BOSTