Sadzik, et al. (2012). "Dynamic CEO compensation". *The Journal of Finance* 67.5, pp. 1603–1647.

Fama, Eugene F and Kenneth R French (1997). "Industry costs of equity". *Journal of financial economics* 43.2, pp. 153–193.

Frydman, Carola and Dirk Jenter (2010). "CEO Compensation". *Annual Review of Financial Economics* 2.1, pp. 75–102.

Gabaix, Xavier and Augustin Landier (2008). "Why has CEO Pay Increased So Much?" *The Quarterly Journal of Economics* 123.1, pp. 49–100.

Gillan, Stuart et al. (2017). "Getting the incentives right: Backfilling and biases in executive compensation data". *Review of Financial Studies*. forthcoming.

Gompers, Paul A (1995). "Optimal investment, monitoring, and the staging of venture capital". *The Journal of Finance* 50.5, pp. 1461–1489.

Grenadier, Steven R and Andrey Malenko (2011). "Real options signaling games with applications to corporate finance". *Review of Financial Studies* 24.12, pp. 3993–4036.

Grenadier, Steven R, Andrey Malenko, and Nadya Malenko (2016). "Timing decisions in organizations: Communication and authority in a dynamic environment". *The American Economic Review* 106.9, pp. 2552–2581.

Grenadier, Steven R and Neng Wang (2005). "Investment timing, agency, and information". *Journal of Financial Economics* 75.3, pp. 493–533.

Gryglewicz, Sebastian and Barney Hartman-Glaser (2016). *Dynamic agency and real options*. Working Paper.

Hall, Brian J and Jeffrey B Liebman (1998). "Are CEOs really paid like bureaucrats?" *The Quarterly Journal of Economics* 113.3, pp. 653–691.

He, Z. (2011). "A model of dynamic compensation and capital structure". *Journal of Financial Economics* 100.2, pp. 351–366.

He, Zhiguo, Si Li, et al. (2014). "Uncertainty, risk, and incentives: Theory and evidence". *Management Science* 60.1, pp. 206–226.

He, Zhiguo, Bin Wei, et al. (2017). "Optimal long-term contracting with learning". *The Review of Financial Studies* 30.6, pp. 2006–2065.

Hirshleifer, David and Yoon Suh (1992). "Risk, managerial effort, and project choice". *Journal of Financial Intermediation* 2.3, pp. 308–345.

Holmstrom, Bengt and Paul Milgrom (1987). "Aggregation and Linearity in the Provision of Intertemporal Incentives". *Econometrica* 55.2, pp. 303–28.

Jensen, Michael C and William H Meckling (1976). "Theory of the firm: Managerial behavior, agency costs and ownership structure". *Journal of Financial Economics* 3.4, pp. 305–360.

Jensen, Michael C and Kevin J Murphy (1990). “Performance Pay and Top-Management Incentives”. *The Journal of Political Economy* 98.2, pp. 225–264.

Kallapur, Sanjay and Mark A Trombley (1999). “The association between investment opportunity set proxies and realized growth”. *Journal of Business Finance and Accounting* 26.3-4, pp. 505–519.

Korteweg, Arthur G and Nicholas Polson (2009). *Corporate credit spreads under parameter uncertainty*. Working Paper.

Lambert, Richard A (1983). “Long-term contracts and moral hazard”. *The Bell Journal of Economics* 14, pp. 441–452.

Lyandres, Evgeny and Alexei Zhdanov (2013). “Investment opportunities and bankruptcy prediction”. *Journal of Financial Markets* 16.3, pp. 439–476.

Murphy, Kevin J (1985). “Corporate performance and managerial remuneration: An empirical analysis”. *Journal of Accounting and Economics* 7.1, pp. 11–42.

— (1999). “Executive compensation”. *Handbook of Labor Economics* 3, pp. 2485–2563.

— (2013). “Executive Compensation: Where We Are, and How We Got There”. *Handbook of the Economics of Finance*. Ed. by Milton Harris George M. Constantinides and Rene M. Stulz. Vol. 2, Part A. Elsevier, pp. 211–356.

Peters, Ryan H. and Lucian A. Taylor (2017). “Intangible capital and the investment-q relation”. *Journal of Financial Economics* 123.2, pp. 251–272.

Purnanandam, Amiyatosh and Uday Rajan (2017). “Growth Option Exercise and Capital Structure”. *Review of Finance*. forthcoming. DOI: 10.1093/rof/rfw062.

Rajgopal, Shivaram and Terry Shevlin (2002). “Empirical evidence on the relation between stock option compensation and risk taking”. *Journal of Accounting and Economics* 33.2, pp. 145–171.

Rhodes–Kropf, Matthew, David T Robinson, and S Viswanathan (2005). “Valuation waves and merger activity: The empirical evidence”. *Journal of Financial Economics* 77.3, pp. 561–603.

Rogerson, William P (1985). “The first-order approach to principal-agent problems”. *Econometrica* 53.6, pp. 1357–1367.

Sannikov, Yuliy (2008). “A continuous-time version of the principal-agent problem”. *The Review of Economic Studies* 75.3, pp. 957–984.

Yermack, David (1995). “Do corporations award CEO stock options effectively?” *Journal of Financial Economics* 39.2, pp. 237–269.

— (1996). “Higher market valuation of companies with a small board of directors”. *Journal of Financial Economics* 40.2, pp. 185–211.

CHAPTER 3

Institutions and Resource Misallocation: Firm-level Evidence

with Bruno Pellegrino

3.1 Introduction

One of the major accomplishments of economic theory in the twentieth century was to shed light on the mechanism by which market forces allocate resources efficiently, and to explain how policy distortions can hinder allocative efficiency.

Answering these questions is important for two reasons. Firstly, resource misallocation might be a significant driver of cross-country differences in productivity and income (Banerjee and Duflo, 2005; Restuccia and Rogerson, 2008); hence, by better understanding resource misallocation we might shed light on the origins of income disparities across countries. Secondly, accurately documenting the extent and sources of distortions in an economy can help design policies to remove them, which in turn can raise aggregate income.

In this paper, we utilize financial and survey data to quantify the fraction of between-firm heterogeneity that can be attributed to several types of distortions. In doing so, we take seriously the difference between ex-ante, perceived distortions, which result in allocative inefficiencies, and ex-post, observed heterogeneity. We find that a significant proportion of the observed heterogeneity cannot be attributed to commonly identified sources of misallocation.

One of the earliest efforts to identify and quantify allocative efficiency is due to Harberger (1954), which in one landmark paper attempted to quantify the effect of market power on capital allocation across US industries. More recently, empirical economists have begun using rich micro-level datasets to estimate aggregate measures of misallocation in the context of heterogeneous firm models. One major contribution is due to Hsieh and Klenow (2009), henceforth HK, who estimate how much manufacturing GDP could be recovered by shifting capital and labor across plants: they estimated that aggregate output from reallocating inputs might be as high as 40% for the United States, and as high as 120% for India and China. Their work serves as an important benchmark for quantifying the aggregate implications of resource misallocation.

The state of the art of current misallocation theory relies crucially on three tenets. One, is that firms are fundamentally heterogenous in their productivity. Two, is that these frictions are best modeled as "wedges" (or equivalently, "shadow taxes" or "iceberg costs", following the trade theory jargon), which appear in the objective function of the firm, but not in the actual firms' profit and loss accounts. Three, is that the magnitude of these wedges varies across individual firms.

A number of empirical studies of misallocation have relied, so far, on what we call an "indirect" approach; the commonality of these studies (Bartelsman, Haltiwanger, and Scarpetta, 2013; Gopinath et al., 2017; Hsieh and Klenow, 2009; Midrigan and Xu, 2014; Restuccia and Rogerson, 2008) is that they start from the basic notion that the key ingredients in the misallocation measurement cookbook - that is, the firm-level wedges - are fundamentally unobserved, and therefore the researcher needs to back them out them from the data by imposing very strong modeling assumptions. In practice, the assumptions consists of attributing to the treatment variable of interest (in this case, the unobserved frictions) some residual variation in data that cannot otherwise be accounted for by the model. The by-product of such residual measurement approach is that the researcher cannot make any conclusive statement about the origin of the inferred firm-level frictions (Haltiwanger, Kulick, and Syverson, 2018).

Some recent studies have tried to move from the indirect approach to a *direct* approach. In order to achieve this, it is necessary for the researcher to have access, together with accounting data, to some firm-level "treatment variable" that can be unambiguously linked to the individual firms' exposure to frictions. One notable study that has adopted this approach is due to Joel M. David, H. A. Hopenhayn, and Venkateswaran (2016), henceforth DHV, who estimated the malallocative impact of informational frictions using the information content of stock market data.

Our study moves the literature forward in this direction. The objective of this research is to measure, using the *direct* approach, the aggregate malallocative impact of four types of institutional frictions that are prominent in the literature[1], namely : (A) bureaucracy and regulations; (B) family control; (C) financial frictions; (D) labor market laws. We feel that these institutional frictions are particularly close in spirit to those subsumed by HK's original paper, which we consider to be the blueprint for misallocation measurement using firm-level data. The theoretical model featured in this paper, as well as the one in DHV, are both based on the one by HK.

What allows us to directly measure the impact of these these institutional distortions is a novel dataset provided to us by the Brussels-based think tank Bruegel: the EFIGE dataset (Altomonte and Aquilante, 2012). The dataset augments accounting data from the Amadeus-Bureau van Dijk databank with survey data from a large representative sample of manufacturing firms from seven European countries (Austria, France, Germany, Hungary, Italy, Spain and the UK). In the survey response data, firms themselves disclose information about the institutional frictions they face.

The comparative richness of our dataset, compared with a novel identification strategy, allows us to identify the impact of these four distortions at the firm level under relatively lax assumptions. In particular, we no longer require the production technology to be homogeneous across firms or to take a known functional form. Also, our model allows for informational frictions in the style of DHV, although their effect is not estimated directly. Hence, we see our study as complementary to theirs.

We estimate a statistically significant impact of financial constraints in Spain, of family

[1]H. Hopenhayn and Rogerson (1993), Eisfeldt and Rampini (2006), Farrell and Lund (2006), Buera and Shin (2013), Caselli and Gennaioli (2013), Gourio and Roys (2014), Midrigan and Xu (2014), Garicano, Lelarge, and Van Reenen (2016), Dias, Robalo Marques, and Richmond (2016), León-Ledesma and Christopoulos (2016), Whited and Zhao (2016), Gopinath et al. (2017), and Adamopoulos et al. (2017)

control in Spain and Hungary, and of labor market regulations in Italy. As was the case for Harberger's landmark study, these results are non-trivial in absolute terms (the total value of forgone output is in the order of several tens of billions of Euros); however, they are relatively small compared to measurements that have been produced in the recent literature. We estimate the gains from input reallocation to be less than 1 percentage point of aggregate output. These findings are reflective of the fact that firm's markups do not appear to vary substantially between firms that are constrained and those that are unconstrained - which is the ultimate testable implication of misallocation in general equilibrium models.

Because the firms' relevant cost base, input substitutability, and demand elasticity varies substantially depending on whether we consider the firms' behavior in the long run or the short run, we also estimate an alternative, "short-run" variant of our model. Estimating our model under this alternate mapping, we find a statistically significant effect of family control in Germany and of labor market regulations in France and Italy. However, reallocation gains still fall below 1 percentage point of GDP for every country in our sample.

We believe these findings provide an important contribution to the current debate on the impact of institutional frictions on aggregate productivity, and, more in general, on the nature of cross-country differences in income. Moreover, the data and techniques presented in this study might contribute to more informed policies and regulations that target misallocation, especially with regards to EU countries.

The rest of the paper is organized as follows. In section 3.2, we present a model of monopolistic competition with heterogeneous firms and frictions; the purpose of the model is to provide a description of how distortions affect the firms' maximization problem; the model produces a set of statistical relationships among variables that we can use to recover the distribution of the firm-level distribution of wedges from the data. In section 3.3, we describe our dataset and map its variables into their counterparts from our theoretical model. In section 3.4, we illustrate how we can use the model to recover the the wedges' distribution from the data and estimate their effect on aggregate output. In section 3.6, we present and discuss estimation results, and use the structural parameters recovered from the data to compute how much higher aggregate output would be in a counterfactual friction-less equilibrium. In section 3.7, we discuss the robustness of our econometric results. In section 3.9, we conclude.

3.2 A (distorted) model economy

In this section we present a model of monopolistic competition that describes how heterogenous policy frictions impact the behavior of firms. We derive equilibrium relationships that we can use to recover the distribution of firm-level distortions from the data. The empirical distribution of the wedges will then inform us of which policy frictions are most relevant in which country, and will allow us to compute their impact on aggregate output.

A set of "upstream" firms $i \in \mathcal{I}$ produces differentiated goods using a production function of capital K, labor L, and intermediate inputs X:

$$Y_i = A_i \cdot F_i(K_i, L_i, M_i) \tag{3.1}$$

A_i is firm i's total factor productivity (TFP): we assume it to be exogenously determined. The production function is assumed to satisfy constant returns to scale (CRS), which implies:

$$\frac{\partial \log F_i}{\partial \log K_i} + \frac{\partial \log F_i}{\partial \log L_i} + \frac{\partial \log F_i}{\partial \log M_i} = 1$$

Notice that the production function is allowed to vary at the firm level. We assume that the total supply of capital, labor and intermediate inputs is completely inelastic at the country level. This assumption is consistent with our objective of studying how the *allocation* (as opposed to the total supply) of inputs affects aggregate output.

There is a single "downstream" firm, which takes the output of the upstream firms as input and produces a consumption good using a constant elasticity of substitution (CES) production function

$$Y = \left(\sum_{i \in \mathcal{I}} e^{\frac{z_i}{\eta}} Y_i^{\frac{\eta-1}{\eta}} \right)^{\frac{\eta}{\eta-1}}$$

where z_i are a series of firm-specific shocks that are independently, identically normally distributed and have mean μ^z and variance Σ^z. Throughout the paper, we will use Σ^j to denote the variance of random variable j, and σ^j to denote its standard deviation.

Intermediate good firms choose their input mix prior to realizing the demand shock. We assume that the final goods firm is distortion-free and seeks to maximize profits

$$\max_{\{Y_i\}_{i\in\mathcal{I}}} \left\{ Y - \sum_{i\in\mathcal{I}} P_i Y_i \right\}$$

taking the prices of the upstream goods P_i as given. Notice we are normalizing the price vector by picking the final good as numeraire. This implies that the demand function faced by each intermediate goods firm is given by:

$$Y_i = \left(\frac{e^{z_i}}{P_i} \right)^{\eta}$$

Intermediate good firms i face homogeneous prices for rented capital (r), labor (w) and intermediate inputs (p) and heterogeneous, random wedges on output $\left(\tau_i^Y\right)$, capital $\left(\tau_i^K\right)$, and labor $\left(\tau_i^L\right)$. Firm i maximizes the following "distorted" expected profit function:

$$\max_{K_i, L_i, X_i} \mathbb{E}_i \left[\exp\left(-\tau_i^Y\right) P_i Y_i - \exp\left(\tau_i^K\right) r K_i - \exp\left(\tau_i^L\right) w L_i - p M_i \right]$$

where $\mathbb{E}_i$ represents taking the conditional expectation with respect to firm i's information set. Moreover, we also assume that the vector $\left(\tau_i^Y\ \tau_i^K\ \tau_i^L\right)$ is independently and identically distributed according to a multivariate normal:

$$\left(\tau_i^Y\ \tau_i^K\ \tau_i^L\right) \sim iid\, N\left(0, \Sigma^{\tau}\right)$$

The wedges reflect the distortionary impact of policies, institutions and financial frictions on the firms' optimization problem. We make the key assumption that the wedge vector $\left(\tau_i^Y\ \tau_i^K\ \tau_i^L\right)$ is statistically independent from the demand shock z. Moreover, in order to simplify the notation going forward, we are going to define implicitly the expected wedges

$\left(\bar{\tau}_i^Y \; \bar{\tau}_i^K \; \bar{\tau}_i^L\right)$ as follows:

$$\begin{aligned}
\exp\left(-\bar{\tau}_i^Y\right) &= \mathbb{E}_i\left[\exp\left(-\tau_i^Y\right)\right] \\
\exp\left(\bar{\tau}_i^K\right) &= \mathbb{E}_i\left[\exp\left(\tau_i^K\right)\right] \\
\exp\left(\bar{\tau}_i^L\right) &= \mathbb{E}_i\left[\exp\left(\tau_i^L\right)\right]
\end{aligned} \tag{3.2}$$

The independence assumption $\tau \perp z$ allows us to separate the expectations and re-write the maximization problem as

$$\max_{K_i, L_i, X_i} \; \exp\left(-\bar{\tau}_i^Y\right) \mathbb{E}_i\left(P_i Y_i\right) - \exp\left(\bar{\tau}_i^K\right) r K_i - \exp\left(\bar{\tau}_i^L\right) w L_i - p M_i$$

Note from the equations above that our model draws an important distinction that the previous literature ignores. We explicitly distinguish the *actual* wedges and what the firms *know* about the wedges (their expectations). The key concept that we are incorporating is that the exact amount of the wedge cannot be objectively measured by anyone, including the firms themselves: the impact of these wedges is not accounted for on the firms' actual balance sheets. Therefore, a wedge can only impact a firm's profit maximization to the extent to which the firm can anticipate the impact of that wedge at the time they decide their input mix. This distinction between *expected* and *actual* wedges is central to our identification strategy, as we shall discuss in Section 3.4.2.

The first order conditions of firm i's profit maximization problem can be written as:

$$\begin{aligned}
\text{MRP}_i^K &\overset{\text{def}}{=} \frac{\eta - 1}{\eta} \cdot \frac{\partial \log F_i}{\partial \log K_i} \cdot \frac{\mathbb{E}_i\left(P_i Y_i\right)}{K_i} &= \exp\left(\bar{\tau}_i^Y + \bar{\tau}_i^K\right) r \\
\text{MRP}_i^L &\overset{\text{def}}{=} \frac{\eta - 1}{\eta} \cdot \frac{\partial \log F_i}{\partial \log L_i} \cdot \frac{\mathbb{E}_i\left(P_i Y_i\right)}{L_i} &= \exp\left(\bar{\tau}_i^Y + \bar{\tau}_i^L\right) w \\
\text{MRP}_i^X &\overset{\text{def}}{=} \frac{\eta - 1}{\eta} \cdot \frac{\partial \log F_i}{\partial \log X_i} \cdot \frac{\mathbb{E}_i\left(P_i Y_i\right)}{M_i} &= \exp\left(\bar{\tau}_i^Y\right) p
\end{aligned} \tag{3.3}$$

the equations above represent the core testable prediction of this class of models: positive (negative) wedges manifest themselves in the data in the form of higher (lower) marginal revenue products. Hence, a positive (negative) wedge reduces (increases) the affected firm's equilibrium size. Additionally, if the wedge is applied specifically to one input, it will reduce (increase) the firm's relative usage of the affected input.

Using the constant returns to scale assumption, we can merge the firm's first order conditions to obtain the following equilibrium law for individual firm i:

$$\frac{\mathbb{E}_i\left(P_iY_i\right)}{\exp\left(\bar{\tau}_i^Y\right)} = \frac{\eta}{\eta-1}\left[\exp\left(\bar{\tau}_i^K\right)rK_i + \exp\left(\bar{\tau}_i^L\right)wL_i + pX_i\right] \tag{3.4}$$

this formula, as we shall see, is going to be central to the empirical part of this study.

The unit cost of output, *net* of distortions (c_i) and *gross* of distortions (C_i), are implicitly defined using the following formulas:

$$c_iY_i = rK + wL_i + pX_i \tag{3.5}$$

$$C_iY_i = \exp\left(\bar{\tau}_i^Y\right)\left[\exp\left(\bar{\tau}_i^K\right)rK_i + \exp\left(\bar{\tau}_i^L\right)wL_i + pX_i\right]$$

The inelastic input supply assumption implies that it is the variance (Σ^τ), not the mean, of the wedges that determines the extent of misallocation, and so we assume without loss of generality that the wedges are mean zero. [2]

Having laid down the theoretical framework, we can now proceed to present our data and map it to the variables in this model. We will then be able to recover the structural parameters of our model - most importantly, the variance-covariance matrix of the wedges. The rest of the equilibrium relationships used in the empirical part of this paper are detailed in appendix 3.13.

[2] To see why this is the case, consider an equilibrium with aggregate input prices ($r = r'$; $w = w'$; $p = p'$). Consider now a mean shift in the distribution of τ from the origin to some other level $\left(\hat{\mu}^K\,\hat{\mu}^L\,\hat{\mu}^Y\right)$, then it is easy to see that, for the alternative input price vector

$$\begin{aligned} r &= r' \cdot \exp\left(-\hat{\mu}^Y - \hat{\mu}^K\right) \\ w &= w' \cdot \exp\left(-\hat{\mu}^Y - \hat{\mu}^L\right) \\ p &= p' \cdot \exp\left(-\hat{\mu}^Y\right) \end{aligned}$$

all input demands are unchanged, implying we have found an equivalent equilibrium. This shows that equilibrium output is invariant to any mean shift of the distribution of τ. This is a consequence of the assumption that input supply is inelastic.

3.3 Data and mapping to model variables

3.3.1 Dataset description

To bring our model to data, we use the EU-EFIGE/Bruegel-UniCredit dataset, which is a firm-level database. The dataset contains data for a stratified sample of 14,759 manufacturing firms from seven European countries (Austria, France, Germany, Hungary, Italy, Spain, UK). The sample is stratified over three size classes (10-49, 50-249, ≥250 employees) and twenty-four manufacturing industries.[3]

The dataset is comprised of two parts. The first part is cross-sectional response data from the EFIGE executives survey, which was conducted by the think tank Bruegel in 2010: firms were asked questions about a wide range of topics, including their organizational structure, ownership, workforce, international activities, and financing. The second part is a firm/year panel of firm financials (including turnover, assets, interest expenditure, profit and labor costs) for the period 2001-2014 merged from the Amadeus dataset, which is managed by the Bureau van Dijk.

For our analysis, we use a cross-section of the Amadeus part of the dataset combined with the EFIGE survey data. Because we want to use data from a point in time which is as close as possible to the time the survey was administered, and at the same time in which the EU economy is near its the long-run equilibrium, we use, for each firm, the latest datapoint available in the 2004-2007 period.

We use three sets of dummy variables that represent the firms' answers to selected questions from the EFIGE survey. The first set of dummies is generated by the firms answering the following multiple choice question:

"E6. Indicate the main factors preventing the growth of your firm:
☐ *financial constraints*
☐ *labour market regulations*
☐ *legislative or bureaucratic restrictions*
☐ *lack of management and/or organizational resources*

[3] Sector definition is NACE rev 2, two-digit level (codes 10 to 33)

☐ *lack of demand*
☐ *other"*

The second set of dummy variables is generated by the firms answering the following multiple choice question:

> *"F16. Which type of information does the bank normally use/ask to assess your firm's credit worthiness?*
> ☐ *Collateral*
> ☐ *Balance sheet information*
> ☐ *Interviews with management on firm's policy and prospects*
> ☐ *Business plan and firms' targets*
> ☐ *Historical records of payments and debt service*
> ☐ *Brand recognition*
> ☐ *Other"*

In both cases, if the firm ticks a certain option, a corresponding dummy variable in the dataset takes value equal to one. The third dummy variable we use from the EFIGE dataset takes value one if the firm answers "yes" to the following yes/no question:

> *"A20. Is your firm directly or indirectly controlled by an individual or family-owned entity?"*

Finally, we also use some data that is external to the EFIGE dataset. In order to compute firm-level *TFPR* using the Hsieh-Klenow approach (which we use as a benchmark) we use sector-level estimates of the production function elasticities in 2007, which we source from the EU KLEMS database (O'Mahony and Timmer, 2009).

3.3.2 Representativeness and coverage

To ensure representativeness of the EFIGE sample, the dataset is equipped with sampling weights. Weighting ensures that the in-sample distribution of firms over industries and size classes matches the population's. By explicitly accounting for firm size, we ensure that our estimates are not the result of some underlying relationship between markup and size. This effect is studied in Edmond, Midrigan, and Xu (2018), and we account for size to ensure that our estimates are not driven by the super-elasticity effects studied in their work.

However, representativeness is not guaranteed when the availability of firms' financials is taken into consideration. The Amadeus database, which is the source of firm financials, has known issues of coverage and sample selection (Kalemli-Ozcan et al., 2015). Specifically, while the (weighted) EFIGE sample is representative, firm financials appear to be missing, for certain countries (Germany, UK) in a non-random way.

We are able to address this issue thanks to the fact that stratification variables (employment size and NACE 2-digit industry) belong to the survey part of the dataset, and are therefore available for all the firms in the sample, regardless of whether BvD financial data for the corresponding firm is available. This allows us to devise a re-weighting scheme for German and British firms, for which we do find evidence of sample selection based on firm size. The application of these weights for all our analyses that require firm financials allows us to preserve the representativeness of our sample with respect to the stratification variables. We discard Austrian firms because financial data is available for too few of them (less than 70). For further details regarding the methodology we used to correct for sample selection, please see Appendix 3.10.

We call the full cross section of firms in the EFIGE dataset, excluding Austria, the "full sample" (14,316 observations); we call instead the subsample of firms for which Amadeus financial data and the relevant survey questions are available at least for one year (6,560 observations) the "BvD Sample". Tables 3.1 and 3.2 show the distribution of firms by country and size category for both the full sample and the BvD sample. Table 3.3 shows the distribution of firms in the BvD sample by country and industry.

3.3.3 Mapping dataset variables to model variables

Having described our theoretical framework and our dataset, we now proceed to map one into another. To begin with, we assume that the output wedge τ^Y can be itself decomposed as the sum of two iid normal wedges: τ^F which captures the effect of family control and management, and τ^G, which reflects asymmetric distortions stemming from bureaucracy and governmental regulation:

$$\tau_i^Y = \tau_i^F + \tau_i^G$$

τ^G is close to HK's original interpretation of output distortions. The inclusion of family control τ^F is instead motivated by the work of Caselli and Gennaioli (2013) and, more generally, a large finance literature on so-called "empire-building" desires by firm managers (Bertrand and Mullainathan, 2003). Our rationale for considering this as a source of distortions is that a family's preference for maintaining control over the firm's management limits the ability of a firm to grow, as doing so may require giving up valuable control rights.[4] Vice-versa, when the firm's is not tightly controlled by a family, it is easier for managers to engage in empire-building, and this can cause the firm to grow beyond its optimal size. The sign and size of the wedge τ^F reflects this ongoing tension between the management's unobserved preference for empire building and the controlling family's desire to maintain control over the firm's operations.

Following this decomposition, the distribution of the firm-level wedges can be therefore be re-defined as follows:

$$\begin{bmatrix} \tau_i^F \\ \tau_i^G \\ \tau_i^K \\ \tau_i^L \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu^F \\ \mu^G \\ \mu^K \\ \mu^L \end{bmatrix} \begin{bmatrix} \Sigma^F & & & \\ \rho^{FG}\sigma^F\sigma^G & \Sigma^G & & \\ \rho^{FK}\sigma^F\sigma^K & \rho^{GK}\sigma^G\sigma^K & \Sigma^K & \\ \rho^{FL}\sigma^F\sigma^L & \rho^{GL}\sigma^G\sigma^L & \rho^{KL}\sigma^K\sigma^L & \Sigma^L \end{bmatrix} \right)$$

In order to incorporate the firm's (and the econometrician's) knowledge of these distortions, let us define the following set of indicator variables: they evaluate to one if the firm's corresponding wedge τ_i overcomes a certain normalized threshold T:

$$D_i^F = \mathbb{I}\left\{\frac{\tau_i^F}{\sigma^F} > T^F\right\} \qquad D_i^G = \mathbb{I}\left\{\frac{\tau_i^G}{\sigma^G} > T^G\right\} \tag{3.6}$$
$$D_i^K = \mathbb{I}\left\{\frac{\tau_i^K}{\sigma^K} > T^K\right\} \qquad D_i^L = \mathbb{I}\left\{\frac{\tau_i^L}{\sigma^L} > T^L\right\}$$

[4] See Pérez-González (2006), Bennedsen et al. (2007), Tsoutsoura (2015), and Villalonga and Amit (2006)

The survey response dummies discussed in Section 3.3 are mapped directly to these four dummies. Specifically: D^F is equal to one if the firm reports being family-controlled; D^G is equal to one if the firm reports its growth to be constrained by bureaucracy/government regulations; D^K is equal to one if the firm reports its growth to be constrained by financial constraints (with exceptions: see below) D^L is equal to one if the firm reports its growth to be constrained by labor market regulations. In other words, these dummy variables embed information, provided by the firms themselves, about which firms face which constraints. The frequency of "one" values for each these dummies is plotted, by country and constraint type, in Figure 3.1.

We are also using another variable from the dataset, which we label D^z: it indicates whether the firm reports "Lack of demand" as a reason for its lack of growth. This variable doesn't reflect the presence of a real wedge, but it is still useful: it will be used to carry out a rough endogeneity test in the empirical section of the paper.

From the Amadeus part of dataset we use operating revenues, fixed assets and cost of labor. These are mapped, respectively, to model variables PY, K, and L. We also use EBITDA, which we use to compute all costs other than capital and labor. Consistently with the KLEMS framework (O'Mahony and Timmer, 2009), this residual costs measure is mapped to the intermediate input X.

This mapping implies that the price of labor and of intermediate inputs is normalized to 1, as in Hsieh and Klenow (2009). This is consistent with our assumption that input markets are competitive, so each unit of labor or intermediate input is paid its marginal revenue product.

We measure the (marginal) rental price of capital r as the sum of the policy rate, a spread that is given by the sum of: 1) the central bank policy rate; 2) the yearly average value of the Bank of America Merrill Lynch Corporate Bonds Master Spread Index; 3) The average depreciation rate in the sample. Because we need a measure of the marginal cost of capital, the implicit assumption is that each additional dollar of productive capital (measured as fixed assets) is financed using non-current debt.

Table 3.4 summarizes the mapping of the variables in the model to the variables in our dataset, and Table 3.5 displays descriptive statistics.

3.4 Methodology

3.4.1 Estimating markups

We estimate markups at the firm level using the supply-side approach outlined by (De Loecker and Warzynski, 2012, henceforth DW), who show how to build a robust metric of markup that only relies on the assumption of firms engaging in cost-minimization.

The first step is to estimate the production function, which we can do thanks to the fact that our firm financials data comes in a panel format, covering the years 2001-2014. Letting lowercase letters denote logged output and input, and adding a time subscript t

$$y_{it} = \omega_{it} + \theta_s^K\, k_{it} + \theta_s^L\, \ell_{it} + \theta_s^M\, m_{it} + \varepsilon_{it}$$

where ω_{it} is firm i's (stochastic) productivity at time t, which we assume to follow a Markov process, θ_s^X is the elasticity of output with respect to input X, which varies at the level of sector $s \ni i$.

We estimate this production function using a control function approach (Olley and Pakes, 1996; Levinsohn and Petrin, 2003). In the underlying structural model, capital is the state variable, labor is the free variable, and material inputs is our "proxy" variable. Then, provided that the material inputs choice is not subject to any dynamic friction, DW show that we can compute the supply-side markup using the following formula.

$$\mu_{it} = \theta_s^M \cdot \frac{P_{it}Y_{it}}{c_{st}M_{it}} \cdot e^{-\varepsilon_{it}}$$

the

3.4.2 Parametric identification

We now illustrate how model parameters can be recovered from the data: specifically, we need to recover the elasticity of substitution η, the parameters of the joint distribution of the firm-level wedges (μ^τ, Σ^τ) and the thresholds T.

We already made the case, in section 3.2, that the mean of the distribution of the vector τ has no effect on resource allocation. As we assumed that the wedges were mean zero, this allowed us to identify the wedges separately from the thresholds. Another possible normalization would have been to assume the thresholds to be equal to zero. In Appendix 3.11 we explain why our choice of normalization is more sensible.

Following this assumption, we have that the percentage of "one" values for each dummy D^j immediately pins down the ratio of the reporting threshold to the variance for the respective constraint $\left(\frac{T^j}{\sigma^j}\right)$:

$$\mathbb{E}\left(D_i^j\right) = \mathbb{P}\left(\frac{\tau_i^j}{\sigma^j} > T^j\right) = 1 - \Phi\left(T^j\right)$$

where $\Phi(\cdot)$ is cumulative normal distribution function. The correlation ρ^{jk} can then be easily recovered by computing the following statistic:

$$\begin{aligned}\mathbb{E}\left(D^j D^k\right) &= \mathbb{P}\left(\frac{\tau^j}{\sigma^j} \geq T^j \text{ and } \frac{\tau^k}{\sigma^k} \geq T^k\right) \\ &= 1 - \left[\Phi\left(T^j\right) + \Phi\left(T^k\right) - \Phi_2\left(T^j, T^k, \rho^{jk}\right)\right]\end{aligned} \tag{3.7}$$

where $\Phi_2(\cdot, \cdot; \rho)$ is the bivariate cumulative standard normal distribution function with correlation ρ. Notice that so far no accounting data was needed in order to identify the thresholds T or the correlations ρ.

At this point, what's still left to identify are the four diagonal elements of the variance-covariance matrix Σ^τ and the demand elasticity of substitution η . We show that these parameters are identified, in the Generalized Method of Moments (GMM) framework, using equation 3.4 as a moment condition.

We first provide some intuition on the mechanics of our identification strategy. The key testable assumption of misallocation is that firms with higher expected wedges ($\bar{\tau}$) should display higher markups, which we denote μas in Asker, Collard-Wexler, and De Loecker (2014). In our dataset, firms for which $D = 1$ have a greater expected wedge. As a consequence, if there is substantial heterogeneity in wedges, these firms will also tend to display a higher markup (revenues/cost ratio). The larger the dispersion of the wedges, the greater the difference in the expected wedges - and therefore in markups - between the "treated" ($D = 1$) and the "control" ($D = 0$) firms. Hence, we can infer the dispersion in wedges by comparing the markups of firms that report a large positive wedge ($D = 1$) to those of the firms that do not.

In order to recover the dispersion of the wedges, by using this insight, we turn to equation 3.4. The equation describes a condition on the firm's markup that all firms have to respect in equilibrium.

The key assumption that allows us to attain identification has to do with information. In particular, we assume that the firm's information set, at the moment of choosing the input mix, only includes the dummy variables $\left(D_i^F \; D_i^G \; D_i^K \; D_i^L\right)$. Intuitively, this means that all the information about the distortions that is used by the firm i in solving its profit-maximization problem is revealed to us in the dummy variables collected via the questionnaire. The firm is neither lying to us, nor it is knowingly withholding information. The implications of this assumption are discussed at length in section 3.5.

This assumption is restrictive, but standard: it falls in line with the general principle of structural estimation, that the optimizing agent should not have a larger information set than the econometrician. The consequence of this assumption is that equation (3.4) can be re-written as:

$$\mathbb{E}\left[\exp\left(z_i\right)\middle|\, D_i\right] = \text{constant}$$

$$\mathbb{E}\left[\frac{\eta-1}{\eta} \cdot \frac{\exp\left(-\bar{\tau}_i^F - \bar{\tau}_i^G\right) P_i Y_i}{\exp\left(\bar{\tau}_i^K\right) rK_i - \exp\left(\bar{\tau}_i^L\right) wL_i - pX_i}\middle|\, D_i\right] = 1 \tag{3.8}$$

The expected wedges $\left(\bar{\tau}_i^F \; \bar{\tau}_i^G \; \bar{\tau}_i^K \; \bar{\tau}_i^L\right)$ are a function of the dummy variables D, the thresholds T and of the diagonal elements of the variance-covariance matrix Σ^τ.

The GMM instruments are the dummy variables D (four) plus a constant term. Hence, the model is at least exactly identified. Because the equation must hold for every configuration of the treatment dummies D_i, their interactions are also, in principle, viable instruments, although they might be redundant. The detailed equations that express the expected wedges s $\bar{\tau}_i$ as a function of the D_i , T and the diagonal elements of Σ^τ are provided in Appendix 3.12.

3.4.3 Estimation methodology

The next step is to translate the identification results from 3.4 into an actual estimation framework. This requires us to first solve a practical numerical problem that has to do with the dimensionality of the wedge vector.

If the wedges $\left(\tau^F \, \tau^G \, \tau^K \, \tau^L\right)$ are all correlated, then we need to repeatedly evaluate a 4-dimensional cumulative normal distribution in order to compute the GMM objective function. Unfortunately, the cumulative multi-normal becomes numerically intractable when the dimensionality is above two. For the estimation we use the statistical package STATA, which indeed only provides an in-built binormal *cdf* evaluator. This means that, in practice, we can only allow two wedges at a time to be correlated.

We argue that a dimensionality of two is sufficient for this setting: to show this, we have computed the correlation matrix of the dummy vector D and reported it in Table 3.6. As can be seen from the Table, there is virtually no correlation between the dummies, except for the $\left(D^G, D^L\right)$ pair, which has a correlation of nearly 50%. As a consequence, we believe it is reasonable to carry out the estimation by modeling explicitly the correlation ρ^{GL} and assuming that all of the other entries of the correlation matrix of τ are zero:

$$\rho^{FG} = \rho^{FK} = \rho^{FL} = \rho^{GK} = \rho^{KL} = 0$$

As we mentioned before, the parameters T^F, T^G, T^K, T^L and ρ^{GL} can be retrieved without using any firm financials. This means that this part of the estimation can be carried out using the "full sample". The second part of the estimation instead, requires financial data. In order to exploit all the available data, we carry out the estimation sequentially[5]:

In Step 1, we estimate the parameter vector

$$\left(T^F, T^G, T^K, T^L, \rho^{GL}\right)$$

using maximum likelihood (MLE) over the full sample and applying the regular EFIGE sampling weights. In Step 2, we estimate parameter vector

$$\left(\sigma^F, \sigma^G, \sigma^K, \sigma^L, \eta\right)$$

using GMM over the BvD sample and applying our weighting scheme that adjusts for sample selection in the Amadeus part of the dataset. We write the moment conditions as a function of the sigmas alone by plugging in the MLE estimates from Step 1. We use, as our GMM instruments, the vector $\left(D^F, D^G, D^K, D^L\right)$, as well as the interaction of D^G and D^L.

3.5 Identification

3.5.1 Information and measurement error

We now wish to discuss more in detail our main identifying assumption and what it implies in terms of our ability to identify the distribution of the firm-level wedges.

Firstly, recall that, in the theoretical model outlined in Section 3.2, we made an explicit distinction between the firm's actual wedge vector τ_i and the expected wedge vector $\overline{\tau}_i$. This captures the fact that, in reality, firms do not know (ex-ante) what obstacles they are going to face. Consider for example a firm trying to obtain a permit to build a new factory: it would

[5]This implies that the GMM standard errors are "conditional" on the MLE parameter estimates. Since the latter are estimated very precisely, we argue that using sequential estimation does not lead to significant downward bias in our GMM standard errors in Step 2.

be unrealistic to assume that the firm knows with exact precision what the non-monetary cost will be (bureaucratic hurdles, management time, etc.). We argue that, incorporating this uncertainty about wedges in our model is not simply a mathematical sophistication, but it has important implications in terms of: 1) what data should we collect in order to identify the malallocative impact of firm-level wedges; 2) how we interpret our econometric results.

Consider the firm's optimality conditions, shown in equation (3.3). What matters in determining the firm's scale and input mix, is not the *actual* wedges τ_i experienced ex-post by the firm, but the firm's *beliefs* about the wedges, which are captured by $\overline{\tau}_i$. This yields to the conclusion that, even if the researcher succeeds in accurately measuring the quantitative ex-post impact of the wedges, this might turn out to be counterproductive in terms of identification, since the what is needed is a measure of the firm's *perceptions* of the wedges, not of the *actual* wedges.

This is the reason why we believe our firm-level data is particularly appropriate to the econometric task at hand, more so than most other quantitative measures of the ex-post wedges. We believe that, unless the quantitative impact of the friction being investigated can be accurately and objectively measured by the firms themselves, survey data should always be preferred.

Next, we consider different types of measurement error that the dummy vector D might incorporate, and how robust our results are to such mismeasurement. As mentioned in (3.4.2), our key identifying assumption is that all information about the wedges that is used by the firms in solving their maximization problem is reported in the dummy vector D_i.

Because we need D to reflect firms' beliefs, any bias in the dummy vector D will not create an identification problem as long as it reflects an incorrect belief of the firm about the wedges. Indeed, if firms hold biased beliefs about the wedges, the "biased" expected wedges will determine the size and direction of misallocation. Bias in D will only create an identification problem if D fails to reflect information that firms are using in their profit-maximization problem.

We argue that the amount of information used by the firms that is omitted in the dummy variables D is likely to be very small. This is because firms are unlikely to know the exact magnitude of the shadow taxes they faces. Consider, for example, administering a survey in which firms are asked to report the tax-equivalent value of the bureaucratic constraints they face. How precise would we expect their estimate to be?

In our view, a more realistic model of how firms incorporate institutional constraints in their decision making is that the firms have a general idea of *which* institutional constraints they face. This type of information is likely captured with a reasonable degree of accuracy by the EFIGE survey. If this is not the case - that is, if firms cannot even report accurately this limited amount of information - then we argue that it is impossible for misallocation to arise in the first place: if firms don't have any information that they can relay to us about the expected sign and size of the wedges, then there is no incremental information about the wedges that they can incorporate in their decision making that will make them deviate from the efficient allocation. A failure of this identifying assumption about the firm's information structure will tend, in general, to attenuate our estimates of the wedge's variance.

While, unfortunately, there is no way to test whether this assumption is correct, we argue our identifying assumptions are very lax compared to those that have to be invoked under the "indirect" measurement framework, namely: 1) the firm knows the exact amount of the wedges and faces no uncertainty; 2) the researcher doesn't know the wedges, but knows everything else about the model economy. Furthermore, the frictions indicated by our survey responses are consistent with prior work on cross-country differences in economic frictions. In 3.3, we plot the results of a probit regression of the country-level likelihood of reporting labor constraints D^L against the Regulation of Labor Score of Botero et al. (2004). Their score, which is based on cross-country differences in legal frameworks, is computed entirely separately from our measure of labor constraints, which is drawn from survey responses. Nevertheless, we find that there is a strong rank-correlation between these measures, and that the effect is linear in the index. Similarly, 3.2 plots the country-level likelihood of reporting bureaucratic restrictions D^G against the Regulation of Entry Score of Djankov et al. (2002). This score is based on the time and cost a hypothetical startup would face in the country, and again exhibits strong rank-correlation with our country-level survey responses regarding legislative / bureaucratic restrictions. These plots serve to validate our survey data and suggest that the survey responses are informative of actual frictions faced by firms in those countries.

3.5.2 Financial constraints: rationing or screening?

Among the dummy variables D, which constitute our observed variation in firm-level frictions, D^K stands out as the one most likely to cause the identifying assumption $z \perp D^\tau$ to fail. This concern is motivated by an extensive literature in financial economics.

One of the key roles of the financial sector, as identified by Levine (2005), is "the production of information ex ante about possible investments and the allocation of capital". Thus, screening is a sign of a functioning financial sector. Credit rationing, in contrast, is a result of market imperfections and a sign of market failure, in that firms have limited credit despite being willing to pay a higher interest rate (Stiglitz and Weiss, 1981; Holmström and Tirole, 1998).

In other words, it is conceivable that some of the firms reporting financial constraints are "false positives", in the sense that they are actually being screened out of credit, although they report being rationed. These false positives pose a threat to identification because their screening is likely to incorporate information about idiosyncratic variation in demand conditions - causing D^K to correlate with z.

This potential endogeneity is a well-known flaw of all self-reported measures of credit constraints. While we are unable to eliminate these concerns completely, we can use information from question f16 of the EFIGE survey (previously described in section 3.3) to mitigate them. In particular, among firms that report "financial constraints" as an obstacle to growth, we revert D^K to zero for firms that reported being asked 3 or more different types of information (out of 7) by their banks (in our dataset, these firms account for approximately 50% of the firms that answered to this survey question).

The idea behind this adjustment, is that if a firm was required to disclose a significant amount of information before it was denied credit, it was probably not rationed, but screened.

Conversely if it was denied credit without being asked any information, it was instead likely the result of rationing.

A reassuring sign (although not a proof) that this adjustment is sufficient to eliminate endogeneity from D^K, is that the variable does not correlate positively with D^z following the adjustment (see Table 3.6). If anything, the two to dummies are slightly negatively correlated.

3.5.3 Generic violations of the GMM exclusion restrictions

Even after adjusting the variable D^K, it is still possible for the dummies D to correlate with the residual z for unknown reasons, and thus cause a failure of our GMM exclusion restriction.

While it is not possible, in general, to test for instrument exclusion restrictions, our data allows in this case to carry out a rough "endogeneity" test: the variable that allows us to do so is the dummy D^z, which we introduced in section 3.3. This additional dummy is generated by the same survey question as $\left(D^G, D^K, D^L\right)$ and captures the individual firms' perceptions about demand (or lack thereof). Recalling that, in our model, the residuals of the moment condition are determined by the random demand shock z, it is reasonable to assume that, if the instruments are truly orthogonal to z, they should be uncorrelated with the dummy D^z.

The correlation matrix in table 3.6 shows that the dummy D^z is virtually uncorrelated with the other dummies D^F, D^G,D^K and D^L: while this is not a test of the exclusion restrictions in the strict sense, we take it as a reassuring sign that, at the least, there are no "red flags"of endogeneity in the data.

3.6 Estimation results

3.6.1 Reduced form evidence

In this section, we present the results from our econometric analysis. We start with some non-parametric evidence on the magnitude of distortions.

The key testable manifestation of heterogenous firm-level distortions is that firms with a higher expected wedge ($D = 1$) should display higher markups (P_i/c_i). Hence, before carrying out any complex estimation, it is useful to investigate how the distribution of the firms' markups varies by conditioning on the dummy vector D_i.

In table (3.7), we display estimation results, country-by-country, from regressing the firms' log markup (Revenues/Costs) on our constraint dummies. All regressions include sector fixed effects and are performed using sampling weights that correct for selection into the Amadeus sample. Our regression analysis indicates that are some differences in the distribution of markups across constrained and unconstrained firms, but these differences are small in economic terms. Considering only coefficients that are statistically significant, we find that Family Control predicts higher markups in Hungary (+1.4%) and Spain (+1.2%) and lower markups in the UK (-2.5%). The bureaucracy dummy D^G predicts lower markups in Germany (-2.2%). The Financial constraints dummy D^K predicts lower markups in Italy (-0.4%) and higher markups in Spain (+1.0%). Finally, the Labor regulation dummy D^L predicts higher markups in Italy (+0.3%). None of the estimated reaction coefficients is larger than 3% in absolute value.

In Figures 3.4-3.5, we display the distribution of the firms' log markup conditioning on different values of the dummies D_i. Each graph pools data from all the firms in our sample, and the markups are netted of country and sector fixed effects. On the whole, the density plots are suggestive that firms that face large positive distortions do display somewhat higher markups overall, but that the magnitude of the effect is modest, especially with respect to bureaucratic constraints $\left(D_i^G\right)$, where hardly any difference at all can be noticed between the markups of firms that report being constrained by bureaucracy and regulations, and those which do not.

3.6.2 Structural estimation

Table 3.8 presents our structural estimates of the distribution of the firm-level wedges.[6] The lower part of table presents estimates from step 1 of the estimation (carried out using MLE). It can be seen that the thresholds T^j , as well as the correlation ρ^{GL}, are all estimated with a high degree of precision, particularly for larger countries. This is highly desirable since, in our sequential estimation framework, the accuracy of the GMM standard errors in Step 2 relies on the accuracy of the estimates from Step 1. The empirical estimate of ρ^{GL} hovers around 45%, and is reasonably stable across countries.

In the upper part of table 3.8 we present results from GMM step 2 estimation. The standard deviations σ inform us directly about the extent of the firm-level distortions and their variation across countries. Because these parameters are positive by construction, the significance test is one sided (the alternative hypothesis is that they are strictly larger than zero).

Most of the variance estimates in the table are either zero or close to zero, with some exceptions. We find a statistically and economically significant impact of financial constraints in Spain: the estimated standard deviation of the wedges is 13%; this implies that a firm that is one standard deviation above the mean of the distribution of the capital wedges (which we previously normalized is zero) can expect to face an effective cost of capital that is 26% higher than a firm that is one standard deviation to the left of the mean.

We also estimate a statistically significant dispersion (0.7%) in family control wedges for Spain and Hungary.

Our country-level estimates of the elasticity of substitution of demand (η) range from

---

[6] The variance-covariance matrix of the GMM parameter vector estimate σ is given by $\mathbb{E}\left[\left(\frac{\partial u}{\partial \sigma} D\right) W \left(\frac{\partial u}{\partial \sigma} D\right)'\right]^{-1}$, where u are the moment equation residuals and W is the GMM weighting matrix. Because the σ parameters are by definition non-negative, the residual function $u(\sigma)$ is smooth and symmetric around a neighborhood of $\sigma = 0$, implying that the gradient $\frac{\partial u}{\partial \sigma}$ tends to zero as σ approaches zero. For this reason, whenever $\sigma = 0$ the standard error is also necessarily zero.

16 to 28. This estimate appears high when compared to estimates obtained from previous empirical work. However, context is needed. Recall that, in the absence of distortions, η determines the firms' markup over the unit cost ratio.

We require our estimate of η to be consistent with the observed markups. We assume, as do HK, that firms can choose capital flexibly (we wish to capture the effect of distortions on the long-run equilibrium). Therefore, our unit cost of output incorporates the cost of capital, and the markup for the typical firm is significantly smaller for us than for other empirical works that assume capital to be fixed or semi-fixed. As a result, our estimates of η has to adjust upwards to adequately describe the data, and to reflect of long-run demand conditions.

3.6.3 Reallocation gains

After estimating the distribution of τ, we compute the output gains from reallocating resources among firms.

In order to compute the counterfactual input demands at the firm level, we need to make some assumptions about the functional form of the production function. Consistently with HK and DHV, we assume it to be of the Cobb-Douglas form. This allows us to compute the firm-level input demand functions (detailed formulas are given in Appendix 3.13).

Because the production function is allowed (on purpose) to vary at the firm level, standard aggregation results based on the Cobb-Douglas form of the production function do not hold. As a consequence, there is no closed-form solution for aggregate prices or for an aggregate efficiency metric (such as aggregate *TFP*). This is where the inelastic input supply assumption comes into play. Using the firm-level input demand functions, we find input prices so that the aggregate demand of capital, labor and intermediate inputs is unchanged after removing the frictions. We do this numerically. Then, we know that any increase in aggregate real output is due exclusively to improvements in efficiency.

Because different firms have different output prices, we cannot simply sum Y_i over $\mathcal{I}$. However, we can use a technique similar to the one used by statistical agencies to evaluate real

GDP growth: we measure the aggregate output gain as a weighted average of the percentage firm-level increase in real output (from the distorted to the distorted equilibrium), weighting each firm by its expected revenues $\mathbb{E}(P_i Y_i)$. The rationale for using expected revenues instead of actual revenues is that the firms' input choices are only optimal ex-ante (i.e. before the shocks are realized).

The estimates are presented in Figure 3.8. Consistently with our estimates of the distribution of τ, reallocation gains are very small, ranging from 0% to less than 1% of aggregate output. They are

3.6.4 Discussion

The most remarkable thing about our estimates of reallocation gains is that, while they are non-trivial in an economic sense, they are small, in relative terms, when compared to the benchmark results of HK. This is at least in part to be expected: the central idea of our study is to not use the entire variation in firms' markup as a measure of distortions, but to only use the share of variance explained by our "Treatment" dummy variables.

There is, however, another factor that is at least as important in determining the difference in results: the fact that, in our case, there is very little variation in the data to be explained to begin with. In this subsection, we argue that this is due to the fact that our analysis focuses on the dispersion in markups as opposed to the dispersion in Total Factor Productivity Revenue (defined as TFP times output price): in particular, we believe that, as a sufficient statistics for the variance in wedges, the first is likely to embed a significant upward bias that does not affect, instead, the variance of the log markup.

We start by noting that, if there is no heterogeneity nor mismeasurement in the firms' production function parameters and input costs, the dispersion in both these statistics captures the dispersion in the wedges (as desired) as well as variation in demand shocks (which in our case is the residual). Consider the simplified case where there are only output wedges: then the cross-sectional variation in firm's markup is determined by the expected wedge $\overline{\tau}_i^Y$ and the shock z_i:

$$\log \frac{P_i Y_i}{c_i Y_i} = \text{constant} + z_i + \bar{\tau}_i^Y$$

the log of *TFPR* is equal to the log markup plus an error term ε_i

$$\log TFPR_i \;\equiv\; \log P_i A_i \;=\; \log\left(\frac{P_i}{c_i}\right) + \varepsilon_i$$

assuming that the production function is Cobb-Douglas (with output/input elasticities α)

$$\varepsilon_i = \alpha_i^K \log\left(\frac{r}{\alpha_i^K}\right) + \alpha_i^L \log\left(\frac{w}{\alpha_i^L}\right) + \alpha_i^X \log\left(\frac{p}{\alpha_i^X}\right)$$

In other words, mismeasurement of production elasticities and input costs affects the variance of $\log TFPR$, but not the variance of the log markup. This needs to be kept into account when comparing our estimates to HK's, since in HK the sector-level expenditure shares are used as the estimate of the output/input elasticities (firms are assumed to have the same production function). Any firm-level heterogeneity across these two dimensions, which cannot be directly measured in HK's data, will be captured be the ε term, increasing the overall dispersion. Last but not least, in HK, because of data availability, value added is used as a proxy for revenues, and the production function only takes labor and capital as inputs. As a consequence, we believe it is possible that their estimate of the variance of wedges might incorporate a significant upward bias.

In order to assess to what extent the differences in our estimates differ from HK due to the choice of using the total markup as opposed to value added-based *TFPR*, we computed the log *TFPR* following the Hsieh-Klenow methodology and plotted distribution next to the log markup, after deducting country and sector fixed effects. The results of this analysis are shown in Figure 3.10: the standard deviation of the log markup is about 0.083, while the standard deviation of the log *TFPR* is 0.532: over 6 times as large. Using the formula provided in HK's paper to compute reallocation gains, we obtain an upper bound for the reallocation gains (across all countries in our sample) that is remarkably close to their original estimate for the USA: 42.5%.[7] By contrast, applying the same formula to our observed dispersion in log markup (and using an average of our estimates of η), we obtain an upper bound of about 8%. This suggests that most of the difference in the size of our estimates is due to the choice of the key statistic for the variance of wedges, and in how it is computed.

As a result, we argue that one reason why our estimates are so different is that $\log TFPR$ is a valid metric of misallocation, only under he assumption that the production function is Cobb-Douglas and homogenous across firms; whereas the dispersion of log markups is valid for any CRS production function. This reflects the fact that our estimates of distortions and reallocation gains are robust to various specifications of the production technology.

Given that we find that a large fraction of the observed dispersion in markups cannot be attributed to distortions, it is natural to ask what are the sources of this large residual variation. This residual variation can be decomposed into within-firm and between-firm components, and the literature has proposed channels for both. Roughly 28% of the residual variation in markups can be attributed to within-firm variation in markups. As discussed in Asker, Collard-Wexler, and De Loecker (2014), within-firm variation in markups arise from dynamic production inputs and their associated adjustment costs. DHV focus on the impact of informational frictions on generating dispersion in marginal products of capital. In our model, this variation would carry over to firm markups, and thus is a likely culprit for residual variation. Systematic differences across firms, arising from differing industries and competitive environments, will naturally generate cross-sectional dispersion in markups as well.

3.7 Robustness

3.7.1 Production function and output price mismeasurement

Our econometric estimates of the distribution of firm-level wedges are, very much by construction, robust to misspecification of the firm-level production function and mismeasure-

[7] $\Delta \log GDP = \frac{\eta}{2} Var\left(\log TFPR\right), \quad \eta = 3$

ment of output prices. These are two of the most common sources of endogeneity in firm-level production models. This is a consequence of the fact that the moment condition used to perform the estimation does not require any information about output prices or the production function. All we require is the constant returns to scale assumption.

The reason why no information about output prices is needed is that the moment condition can be written entirely in terms of nominal amounts (revenues and expenditures). Standard procedures to estimate TFPQ and TFPR at the firm level require that the researcher deflate expenditures into quantity figures. This can be problematic since prices are not observed at the firm level; hence, the researcher frequently has to use sector-level price deflators, which introduces measurement error.

Finally, our residual source of variation (the moment condition residuals) is z_i: our estimation strategy is devised so that it does not enter our estimate of the distribution of firm-level wedges (unless of course there's a violation of the GMM orthogonality condition). In other words - all residual variation in the model is attributed to informational frictions (which are the focus of DHV). This last feature captures the essence of the direct measurement approach - which is to segregate the estimated impact of distortions from unexplained variation in the data.

3.7.2 Inputs fixedness and elasticity of substitution: long-run v/s short-run

In our baseline model, as in HK, all inputs are considered variable, including capital an non-production labor, and the production function is Cobb-Douglas. This might be a reasonable assumption if we are considering a long-run equilibrium, but not if we are trying to capture distortions that affect firms' size over the short run. In order to clarify this point, let us discuss the basic features of of long-run and short-run equilibria.

Firstly, in the short run, the elasticity of substitution between inputs is very low or zero (firms cannot vary their production technology in the short term). Therefore, while a Cobb-Douglas production function might be a reasonable representation of the a manufacturing firm's technology in the long run, a Leontief production function is a more reasonable representation of the firms' actual production technology in the short term.

Secondly, many inputs, notably capital, are fixed in the short run. Hence, while it might be reasonable to include capital among the variable inputs in the long run, the productivity-relevant cost base has to be narrower in the short run. The flip-side of a wider cost base in the long run is smaller markups, which in turn reflective the fact that the long-run demand curve is flat or nearly flat.

It is therefore possible that our negative results from section 3.6 might be caused by the fact that we are analyzing the firm's markups over the wrong time horizon. For this reason, we consider an alternative version of our model, which we call the "short run" model, which utilizes a different mappings of the model costs to the firms' costs.

This alternative mapping is also shown in Table 3.4. The firm's total variable costs are mapped to the firm's "Costs of goods sold" (COGS). Financial constraints are absent (since capital is not part of COGS). Labor market frictions are modeled as wedges that are symmetrically applied to all variable inputs (i.e. they are equivalent to output wedges): this is due to the fact that, in the short term, there is no substitutability between labor and other inputs.

Since the results of this alternative estimation, which we call the "Short-run model", carry a different interpretation from those of the long-run model (the one in which all costs are considered variable), they are presented and discussed separately.

Parameter estimates for this alternative model are presented in Table 3.9. In the short run, we find a statistically significant effect of family control in Germany and of labor market regulations in France and Italy.

3.7.3 Sector-level effects

One further potential pathology we want to address is the possibility that demand shocks are correlated, among firms, at the sector level. This possibility was already partially addressed by our avoiding to use data from recession years. However, if this was not sufficient, systematic residual variation in sector-level demand would obviously have the consequence of biasing our standard errors (due to clustering). Additional, if such systematic variation in demand conditions at the sector level happened to be correlated with our instruments (D), this would constitute a failure of our identification assumption.

In order to address these potential issues, we have devised a way to incorporate sector fixed effects in our estimation framework using Inverse Probability Weighting (IPW). The idea is to re-weight the data in order to account not just for sample selection, but also for selection into "treatment" of firms from different sectors. The re-weighting of the data basically ensures that different "treatment groups" (groups of firms with different values of the dummies D) are balanced with respect to the distribution of firms across sectors. In other words, after re-weighting, the firm's sector does not predict its values of the instruments D, although the distribution of firms is still representative of the population. This re-weighting procedure is explained in Appendix 3.10.

We re-estimated the both the long-run and the short-run variants of our model using IPW to control for sector fixed effects. The results of this exercise are presented in Appendix 3.14, Tables 3.10 and 3.11. As can be seen, our results only vary marginally.

3.7.4 Alternative models of Demand

Our model is based on that of HK, and assumes a constant elasticity of substitution across goods. As discussed in Joel M David and Venkateswaran (2017), the CES demand aggregator is unique in that it implies a constant markup for all firms. Although not quantitatively important for our results, Edmond, Midrigan, and Xu (2018) note that, in general, allocative efficiency does not occur when markups systematically vary with size. Although they model this effect using a Kimball aggregator, the observation that systematic variation in the elasticity of substitution can cause heterogeneity in markups applies in a variety of settings. Our choice to allow for unobserved heterogeneity in firm production technologies precludes our ability to explicitly model alternative aggregators, and so we instead do our best to control for size in our empirical work. Though we do not observe a strong positive relationship between markups and firm size in our data, we further control for any potential relationship by re-weighting our sample to explicitly account for potential variations in size. This has the effect that our estimates are of the component of markup dispersion in our data that can be attributed to our survey measures after controlling for firm size. Given the relationship between markups and size is not strong in our data, we do not believe this is significantly biasing our estimates of reallocation gains downwards. Thus we do our best to ensure our results are not too sensitive to our specification of a CES aggregator.

3.7.5 Using MLE instead of GMM

The GMM model outlined in section (3.4) uses the assumption that the wedges are jointly normally distributed, but not the normality assumption on the firm level shocks z. By using the parametric assumption on the distribution of the shocks, we can obtain an alternative, maximum likelihood estimator for the diagonal elements of the covariance matrix Σ^{τ}. The estimator maximizes the following likelihood function:

$$\log \phi \left(\frac{\Sigma^z}{2} + \log \frac{\eta - 1}{\eta} + \log \frac{\exp\left(-\bar{\tau}_i^F - \bar{\tau}_i^G\right) P_i Y_i}{\exp\left(\bar{\tau}_i^K\right) r K_i + \exp\left(\bar{\tau}_i^L\right) w L_i + p X_i}; \Sigma^z \right)$$

where $\phi\left(\cdot; \Sigma\right)$ is the standard normal *pdf* with mean zero and variance Σ.

We repeated our estimation in Appendix 3.15, tables 3.12-3.13, using MLE in both steps of the estimation procedure, rather than only in the first one. Our estimates barely changed.

3.8 Alternative Explanations

One concern is that our survey data is subject to measurement errors and therefore fails to capture much of the firm-level variation in exposure to frictions. So long as managers are reporting truthfully, the survey responses capture the distortions managers take into consideration when operating their firms, and thus measures constraints more directly than by relying on realized firm metrics. Given that we find managers' reports of weak demand are unrelated to their reports of constraints, we are not unduly concerned that managers are making false representations in their survey responses. However, noisy responses still have the effect of biasing our estimates towards zero, and may be part of the reason for our low estimates of the reallocation gains.

Another explanation for the magnitude of our estimates may be that our attempts to control for size and sector effects via reweighing are insufficient, and there is some additional relationship between size and the incidence of distortions that we have not captured. If this were the case, then smaller firms might be expected to grow even more in the distortionless economy, thus resulting in higher reallocation gains. While we attempt to check for this possibility via alternative weights, our attempts are limited by the data that we have, and we hope that future work can bring new, more comprehensive data to address these questions.

Finally, another limitation of our study is that, in our model, the production technology is exogenously determined. This has the implication that, in the distortionless economy, firms are constrained to using the same production technology as the one they used in the distorted one. When firms have access to and can choose among multiple production technologies, the removal of distortionary policies influences the decision of technology adoption, which may result in larger than estimated gains from the elimination of said policies.

3.9 Conclusions

We have used, for the first time, a combination of firm-level survey and accounting data to estimate the malallocative effects of four types of frictions: financial constraints, labor regulations, family control and bureaucracy. While we did find statistical evidence of an effect of financial frictions in Spain, labor regulation in France and Italy, and of family control in Germany and (to a much lesser extent) in Hungary, our data analysis suggests that the malallocative effect of these frictions can only explain a small fraction of cross-country differences in productivity. However, it appears that when information reported by firms themselves is used to estimate the impact of family control, financial frictions, labor regulation and bureaucracy, misallocation from these four types of frictions appears to be much smaller than previously suggested in the literature.

To conclude, we distill our methodological contribution into four key principles. Firstly, intermediate inputs should not be omitted from the production function; nor should they be subtracted from revenues in order to map output to value added. Secondly, the dispersion of log margin should be preferred over the dispersion of logTFPR as a sufficient statistic of misallocation: under relatively lax assumptions, the first is robust from misspecification of the firm-level production, while the latter is not. Thirdly, the elasticity of substitution of demand should be consistent with the relevant cost base (i.e. a demand elasticity of 3 is not appropriate if the cost base includes the capital input). Fourthly, the researcher should be wary of the distinction between the "true" wedges and the wedges perceived by the firms: where the two diverge, the latter determine the extent of misallocation.

Figure 3.1: Constraints to Growth in the EFIGE survey

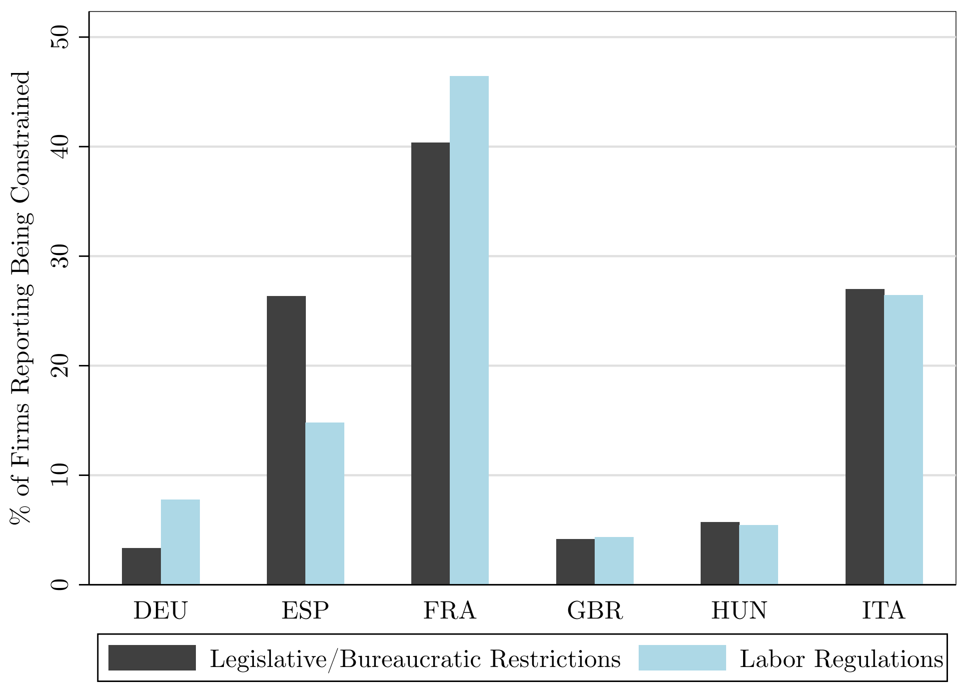


Figure 3.2: External Validation of the "Bureaucracy" dummy D^B

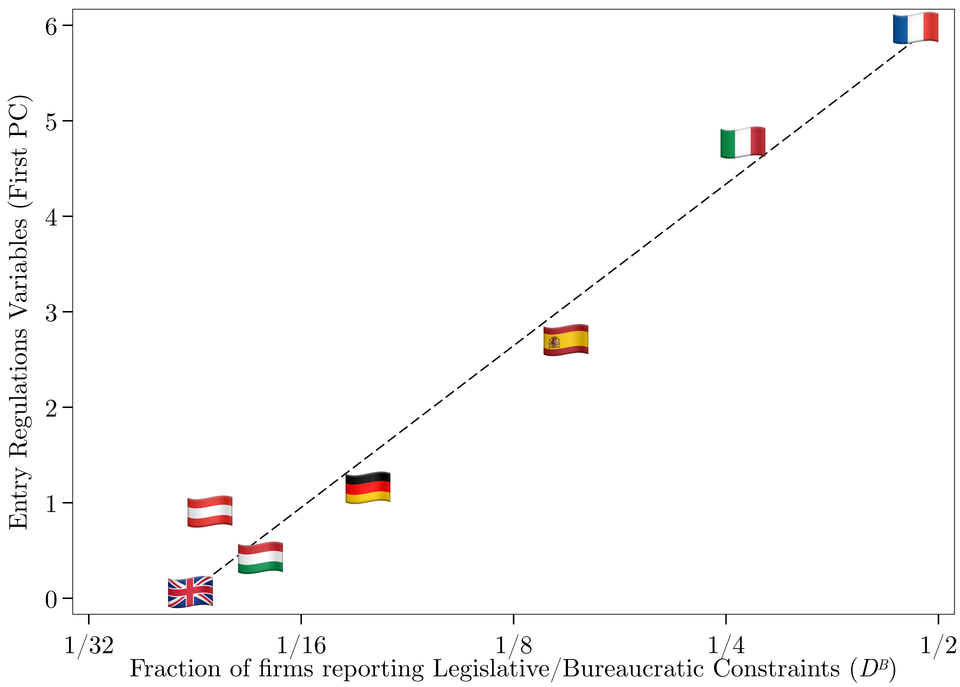


Figure 3.3: External validation of the "Labor Market Regulations" dummy D^L

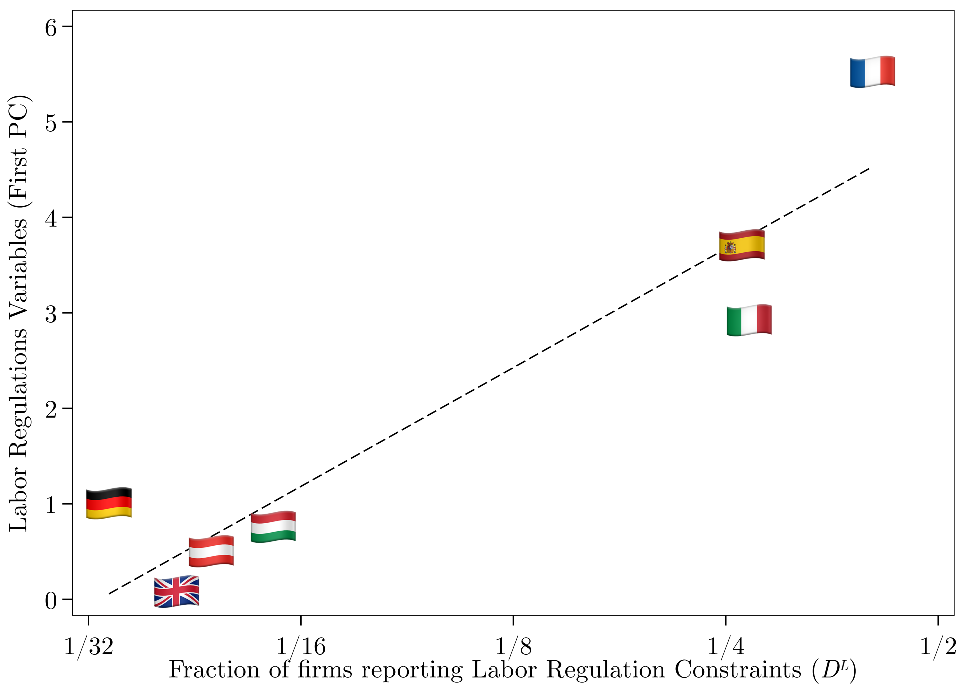
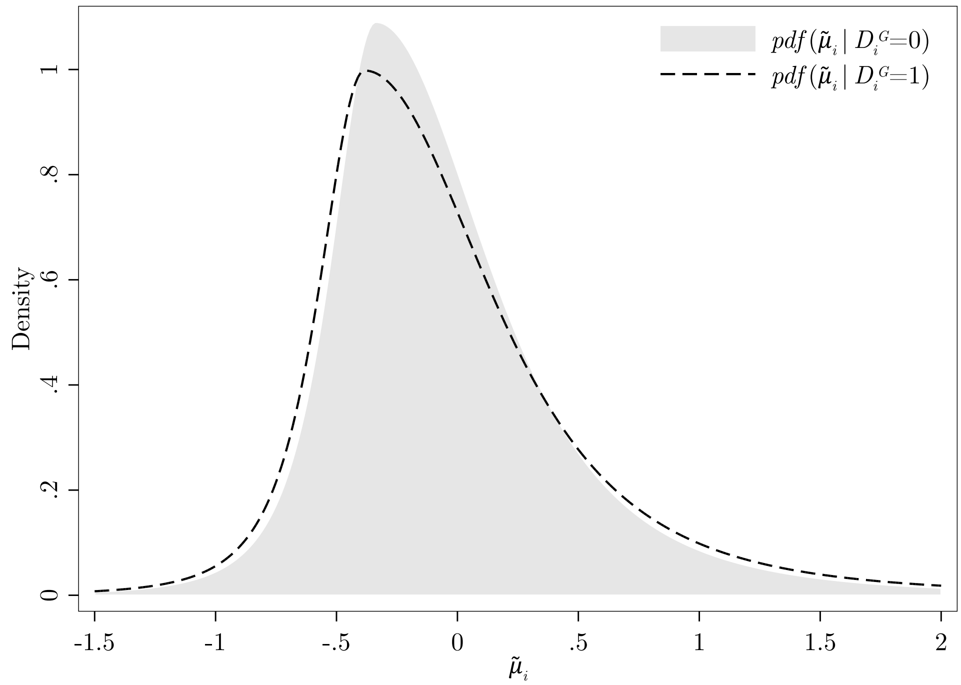


Figure 3.4: Markups and Legislative/Bureaucratic Constraints



NOTES: The picture above displays the conditional density distribution of log markups, residualized on country and sector fixed effects. The dotted line displays the density for firms reporting "Legislative/bureaucratic Constraints" in the EFIGE survey. The gray area displays the density for firms that do not report such constraints. Markups estimated using the method of De Loecker and Warzynski (2012).

Figure 3.5: Markups and Labor Market Regulations

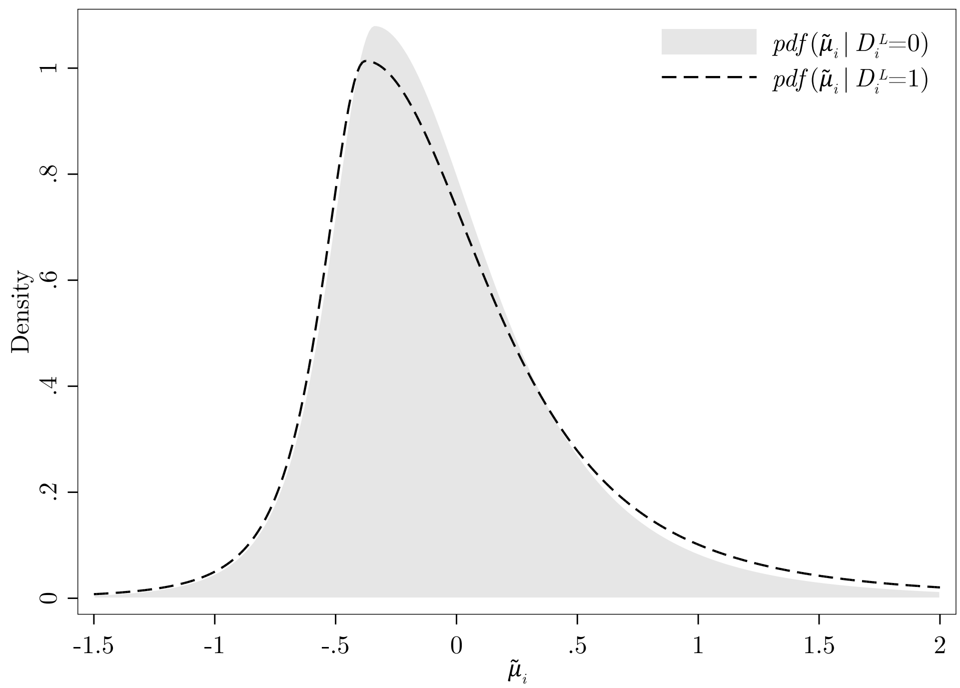


Figure 3.6: Productivity and Legislative/Bureaucratic Constraints

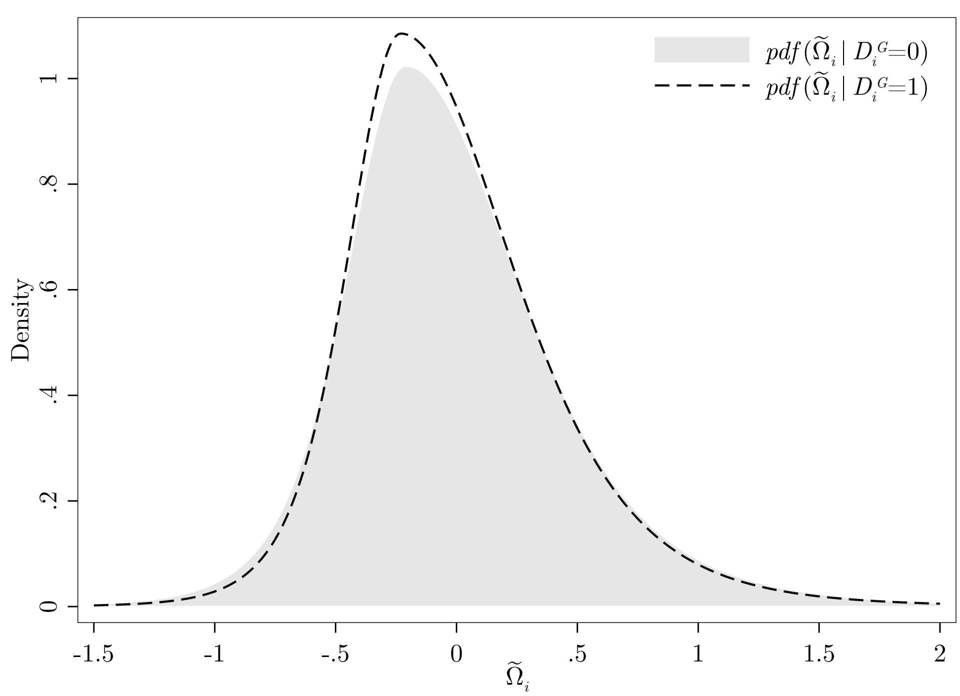


Figure 3.7: Productivity and Labor Market Regulations

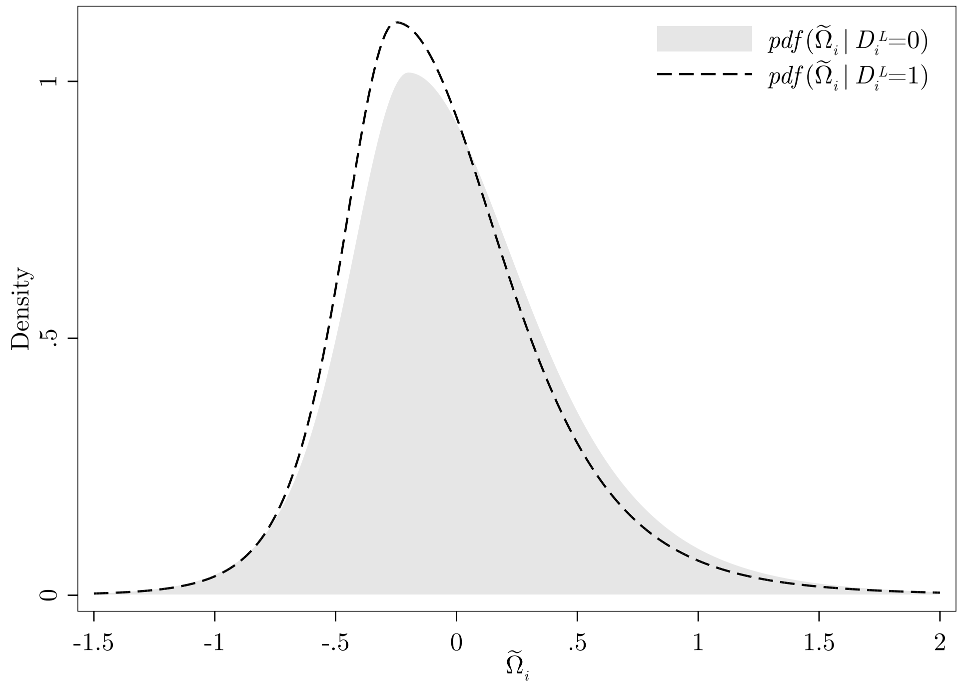
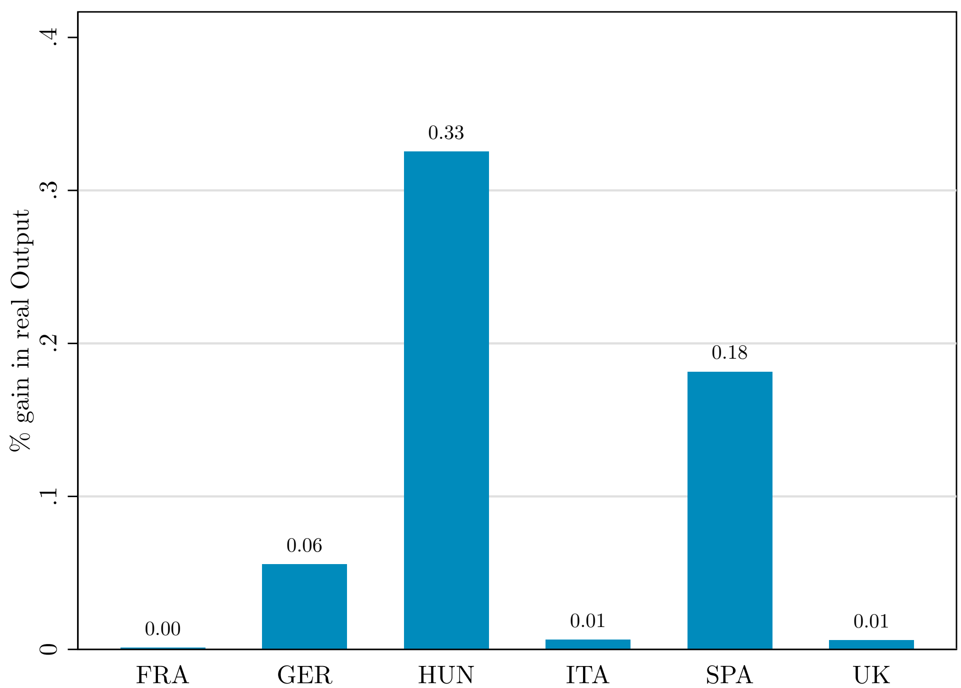


Figure 3.8: Reallocation Gains (by country) - long run model



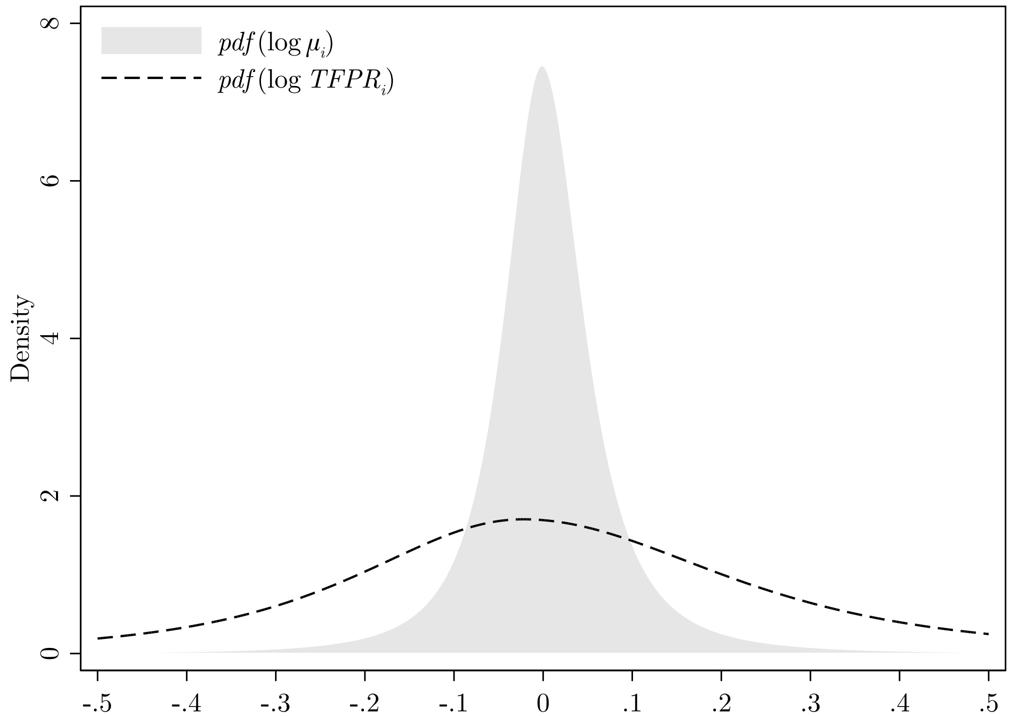


Figure 3.10: Comparison of the pdf of log markup and logTFPR (HK methodology)

Figure 3.9: Reallocation Gains (by country) - short run model

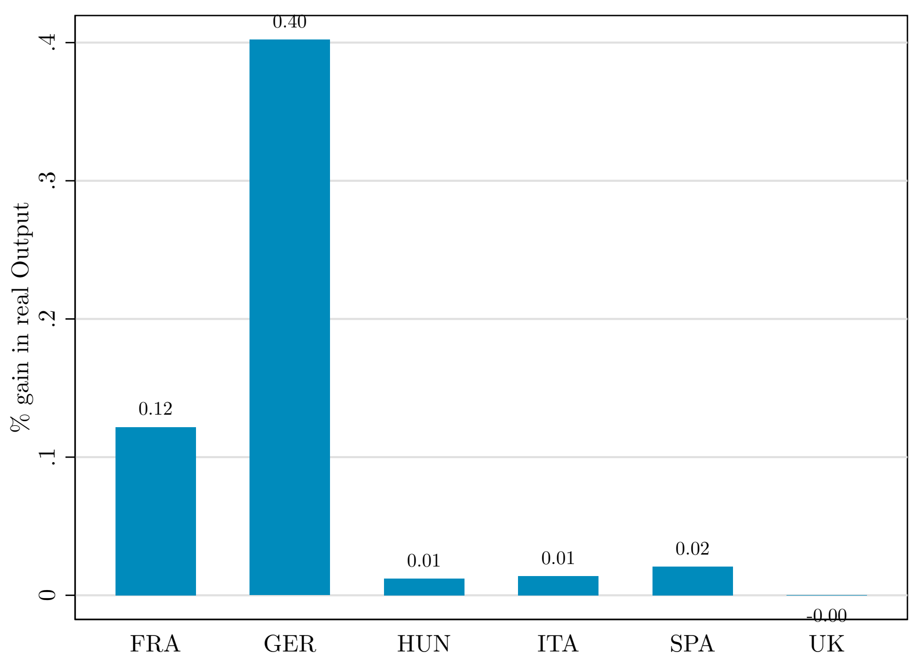


Table 3.1: Sample size and composition by country and employment - Full sample

Size Class	FRA	DEU	HUN	ITA	ESP	GBR	Total
10-19 persons employed	991	695	148	1,032	1,025	626	4,517
	33.7%	24.0%	30.6%	34.4%	36.5%	30.7%	31.9%
20-49 persons employed	1,138	1,118	174	1,396	1,233	793	5,852
	38.7%	38.6%	36.0%	46.5%	44.0%	38.9%	41.3%
50-249 persons employed	599	783	117	428	401	513	2,841
	20.4%	27.0%	24.2%	14.3%	14.3%	25.2%	20.1%
>250 persons employed	211	300	45	144	146	107	953
	7.2%	10.4%	9.3%	4.8%	5.2%	5.2%	6.7%
Total	2,939	2,896	484	3,000	2,805	2,039	14,163

Table 3.2: Sample size and composition by country and employment - BvD sample

Size Class	FRA	DEU	HUN	ITA	ESP	GBR	Total
10-19 persons employed	941	108	144	996	984	90	3,263
	34.1%	10.6%	31.5%	34.9%	37.1%	11.1%	30.9%
20-49 persons employed	1,070	350	166	1,335	1,166	251	4,338
	38.8%	34.2%	36.3%	46.7%	44.0%	30.9%	41.1%
50-249 persons employed	549	387	105	400	371	384	2,196
	19.9%	37.8%	23.0%	14.0%	14.0%	47.3%	20.8%
>250 persons employed	196	178	42	125	132	86	759
	7.1%	17.4%	9.2%	4.4%	5.0%	10.6%	7.2%
Total	2,756	1,023	457	2,856	2,653	811	10,556

Table 3.3: Sample size and composition by country and industry - BvD sample

Sector	Description	FRA	DEU	ESP	GBR	HUN	ITA	Total
10-12	Food products, beverages and tobacco	198	94	441	71	61	234	1,099
13-15	Textiles, wearing apparel, leather and related products	195	36	139	36	28	389	823
16-18	Wood and paper products; printing and reproduction of recorded media	304	96	292	110	60	250	1,112
19-21	Fuel, chemicals and chemical products	87	48	115	48	17	105	420
22-23	Rubber and plastics products, and other non-metallic mineral products	359	123	293	69	64	326	1,234
24-25	Basic metals and fabricated metal products, ex. machinery and equipment	865	232	585	113	105	649	2,549
26-27	Electrical and optical equipment	277	130	102	119	35	223	886
28	Machinery and equipment n.e.c.	239	171	282	93	40	356	1,181
29-30	Transport equipment	104	35	96	27	14	82	358
31-33	Other manufacturing; repair and installation of machinery and equipment	128	58	308	125	33	242	894
10-33	All Industries	2,756	1,023	2,653	811	457	2,856	10,556

Table 3.4: Mapping of model variables to data

Variable		Concept	Source
$P_{it}Y_{it}$	$\leftarrow$	Revenues	BvD Amadeus/ORBIS
$P_t^K K_{it}$	$\leftarrow$	Fixed assets (lagged)	BvD Amadeus/ORBIS
$W_t L_{it}$	$\leftarrow$	Costs of employees	BvD Amadeus/ORBIS
$C_t X_{it}$	$\leftarrow$	Material Inputs	BvD Amadeus/ORBIS
P_t	$\leftarrow$	Gross Output Price Index (GO_P)	OECD StAn / EU KLEMS
P_t^K	$\leftarrow$	GFCF Price Index (Ip_GFCF)	OECD StAn / EU KLEMS
W_t	$\leftarrow$	Labor Price Index (LAB/LAB_QI)	OECD StAn / EU KLEMS
C_t	$\leftarrow$	Intermediate Inputs Price Index (II_P)	OECD StAn / EU KLEMS
R_t	$\leftarrow$	Policy rate + SAFE Spread	ECB Survey on the Access to Finance of Enterprises
D_i^F	$\leftarrow$	Firm is controlled by a family-owned entity	EFIGE Survey
D_i^G	$\leftarrow$	Constraints to growth: Bureaucratic restrictions	EFIGE Survey
D_i^K	$\leftarrow$	Constraints to growth: Financial constraints*	EFIGE Survey
D_i^L	$\leftarrow$	Constraints to growth: Labor regulation	EFIGE Survey
D_i^z	$\leftarrow$	Constraints to growth: Lack of demand	EFIGE Survey

Table 3.5: Descriptive Statistics

Variable	Obs	Mean	StDev	Min	Max
D^F	14,316	0.699	0.459	0.000	1.000
D^G	12,001	0.214	0.410	0.000	1.000
D^K	12,001	0.092	0.290	0.000	1.000
D^L	12,001	0.195	0.396	0.000	1.000
D^z	11,589	0.427	0.495	0.000	1.000
r	6,560	0.274	0.142	0.039	0.989
PY	6,560	26,452.002	192,748.328	161.000	6,925,000.000
rK	6,560	1,197.564	10,400.587	0.083	396,698.906
wL	6,560	3,942.671	24,784.824	1.000	1,088,197.000
pX	6,560	20,491.098	161,787.781	55.000	6,736,910.000

Table 3.6: Correlation matrix of constraint dummies

	D^F	D^G	D^K	D^L	D^z
D^F	1.000				
D^G	0.004	1.000			
D^K	0.049	0.005	1.000		
D^L	0.029	0.432	0.002	1.000	
D^z	-0.021	0.037	-0.114	0.066	1.000

Table 3.7: OLS Regressions

Dependent Variable: $\log\left(\frac{P_i}{c_i}\right)$							
	Independent Variable	FRA	GER	HUN	ITA	SPA	UK
D^F	Family control wedges (Y)	0.002	0.001	0.014***	-0.001	0.012***	-0.025***
		(0.002)	(0.004)	(0.005)	(0.002)	(0.004)	(0.008)
D^G	Bureaucracy wedges (Y)	-0.001	-0.022***	-0.006	0.001	0.004	0.029
		(0.002)	(0.008)	(0.012)	(0.002)	(0.005)	(0.026)
D^K	Financial constraints (K)	-0.006*	-0.004	-0.000	-0.004**	0.010**	0.021
		(0.004)	(0.005)	(0.009)	(0.002)	(0.004)	(0.021)
D^L	Labor regulation wedges (L)	-0.001	0.018	-0.015	0.003**	0.001	-0.028
		(0.002)	(0.012)	(0.011)	(0.002)	(0.004)	(0.021)
	Sector Controls	Yes	Yes	Yes	Yes	Yes	Yes
	Observations	2,355	637	360	2,384	407	417
	R^2	0.072	0.228	0.098	0.102	0.094	0.212

Table 3.8: Demand and firm-level wedges distribution parameters: long-run model

	Parameter	Method	FRA	GER	HUN	ITA	SPA	UK
Demand (substitution) elasticity	η	GMM	29.120***	26.904***	107.403***	29.140***	50.627***	31.615***
			(0.879)	(2.101)	(28.240)	(0.666)	(4.418)	(6.959)
Family control wedges (Y)	σ^F	GMM	0.000	0.004	0.007**	0.000	0.007***	0.000
			(0.001)	(0.004)	(0.003)	(0.000)	(0.002)	(0.000)
Bureaucracy wedges (Y)	σ^G	GMM	0.000	0.000	0.000	0.000	0.000	0.004
			(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.009)
Financial constraints (K)	σ^K	GMM	0.000	0.000	0.000	0.000	0.130***	0.080
			(0.000)	(0.000)	(0.000)	(0.000)	(0.045)	(0.146)
Labor regulation wedges (L)	σ^L	GMM	0.000	0.000	0.000	0.010*	0.004	0.000
			(0.000)	(0.000)	(0.000)	(0.005)	(0.009)	(0.000)
Observations	N	GMM	2,355	637	360	2,384	407	417
Family control wedges (Y)	T^F/σ^F	MLE	-0.191***	-0.972***	-0.148**	-0.736***	-0.710***	-0.361***
			(0.027)	(0.034)	(0.067)	(0.028)	(0.065)	(0.030)
Bureaucracy wedges (Y)	T^G/σ^G	MLE	0.113***	1.432***	1.628***	0.622***	1.059***	1.726***
			(0.027)	(0.042)	(0.111)	(0.028)	(0.076)	(0.052)
Financial constraints (K)	T^K/σ^K	MLE	1.637***	0.902***	1.585***	1.256***	1.070***	1.893***
			(0.045)	(0.035)	(0.108)	(0.036)	(0.071)	(0.058)
Labor regulation wedges (L)	T^L/σ^L	MLE	0.267***	1.839***	1.585***	0.593***	0.604***	1.728***
			(0.027)	(0.052)	(0.112)	(0.028)	(0.067)	(0.052)
Correlation of $\left(\tau^G, \tau^L\right)$	ρ^{GL}	MLE	0.706***	0.572***	0.431**	0.522***	0.510***	0.241**
			(0.021)	(0.056)	(0.168)	(0.030)	(0.079)	(0.094)
Observations	N	MLE	2,973	2,935	488	3,020	518	2,067

Table 3.9: Demand and firm-level wedges distribution parameters: short-run model

	Parameter	Method	FRA	GER	HUN	ITA	SPA	UK
Demand (substitution) elasticity	η	GMM	2.521***	2.982***	3.322***	3.287***	2.755***	3.266***
			(0.021)	(0.101)	(0.087)	(0.033)	(0.057)	(0.118)
Family control wedges (Y)	σ^F	GMM	0.003	0.041	0.007	0.000	0.000	0.000
			(0.007)	(0.028)	(0.015)	(0.000)	(0.000)	(0.000)
Bureaucracy wedges (Y)	σ^G	GMM	0.000	0.000	0.000	0.000	0.000	0.000
			(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Labor regulation wedges (Y)	σ^L	GMM	0.023***	0.000	0.007	0.010*	0.014	0.000
			(0.007)	(0.000)	(0.024)	(0.006)	(0.015)	(0.000)
Observations	N	GMM	2,355	637	360	2,384	407	417
Family control wedges (Y)	T^F/σ^F	MLE	-0.191***	-0.972***	-0.148**	-0.736***	-0.710***	-0.361***
			(0.027)	(0.034)	(0.067)	(0.028)	(0.065)	(0.030)
Bureaucracy wedges (Y)	T^G/σ^G	MLE	0.113***	1.432***	1.628***	0.622***	1.059***	1.726***
			(0.027)	(0.042)	(0.111)	(0.028)	(0.076)	(0.052)
Labor regulation wedges (Y)	T^L/σ^L	MLE	0.267***	1.839***	1.585***	0.593***	0.604***	1.728***
			(0.027)	(0.052)	(0.112)	(0.028)	(0.067)	(0.052)
Correlation of (τ^G, τ^L)	ρ^{GL}	MLE	0.706***	0.572***	0.431**	0.522***	0.510***	0.241**
			(0.021)	(0.056)	(0.168)	(0.030)	(0.079)	(0.094)
Observations	N	MLE	2,973	2,935	488	3,020	518	2,067

3.10 Appendix: sample selection and representativeness

In this appendix, we explain how we use Inverse Probability Weighting (IPW) to correct for sample selection for two countries in our dataset (Germany and the UK) for which this problem appears to be important. We also present graphs showing the results of this reweighting procedure.

As said previously, the Bruegel-Unicredit EFIGE dataset is comprised of two parts. The first is cross-sectional (firm-level) survey data from the EFIGE survey. The second is firm financials panel data that has been merged from the BvD Amadeus databank. The EFIGE sample of firms is significantly smaller than the Amadeus dataset. There are about 14,000 firms in EFIGE, and several hundred thousands in the Amadeus databank. The ones that are matched to EFIGE survey data represent a small subsample of Amadeus.

The EFIGE survey was administered and equipped with sampling weights in a way that would ensure that the (weighted) distribution of firms across size (number of employees) and industry matches the one reported by Eurostat for each of the 7 countries in the survey.

To state this explicitly, let N be a firm-level variable describing the size of the firm in terms of number of employees, S a categorical firm-level variable indicating the sector in which the firm operates, and E a dummy variable that indicates whether the firm was sampled for the EFIGE survey. Then the sampling weight of a firm of size N operating in sector S is:

$$w_E(N,S) = \frac{f(N,S)}{f(N,S\,|E=1)}$$

where $f(N,S)$ is the size/sector probability distribution of firms in the population (assumed to be the same as the one reported by Eurostat) and $f(N,S|\,E)$ is the size/sector probability distribution of firms conditioning on whether they were included in the EFIGE sample.

The Amadeus part of the dataset has known issues of representation and sample selection. Specifically, financial data is missing for a number of firms, for reasons not explicitly stated by the vendor. Because, generally, observations are dropped in analyses that use financial data (such as our own GMM procedure in this paper), this poses a problem of sample selection.

Another potential source of sample selection that is unrelated to the construction of the Amadeus dataset has to do with the possibility that some survey data used in the estimation might be missing from the EFIGE data to begin with.

Going forward, when we refer to “representativeness”, it must be understood that we refer to representativeness with respect to the original survey stratification variables (size and sector).

First, we note that data coverage differs significantly, and for different reasons, across countries. Coverage is about 95% for France, Hungary and Italy, about 25% for Germany and UK, and about 18% for Spain. For Germany and the UK the low coverage is due to missing data from the Amadeus databank. For Spain, instead, the low coverage is due to some EFIGE survey questions missing (specifically, questions that are used in the estimation).

The thing that allows us to analyze and correct for sample selection is that the stratification variables (size and sector) are observed for all firms in the EFIGE survey. Because of this fact we can compare the size/sector distribution of firms conditional on BvD financials data availability $f\left(N, S\,|B=1\right)$ with the unconditional distribution $f\left(N, S\right)$.

In Figures A1 and A3, we compare graphically the marginal conditional and unconditional distributions

$$f\left(N\,|B=1\right) \qquad v/s \qquad f\left(N\right)$$
$$f\left(S\,|B=1\right) \qquad v/s \qquad f\left(S\right)$$

these are estimating using the EFIGE sampling weights (w_E). What can clearly be seen from these graphs is that there’s clear evidence of selection on size for German and British firms, and somewhat less clear evidence of selection on sector. There is no evidence of sample selection for other countries.

Based on this analysis, we have produced a probability weight variable specifically for German and UK firms, which is used for all analyses that require availability of financial data. Using Bayes' rule, the variable can be computed as follows:

$$\begin{aligned} w_B(N,S) &= w_E(N,S) \cdot \frac{f(N,S)}{f(N,S\,|B=1)} \\ &= w_E(N,S) \cdot \frac{f(N,S)}{f(B=1\,|N,S)} \cdot \frac{f(B=1)}{f(N,S)} \\ &= w_E(N,S) \cdot \frac{f(B=1)}{f(B=1\,|N,S)} \end{aligned}$$

In order to estimate $f(B=1)$ and $f(B=1\,|N,S)$, we estimate two probit models in which observations are weighted using EFIGE sampling weights (w_E): the first only includes a constant as independent variable, while the second includes the (log) number of employees and sector fixed effects.

Following the computation of the weighting variable, we reproduce graphs A1 and A3 in order to verify the effectiveness of the reweighing: the only difference being that this time we estimate the conditional distributions $f(N\,|B=1)$ and $f(S\,|B=1)$ for Germany and UK using the new weight variable w_B instead of w_E.

It can be seen that, following the re-weighting, the conditional distribution of firms across sizes and sectors matches closely the un-conditional ones, suggesting that the re-weighting has computed correctly. The reason that the two distributions are not exactly the same following the reweighing is due to the fact the probit regression in which we estimated $f(B=1\,|N,S)$ did not include the interaction of log employees and sector fixed effects (we excluded it in order not to over-fit the model).

Taking this logic one step further, we have also produced a third set of weights (for all countries in the dataset) which we call "control weights". The aim of these weights is to provide a simple and intuitive way to control for size and sector when estimating the effect of variables D^K, D^L, D^Y using GMM, without having to modify the moment conditions. Intuitively, by applying these weights during GMM estimation, we ensure that the 8 "treatment groups" of firms (each potential value of the D vector defines a group) are balanced in terms of size and sector distribution of firms.

$$w_C\left(N,S,D\right)=\frac{f\left(D\right)}{f\left(D\left|N,S\right.\right)}w_E\left(N,S\right)$$

The distributions $f\left(D\right)$ and $f\left(D\left|N,S\right.\right)$ are estimated using a multinomial logit regression of the vector $D=\left(D^K\ D^L\ D^Y\right)$, expressed as a categorical variable, on country fixed effects, with and without S and log$N$ as explanatory variables.

FIGURE A1: FIRMS' SIZE DISTRIBUTION, BEFORE RE-WEIGHTING

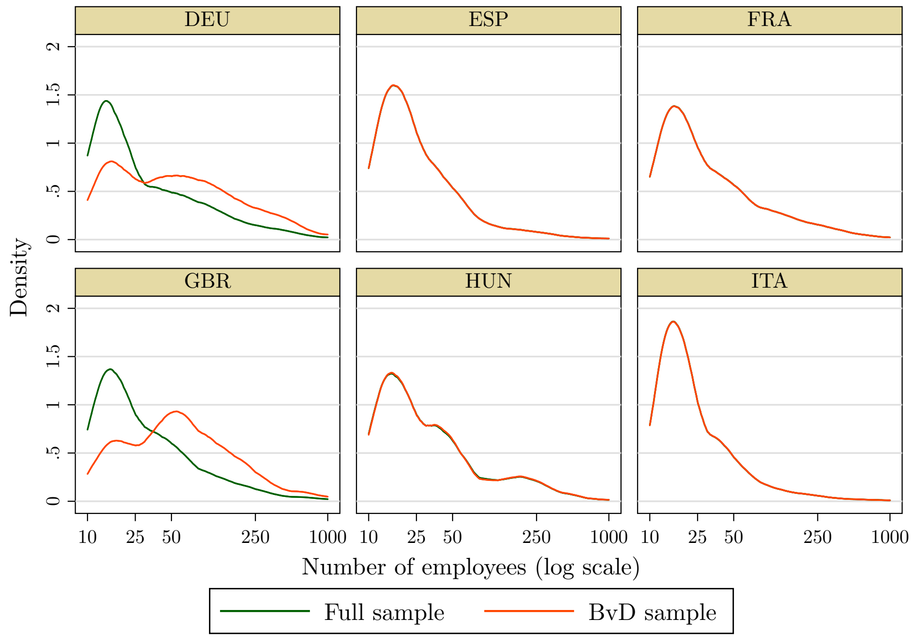


FIGURE A2: FIRMS' SIZE DISTRIBUTION, AFTER RE-WEIGHTING

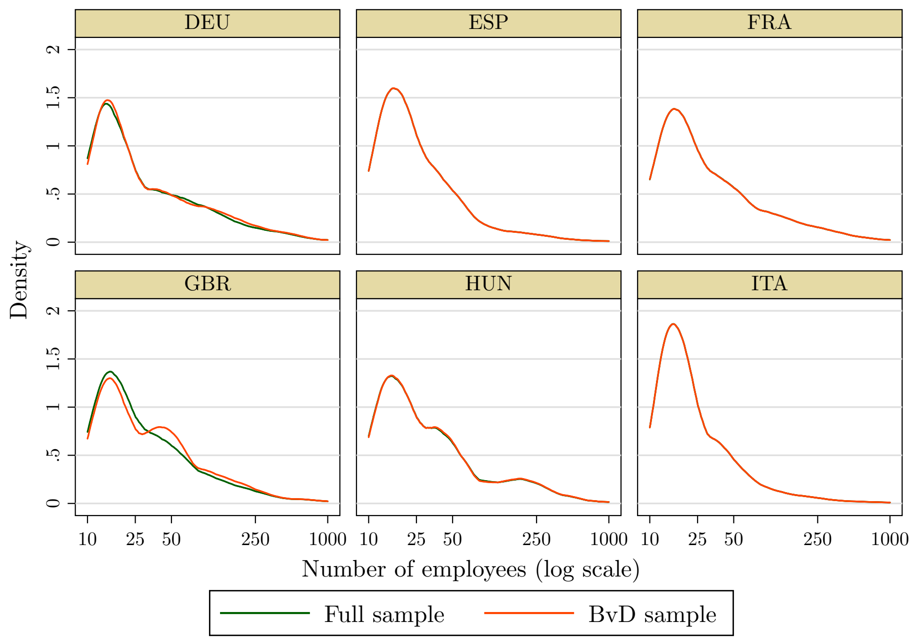


Figure A3: firms' sectoral distribution, before re-weighting

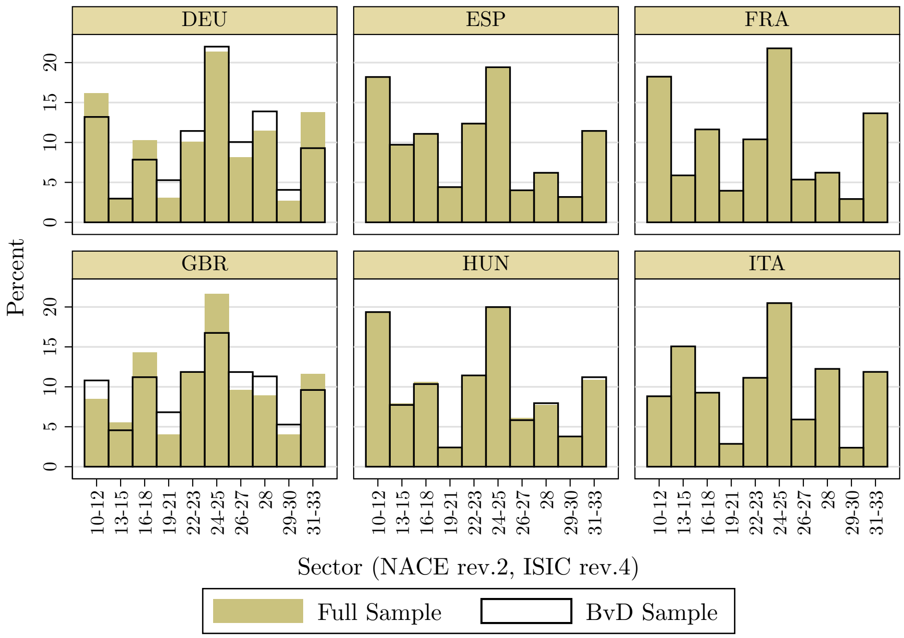
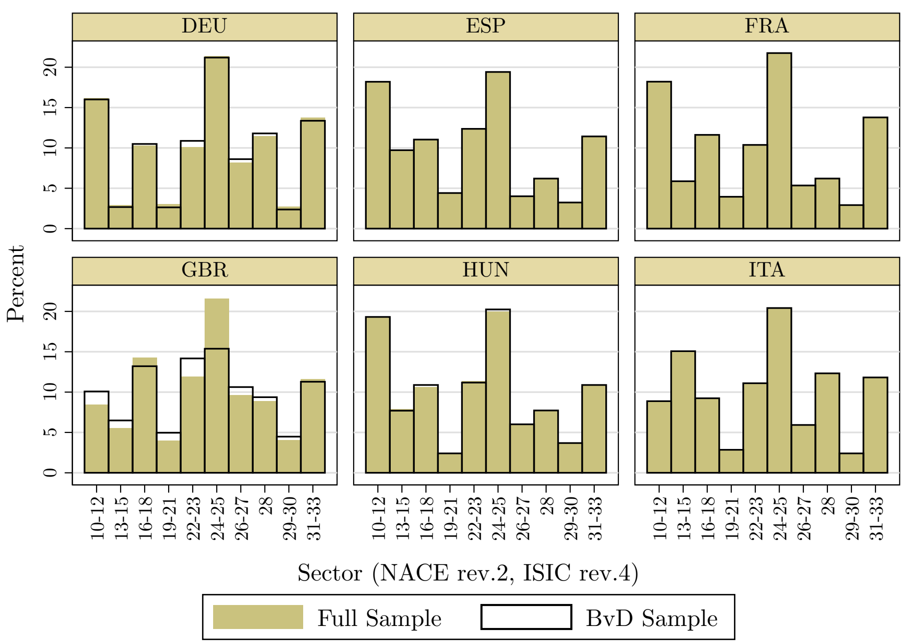


Figure A4: firms' sectoral distribution, after re-weighting



3.11 Appendix: normalization choice for the distribution of wedges

In this appendix, we discuss in further detail our choice of normalizing the distribution of firm-level wedges to be centered around the origin. As mentioned previously, we interpret the constraint dummy D_i as follows (we ignore going forward the distinction between output, capital and labor wedges): it is equal to one if the wedge on firm i that is positive and large . Specifically, the firm reports being constrained if the wedge is larger than a certain reporting threshold:

$$D_i \triangleq \mathbb{I}\left\{\tau_i > T\right\}$$

where $\mathbb{I}\left\{\cdot\right\}$ denotes the indicator function. The threshold T is a parameter to be estimated that is not identified separately from the mean of τ_i's distribution; hence, a normalization assumption such as $\mu^\tau = 0$ is required to identify T; under this mapping, the percentage of constrained firms identifies the ratio of the threshold to the standard deviation:

$$\mathbb{P}\left(\tau > 0\right) = 1 - \Phi\left(\frac{T}{\sigma^\tau}\right)$$

However, this is not the only modeling option available. A second, more trivial interpretation of these dummy variables is that it is equal to one when the firm faces a positive wedge:

$$D_i \triangleq \mathbb{I}\left\{\tau > 0\right\}$$

In the latter case, we can assume the distribution of τ_i to have a mean different from zero that can be estimated; consequently, the percentage of constrained firms in the population identifies the mean/standard deviation ratio of the distribution of τ:

$$\mathbb{P}\left(\tau > 0\right) = \Phi\left(\frac{\mu^\tau}{\sigma^\tau}\right)$$

In order to justify our choice to use the first mapping, we first need to explain how such mapping affects the inferred distribution of wedges. The consequences of the modeling choice is exemplified by the graph below. Here, we assume that, for a hypothetical set of firms and a generic wedge τ,

$$\mathbb{P}(\tau > 0) = 10\% \qquad \text{and} \qquad \sigma^{\tau} = 1$$

The blue curve shows the distribution of τ that is inferred under the latter mapping, while the orange curve represents the one that is inferred under the "threshold" model that we chose to use. The highlighted areas below the two curves correspond to the firms that report $D_i = 1$. The two distributions are similar in that they are both Gaussian with variance one, and differ in that the one plotted in orange is symmetric around zero, while the one in blue is shifted to the left.

Figure B1: Difference between two possible normalizations

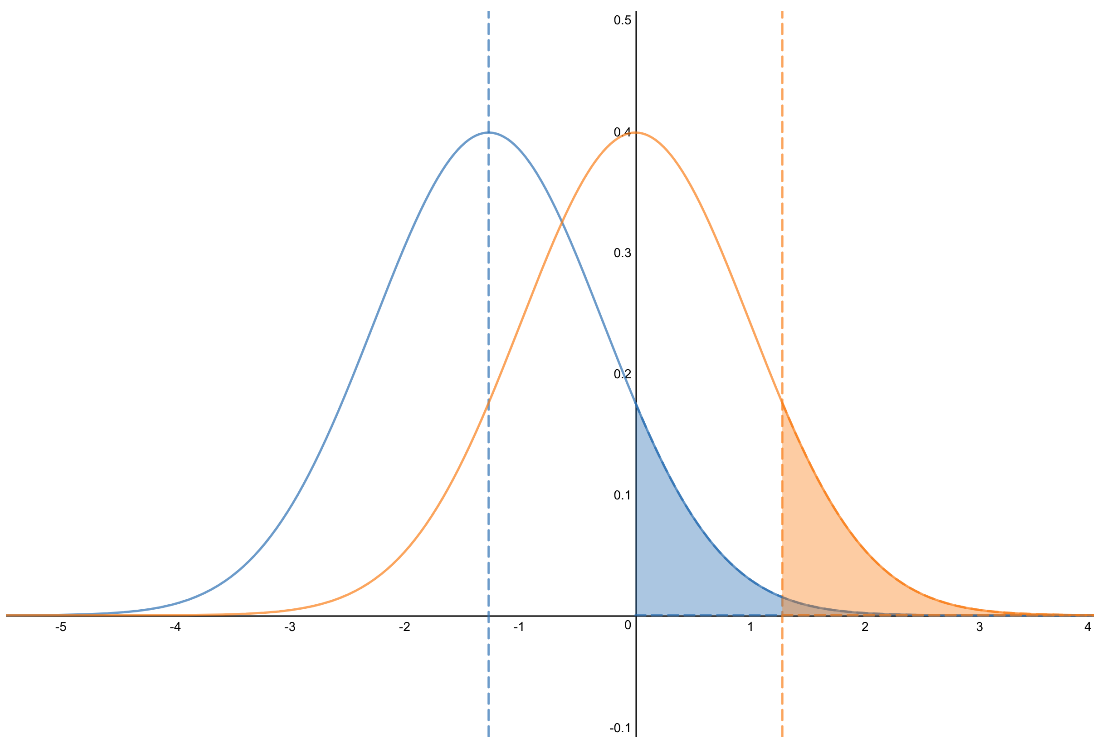


Figure 3.11

Following this explanation, we can now illustrate our rationale for opting for the latter modeling assumption. Notice that one of the consequences of making the first modeling assumption is that a value of $\mathbb{P}(\tau > 0)$ smaller than 50% is necessarily associated with a negative value of μ^τ. Then consider that, for all combination of countries and types of constraints in our dataset, except one, this percentage is lower than 50%. This would imply that, under the first modeling choice, the average wedge would nearly always be negative, In other words, it would mean that financial frictions and government policies would, on average, facilitate firms' growth, an undoubtedly odd finding.

Another reason why we choose the second mapping has to do with how different values of $\mathbb{P}(\tau > 0)$ translate, from an intuitive standpoint, into different shapes of the distribution of τ.

Under the alternative mapping, for a fixed value of μ^τ, a lower value of $\mathbb{P}(\tau > 0)$ is generally interpreted as a more negative wedge for the average firm (in other words: the distribution shifts left). Under the chosen mapping, for a given threshold T, a lower value of $\mathbb{P}(\tau > 0)$ reflects instead fewer firms facing a large positive wedge. We find the latter to be a more natural and intuitively appealing interpretation.

Finally, the third and last reason why we chose the threshold model is that it has the nice property that, for a given value of $\mathbb{P}(\tau > 0)$, a larger estimate of the standard deviation σ^τ does not automatically translate into a more negative mean μ^τ, which is instead the case for the alternative model. This is due to the fact that, in the alternative model, the ratio $\frac{\mu^\tau}{\sigma^\tau}$ is pinned down by the empirical value of $\mathbb{P}(\tau > 0)$.

3.12 Appendix: moment conditions for GMM estimation

In this appendix, we illustrate in more detail the methodology of our econometric estimation. We start by stating the following basic result about expectations.

Lemma 1. *Let X be a random variable distributed according to a multivariate log-normal distribution with parameters μ and $\mathbf{\Sigma}$. Then,*

$$\mathbb{E}\left[e^{t'X}|\mathbf{a} \leq X \leq \mathbf{b}\right] = \frac{\mathbb{P}\left[\mathbf{a} - \Sigma t \leq X \leq \mathbf{b} - \Sigma t\right]}{\mathbb{P}\left[\mathbf{a} \leq X \leq \mathbf{b}\right]}\mathbb{E}\left[e^{t'X}\right]$$

Using this result, we can now go on and write out extensively the moment conditions we use to estimate our model. As stated previously, we use GMM to retrieve parameters of the distribution of the vector $\left(\tau_i^Y \; \tau_i^K \; \tau_i^L\right)$. Throughout, subscripts on variables index firm-specific realizations, while superscripts index distortions. The moment condition that we get as a result of partial equilibrium in our model is:

$$\mathbb{E}\left[\frac{\eta-1}{\eta} \cdot \frac{P_i Y_i}{\exp\left(\overline{\tau}_i^F + \overline{\tau}_i^G\right)} - \exp\left(\overline{\tau}_i^K\right) r_i K_i - \exp\left(\overline{\tau}_i^L\right) w L_i - pX_i \,\middle|\, D_i\right] = 0, \qquad (3.9)$$

recalling that we define

$$\begin{aligned}\overline{\tau}_i^j &= -\log \mathbb{E}_i\left[\exp\left(-\tau_i^j\right)\right] \text{for } j \in \{F, G\} \\ \overline{\tau}_i^\ell &= \log \mathbb{E}_i\left[\exp\left(\tau_i^\ell\right)\right] \text{for } j \in \{K, L\}\end{aligned}$$

where the expectation is taken with respect to firm i's information set. Assume now that the vector of input distortions τ is distributed according to a multivariate log-normal, with mean 0 and variance-covariance matrix Σ. Furthermore, assume that we observe indicator variables $D^j \triangleq \mathbb{I}_{\tau^j > T^j}$, so that firm i's information set is $\mathcal{I}_i = \{D_i^j\}$, where $j \in \{F,\, G,\, K,\, L\}$.

We can immediately recover the ratio T^i/σ^i from the frequency of the indicator variables

D^i, using the relation

$$\begin{aligned}\mathbb{E}\left[D^i\right] &= \mathbb{P}\left[\tau^i \geq T^i\right] \\ &= 1 - \Phi\left(\frac{T^i}{\sigma^i}\right)\end{aligned}$$

where Φ denotes the cumulative distribution function of a standard normal. We recover the off-diagonal elements of the correlation matrix $\rho(\Sigma)$ using the relation

$$\mathbb{E}\left[D^i D^j\right] = \mathbb{P}\left[\tau^i \geq T^i \ \wedge \ \tau^j \geq T^j\right] \tag{3.10}$$

$$= 1 - \mathbb{P}\left[\tau^i < T^i \ \vee \ \tau^j < T^j\right] \tag{3.11}$$

$$= 1 - \left[\Phi\left(\frac{T^i}{\sigma^i}\right) + \Phi\left(\frac{T^j}{\sigma^j}\right) - \Phi_2\left(\frac{T^i}{\sigma^i}, \frac{T^j}{\sigma^j}; \rho_{ij}\right)\right] \tag{3.12}$$

Having estimated the parameters of the correlation matrix $\rho(\Sigma)$, we form the moment conditions used in our GMM estimation procedure. To recover the expectations of a truncated log-normal random variable, we use Lemma 1 above.

For example, to calculate $\exp\left(-\overline{\tau}^G\right)$ given $D^G = 0$, $D^L = 0$, we apply Lemma 1 using $t = \begin{pmatrix} 0 & -1 & 0 & 0 \end{pmatrix}'$, so that

$$\exp\left(-\overline{\tau}_i^G\right) = \frac{\mathbb{P}\left[\tau^G < T^G + \Sigma_{GG},\, \tau^L < T^L + \Sigma_{GL}\right]}{\mathbb{P}\left[\tau^G < T^G,\, \tau^L < T^L\right]} \exp\left(\frac{1}{2}\Sigma_{GG}\right) \tag{3.13}$$

$$= \frac{\Phi_2\left(\frac{T^G + \Sigma_{GG}}{\sqrt{\Sigma_{GG}}}, \frac{T^L + \Sigma_{GL}}{\sqrt{\Sigma_{LL}}};\ \rho_{GL}\right)}{\Phi_2\left(\frac{T^G}{\sqrt{\Sigma_{GG}}}, \frac{T^L}{\sqrt{\Sigma_{LL}}};\ \rho_{GL}\right)} \exp\left(\frac{1}{2}\Sigma_{GG}\right) \tag{3.14}$$

where Σ_{ij} denotes the ij-th element of the variance-covariance matrix Σ. Following the same procedure, we have

$$\exp\left(-\overline{\tau}^{i}\right)=\begin{cases}\dfrac{\Phi_2\left(\frac{T^i+\Sigma_{ii}}{\sqrt{\Sigma_{ii}}},\frac{T^j+\Sigma_{ij}}{\sqrt{\Sigma_{jj}}};\rho_{ij}\right)}{\Phi_2\left(\frac{T^i}{\sqrt{\Sigma_{ii}}},\frac{T^j}{\sqrt{\Sigma_{jj}}};\rho_{ij}\right)}\exp\left(\frac{1}{2}\Sigma_{ii}\right) & \text{if } D^i=0,\ D^j=0,\\ \dfrac{\Phi\left(\frac{T^i+\Sigma_{ii}}{\sqrt{\Sigma_{ii}}}\right)-\Phi_2\left(\frac{T^i+\Sigma_{ii}}{\sqrt{\Sigma_{ii}}},\frac{T^j+\Sigma_{ij}}{\sqrt{\Sigma_{jj}}};\rho_{ij}\right)}{\Phi\left(\frac{T^i}{\sqrt{\Sigma_{ii}}}\right)-\Phi_2\left(\frac{T^i}{\sqrt{\Sigma_{ii}}},\frac{T^j}{\sqrt{\Sigma_{jj}}};\rho_{ij}\right)}\exp\left(\frac{1}{2}\Sigma_{ii}\right) & \text{if } D^i=0,\ D^j=1,\\ \dfrac{\Phi\left(\frac{T^j+\Sigma_{ij}}{\sqrt{\Sigma_{jj}}}\right)-\Phi_2\left(\frac{T^i+\Sigma_{ii}}{\sqrt{\Sigma_{ii}}},\frac{T^j+\Sigma_{ij}}{\sqrt{\Sigma_{jj}}};\rho_{ij}\right)}{\Phi\left(\frac{T^j}{\sqrt{\Sigma_{jj}}}\right)-\Phi_2\left(\frac{T^i}{\sqrt{\Sigma_{ii}}},\frac{T^j}{\sqrt{\Sigma_{jj}}};\rho_{ij}\right)}\exp\left(\frac{1}{2}\Sigma_{ii}\right) & \text{if } D^i=1,\ D^j=0,\\ \dfrac{1-\left(\Phi\left(\frac{T^i+\Sigma_{ii}}{\sqrt{\Sigma_{ii}}}\right)+\Phi\left(\frac{T^j+\Sigma_{ij}}{\sqrt{\Sigma_{jj}}}\right)-\Phi_2\left(\frac{T^i+\Sigma_{ii}}{\sqrt{\Sigma_{ii}}},\frac{T^j+\Sigma_{ij}}{\sqrt{\Sigma_{jj}}};\rho_{ij}\right)\right)}{1-\left(\Phi\left(\frac{T^i}{\sqrt{\Sigma_{ii}}}\right)+\Phi\left(\frac{T^j}{\sqrt{\Sigma_{jj}}}\right)-\Phi_2\left(\frac{T^i}{\sqrt{\Sigma_{ii}}},\frac{T^j}{\sqrt{\Sigma_{jj}}};\rho_{ij}\right)\right)}\exp\left(\frac{1}{2}\Sigma_{ii}\right) & \text{if } D^i=1,\ D^j=1\end{cases}\tag{3.15}$$

for $i\in\{F,G\}$ and $j\in\{F,\ G,\ K,\ L\}$, and

$$\exp\left(\overline{\tau}^{i}\right)=\begin{cases}\dfrac{\Phi_2\left(\frac{T^i-\Sigma_{ii}}{\sqrt{\Sigma_{ii}}},\frac{T^j-\Sigma_{ij}}{\sqrt{\Sigma_{jj}}};\rho_{ij}\right)}{\Phi_2\left(\frac{T^i}{\sqrt{\Sigma_{ii}}},\frac{T^j}{\sqrt{\Sigma_{jj}}};\rho_{ij}\right)}\exp\left(\frac{1}{2}\Sigma_{ii}\right) & \text{if } D^i=0,\ D^j=0,\\ \dfrac{\Phi\left(\frac{T^i-\Sigma_{ii}}{\sqrt{\Sigma_{ii}}}\right)-\Phi_2\left(\frac{T^i-\Sigma_{ii}}{\sqrt{\Sigma_{ii}}},\frac{T^j-\Sigma_{ij}}{\sqrt{\Sigma_{jj}}};\rho_{ij}\right)}{\Phi\left(\frac{T^i}{\sqrt{\Sigma_{ii}}}\right)-\Phi_2\left(\frac{T^i}{\sqrt{\Sigma_{ii}}},\frac{T^j}{\sqrt{\Sigma_{jj}}};\rho_{ij}\right)}\exp\left(\frac{1}{2}\Sigma_{ii}\right) & \text{if } D^i=0,\ D^j=1,\\ \dfrac{\Phi\left(\frac{T^j-\Sigma_{ij}}{\sqrt{\Sigma_{jj}}}\right)-\Phi_2\left(\frac{T^i-\Sigma_{ii}}{\sqrt{\Sigma_{ii}}},\frac{T^j-\Sigma_{ij}}{\sqrt{\Sigma_{jj}}};\rho_{ij}\right)}{\Phi\left(\frac{T^j}{\sqrt{\Sigma_{jj}}}\right)-\Phi_2\left(\frac{T^i}{\sqrt{\Sigma_{ii}}},\frac{T^j}{\sqrt{\Sigma_{jj}}};\rho_{ij}\right)}\exp\left(\frac{1}{2}\Sigma_{ii}\right) & \text{if } D^i=1,\ D^j=0,\\ \dfrac{1-\left(\Phi\left(\frac{T^i-\Sigma_{ii}}{\sqrt{\Sigma_{ii}}}\right)+\Phi\left(\frac{T^j-\Sigma_{ij}}{\sqrt{\Sigma_{jj}}}\right)-\Phi_2\left(\frac{T^i-\Sigma_{ii}}{\sqrt{\Sigma_{ii}}},\frac{T^j-\Sigma_{ij}}{\sqrt{\Sigma_{jj}}};\rho_{ij}\right)\right)}{1-\left(\Phi\left(\frac{T^i}{\sqrt{\Sigma_{ii}}}\right)+\Phi\left(\frac{T^j}{\sqrt{\Sigma_{jj}}}\right)-\Phi_2\left(\frac{T^i}{\sqrt{\Sigma_{ii}}},\frac{T^j}{\sqrt{\Sigma_{jj}}};\rho_{ij}\right)\right)}\exp\left(\frac{1}{2}\Sigma_{ii}\right) & \text{if } D^i=1,\ D^j=1\end{cases}\tag{3.16}$$

for $i\in\{K,L\}$ and $j\in\{F,\ G,\ K,\ L\}$.

3.13 Appendix: additional equilibrium relationships

In this appendix, we derive additional equilibrium relationships which hold when the production function is Cobb-Douglas - that is, when it has the following functional form as in HK:

$$Y_i = A_i K_i^{\alpha_i^K} L_i^{\alpha_i^L} X_i^{\alpha_i^X}$$
$$\text{with} \quad \alpha_i^K + \alpha_i^L + \alpha_i^X = 1$$

The profit-maximizing output level is given by:

$$Y_i = \left[\frac{\eta - 1}{\eta} \cdot \frac{\mathbb{E}\left(e^{z_i}\right)}{C_i}\right]^\eta$$

while the expected output price respects the well-known markup pricing formula, amended to include the effect of distortions:

$$\mathbb{E}_i\left(P_i\right) = \frac{\eta}{\eta - 1} C_i$$

The factor demand functions are given by:

$$\begin{bmatrix} K_i \\ L_i \\ X_i \end{bmatrix} = \begin{bmatrix} \exp\left(-\bar{\tau}_i^K\right) \alpha_i^K / r_i \\ \exp\left(-\bar{\tau}_i^L\right) \alpha_i^L / w \\ \alpha_i^X / p \end{bmatrix} \frac{C_i^{1-\eta}}{\exp\left(\bar{\tau}_i^Y\right)} \left[\frac{\eta - 1}{\eta} \mathbb{E}\left(e^{z_i}\right)\right]^\eta$$

The retrieval of the distribution of the τ vector allows us to compute the expected wedges $\bar{\tau}_i$ for every firm, which in turn pins down the production function elasticities:

$$\begin{aligned} \alpha_i^K &= \frac{\exp\left(\bar{\tau}_i^K\right) r_i K_i}{\exp\left(\bar{\tau}_i^K\right) r_i K_i + \exp\left(\bar{\tau}_i^L\right) w L_i + p X_i} \\ \alpha_i^L &= \frac{\exp\left(\bar{\tau}_i^L\right) w L_i}{\exp\left(\bar{\tau}_i^K\right) r_i K_i + \exp\left(\bar{\tau}_i^L\right) w L_i + p X_i} \\ \alpha_i^X &= \frac{p X_i}{\exp\left(\bar{\tau}_i^K\right) r_i K_i + \exp\left(\bar{\tau}_i^L\right) w L_i + p X_i} \end{aligned} \tag{3.17}$$

3.14 Appendix: GMM estimates with sector controls

Table 3.10: Demand and firm-level wedges distribution parameters: long-run model

	Parameter	Method	FRA	GER	HUN	ITA	SPA	UK
Demand (substitution) elasticity	η	GMM	29.028***	26.874***	110.077***	29.153***	52.747***	31.476***
			(0.904)	(2.062)	(30.120)	(0.681)	(4.733)	(6.615)
Family control wedges (Y)	σ^F	GMM	0.000	0.006	0.009***	0.000	0.007***	0.000
			(0.001)	(0.004)	(0.003)	(0.000)	(0.002)	(0.000)
Bureaucracy wedges (Y)	σ^G	GMM	0.000	0.000	0.000	0.000	0.000	0.004
			(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.009)
Financial constraints (K)	σ^K	GMM	0.000	0.000	0.007	0.000	0.141***	0.043
			(0.000)	(0.000)	(0.108)	(0.000)	(0.053)	(0.138)
Labor regulation wedges (L)	σ^L	GMM	0.000	0.000	0.000	0.012**	0.000	0.000
			(0.000)	(0.000)	(0.000)	(0.005)	(0.009)	(0.000)
Observations	N	GMM	2,355	637	360	2,384	407	417
Family control wedges (Y)	T^F	MLE	-0.191***	-0.972***	-0.148**	-0.736***	-0.710***	-0.361***
			(0.027)	(0.034)	(0.067)	(0.028)	(0.065)	(0.030)
Bureaucracy wedges (Y)	T^G	MLE	0.113***	1.432***	1.628***	0.622***	1.059***	1.726***
			(0.027)	(0.042)	(0.111)	(0.028)	(0.076)	(0.052)
Financial constraints (K)	T^K	MLE	1.547***	0.899***	1.585***	1.250***	1.027***	1.876***
			(0.041)	(0.035)	(0.108)	(0.036)	(0.070)	(0.057)
Labor regulation wedges (L)	T^L	MLE	0.267***	1.839***	1.585***	0.593***	0.604***	1.728***
			(0.027)	(0.052)	(0.112)	(0.028)	(0.067)	(0.052)
Correlation of (τ^G, τ^L)	ρ^{GL}	MLE	0.706***	0.572***	0.431**	0.522***	0.510***	0.241**
			(0.021)	(0.056)	(0.168)	(0.030)	(0.079)	(0.094)
Observations	N	MLE	2,973	2,935	488	3,020	518	2,067

Table 3.11: Demand and firm-level wedges distribution parameters: short-run model

	Parameter	Method	FRA	GER	HUN	ITA	SPA	UK
Demand (substitution) elasticity	η	GMM	2.528***	2.971***	3.294***	3.295***	2.751***	3.288***
			(0.022)	(0.100)	(0.087)	(0.033)	(0.058)	(0.116)
Family control wedges (Y)	σ^F	GMM	0.003	0.043	0.002	0.001	0.000	0.000
			(0.007)	(0.031)	(0.015)	(0.006)	(0.000)	(0.000)
Bureaucracy wedges (Y)	σ^G	GMM	0.000	0.000	0.000	0.000	0.000	0.000
			(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Labor regulation wedges (Y)	σ^L	GMM	0.018***	0.000	0.003	0.010*	0.016	0.000
			(0.007)	(0.000)	(0.025)	(0.006)	(0.015)	(0.007)
Observations	N	GMM	2,355	637	360	2,384	407	417
Family control wedges (Y)	T^F	MLE	-0.191***	-0.972***	-0.148**	-0.736***	-0.710***	-0.361***
			(0.027)	(0.034)	(0.067)	(0.028)	(0.065)	(0.030)
Bureaucracy wedges (Y)	T^G	MLE	0.113***	1.432***	1.628***	0.622***	1.059***	1.726***
			(0.027)	(0.042)	(0.111)	(0.028)	(0.076)	(0.052)
Labor regulation wedges (Y)	T^L	MLE	0.267***	1.839***	1.585***	0.593***	0.604***	1.728***
			(0.027)	(0.052)	(0.112)	(0.028)	(0.067)	(0.052)
Correlation of (τ^G, τ^L)	ρ^{GL}	MLE	0.706***	0.572***	0.431**	0.522***	0.510***	0.241**
			(0.021)	(0.056)	(0.168)	(0.030)	(0.079)	(0.094)
Observations	N	MLE	2,973	2,935	488	3,020	518	2,067

3.15 Appendix: Maximum likelihood estimates

Table 3.12: Demand and firm-level wedges distribution parameters: long-run model

	Parameter	Method	FRA	GER	HUN	ITA	SPA	UK
Demand (substitution) elasticity	η	MLE	29.169***	25.555***	110.228***	29.098***	50.326***	31.719***
			(0.889)	(2.202)	(29.934)	(0.665)	(4.393)	(6.982)
Family control wedges (Y)	σ^F	MLE	0.000	0.004	0.007**	0.000	0.007***	0.000
			(0.001)	(0.004)	(0.003)	(0.000)	(0.002)	(0.000)
Bureaucracy wedges (Y)	σ^G	MLE	0.000	0.000	0.000	0.000	0.001	0.003
			(0.000)	(0.000)	(0.000)	(0.001)	(0.003)	(0.009)
Financial constraints (K)	σ^K	MLE	0.030	0.001	0.000	0.000	0.109***	0.000
			(0.034)	(0.061)	(0.000)	(0.000)	(0.037)	(0.000)
Labor regulation wedges (L)	σ^L	MLE	0.000	0.000	0.000	0.011	0.000	0.000
			(0.000)	(0.000)	(0.000)	(0.006)	(0.000)	(0.000)
Observations	N	MLE	2,355	637	360	2,384	407	417
Family control wedges (Y)	T^F	MLE	-0.191***	-0.972***	-0.148**	-0.736***	-0.710***	-0.361***
			(0.027)	(0.034)	(0.067)	(0.028)	(0.065)	(0.030)
Bureaucracy wedges (Y)	T^G	MLE	0.113***	1.432***	1.628***	0.622***	1.059***	1.726***
			(0.027)	(0.042)	(0.111)	(0.028)	(0.076)	(0.052)
Financial constraints (K)	T^K	MLE	1.637***	0.902***	1.585***	1.256***	1.070***	1.893***
			(0.045)	(0.035)	(0.108)	(0.036)	(0.071)	(0.058)
Labor regulation wedges (L)	T^L	MLE	0.267***	1.839***	1.585***	0.593***	0.604***	1.728***
			(0.027)	(0.052)	(0.112)	(0.028)	(0.067)	(0.052)
Correlation of (τ^G, τ^L)	ρ^{GL}	MLE	0.706***	0.572***	0.431**	0.522***	0.510***	0.241**
			(0.021)	(0.056)	(0.168)	(0.030)	(0.079)	(0.094)
Observations	N	MLE	2,973	2,935	488	3,020	518	2,067

Table 3.13: Demand and firm-level wedges distribution parameters: short-run model

	Parameter	Method	FRA	GER	HUN	ITA	SPA	UK
Demand (substitution) elasticity	η	MLE	2.517***	2.959***	3.312***	3.290***	2.747***	3.252***
			(0.021)	(0.104)	(0.087)	(0.033)	(0.058)	(0.143)
Family control wedges (Y)	σ^F	MLE	0.004	0.042*	0.009	0.000	0.000	0.000
			(0.007)	(0.025)	(0.014)	(0.000)	(0.000)	(0.000)
Bureaucracy wedges (Y)	σ^G	MLE	0.000	0.000	0.000	0.000	0.000	0.000
			(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Labor regulation wedges (Y)	σ^L	MLE	0.025***	0.000	0.016	0.010*	0.017	0.000
			(0.006)	(0.000)	(0.026)	(0.005)	(0.014)	(0.000)
Observations	N	MLE	2,355	637	360	2,384	407	417
Family control wedges (Y)	T^F	MLE	-0.191***	-0.972***	-0.148**	-0.736***	-0.710***	-0.361***
			(0.027)	(0.034)	(0.067)	(0.028)	(0.065)	(0.030)
Bureaucracy wedges (Y)	T^G	MLE	0.113***	1.432***	1.628***	0.622***	1.059***	1.726***
			(0.027)	(0.042)	(0.111)	(0.028)	(0.076)	(0.052)
Labor regulation wedges (Y)	T^L	MLE	0.267***	1.839***	1.585***	0.593***	0.604***	1.728***
			(0.027)	(0.052)	(0.112)	(0.028)	(0.067)	(0.052)
Correlation of (τ^G, τ^L)	ρ^{GL}	MLE	0.706***	0.572***	0.431**	0.522***	0.510***	0.241**
			(0.021)	(0.056)	(0.168)	(0.030)	(0.079)	(0.094)
Observations	N	MLE	2,973	2,935	488	3,020	518	2,067

3.16 Bibliography

Adamopoulos, Tasso et al. (2017). *Misallocation, selection and productivity: A quantitative analysis with panel data from China*. Tech. rep. National Bureau of Economic Research.

Altomonte, Carlo and Tommaso Aquilante (2012). "The EU-EFIGE/Bruegel-unicredit dataset". *Bruegel Working Paper*.

Asker, John, Allan Collard-Wexler, and Jan De Loecker (2014). "Dynamic inputs and resource (mis) allocation". *Journal of Political Economy* 122.5, pp. 1013–1063.

Banerjee, Abhijit V and Esther Duflo (2005). "Growth theory through the lens of development economics". *Handbook of economic growth* 1, pp. 473–552.

Bartelsman, Eric, John Haltiwanger, and Stefano Scarpetta (2013). "Cross-country differences in productivity: The role of allocation and selection". *The American Economic Review* 103.1, pp. 305–334.

Bennedsen, Morten et al. (2007). "Inside the family firm: The role of families in succession decisions and performance". *The Quarterly Journal of Economics* 122.2, pp. 647–691.

Bertrand, Marianne and Sendhil Mullainathan (2003). "Enjoying the quiet life? Corporate governance and managerial preferences". *Journal of political Economy* 111.5, pp. 1043–1075.

Botero, Juan C et al. (2004). "The regulation of labor". *The Quarterly Journal of Economics* 119.4, pp. 1339–1382.

Buera, Francisco J. and Yongseok Shin (2013). "Financial Frictions and the Persistence of History: A Quantitative Exploration". *Journal of Political Economy* 121.2, pp. 221–272. DOI: 10.1086/670271.

Caselli, Francesco and Nicola Gennaioli (2013). "Dynastic management". *Economic Inquiry* 51.1, pp. 971–996.

David, Joel M., Hugo A. Hopenhayn, and Venky Venkateswaran (2016). "Information, Misallocation and Aggregate Productivity*". *The Quarterly Journal of Economics*. DOI: 10.1093/qje/qjw006.

David, Joel M and Venky Venkateswaran (2017). *The sources of capital misallocation*. Tech. rep. National Bureau of Economic Research.

De Loecker, Jan and Frederic Warzynski (2012). "Markups and firm-level export status". *American economic review* 102.6, pp. 2437–71.

Dias, Daniel A, Carlos Robalo Marques, and Christine Richmond (2016). "A tale of two sectors: why is misallocation higher in services than in manufacturing?"

Djankov, Simeon et al. (2002). "The regulation of entry". *The Quarterly Journal of Economics* 117.1, pp. 1–37.

Edmond, Chris, Virgiliu Midrigan, and Daniel Yi Xu (2018). *How Costly Are Markups?* Tech. rep. National Bureau of Economic Research.

Eisfeldt, Andrea L and Adriano A Rampini (2006). "Capital reallocation and liquidity". *Journal of monetary Economics* 53.3, pp. 369–399.

Farrell, Diana and Susan Lund (2006). "China's and India's financial systems: A barrier to growth". *McKinsey Quarterly*, pp. 1–12.

Garicano, Luis, Claire Lelarge, and John Van Reenen (2016). "Firm size distortions and the productivity distribution: Evidence from France". *The American Economic Review* 106.11, pp. 3439–3479.

Gopinath, Gita et al. (2017). "Capital allocation and productivity in South Europe". *The Quarterly Journal of Economics*, qjx024.

Gourio, François and Nicolas Roys (2014). "Size-dependent regulations, firm size distribution, and reallocation". *Quantitative Economics* 5.2, pp. 377–416.

Haltiwanger, John, Robert Kulick, and Chad Syverson (Jan. 2018). *Misallocation Measures: The Distortion That Ate the Residual*. Working Paper 24199. National Bureau of Economic Research. DOI: 10.3386/w24199.

Harberger, Arnold C (1954). "Monopoly and Resource Allocation". *The American Economic Review*, pp. 77–87.

Holmström, Bengt and Jean Tirole (1998). "Private and public supply of liquidity". *Journal of Political Economy* 106.1, pp. 1–40.

Hopenhayn, Hugo and Richard Rogerson (1993). "Job turnover and policy evaluation: A general equilibrium analysis". *Journal of Political Economy*, pp. 915–938.

Hsieh, Chang-Tai and Peter J. Klenow (2009). “Misallocation and Manufacturing TFP in China and India”. *The Quarterly Journal of Economics* 124.4, pp. 1403–1448. DOI: 10.1162/qjec.2009.124.4.1403.

Kalemli-Ozcan, Sebnem et al. (2015). *How to construct nationally representative firm level data from the ORBIS global database*. Tech. rep. National Bureau of Economic Research.

León-Ledesma, Miguel A and Dimitris Christopoulos (2016). “Misallocation, Access to Finance, and Public Credit: Firm-Level Evidence”. *Asian Development Review* 33.2, pp. 119–143.

Levine, Ross (2005). “Finance and growth: theory and evidence”. *Handbook of economic growth* 1, pp. 865–934.

Levinsohn, James and Amil Petrin (2003). “Estimating production functions using inputs to control for unobservables”. *The review of economic studies* 70.2, pp. 317–341.

Midrigan, Virgiliu and Daniel Yi Xu (2014). “Finance and misallocation: Evidence from plant-level data”. *The American Economic Review* 104.2, pp. 422–458.

O’Mahony, Mary and Marcel P Timmer (2009). “Output, input and productivity measures at the industry level: the EU KLEMS database”. *The Economic Journal* 119.538.

Olley, G. Steven and Ariel Pakes (1996). “The Dynamics of Productivity in the Telecommunications Equipment Industry”. *Econometrica* 64.6, pp. 1263–1297. ISSN: 00129682, 14680262.

Pérez-González, Francisco (Dec. 2006). “Inherited Control and Firm Performance”. *American Economic Review* 96.5, pp. 1559–1588. DOI: 10.1257/aer.96.5.1559.

Restuccia, Diego and Richard Rogerson (2008). “Policy distortions and aggregate productivity with heterogeneous establishments”. *Review of Economic dynamics* 11.4, pp. 707–720.

Stiglitz, Joseph E and Andrew Weiss (1981). “Credit rationing in markets with imperfect information”. *The American economic review* 71.3, pp. 393–410.

Tsoutsoura, Margarita (2015). “The effect of succession taxes on family firm investment: Evidence from a natural experiment”. *The Journal of Finance* 70.2, pp. 649–688.

Villalonga, Belen and Raphael Amit (2006). “How do family ownership, control and management affect firm value?” *Journal of Financial Economics* 80.2, pp. 385–417.

Whited, Toni M and Jake Zhao (2016). “The misallocation of finance”.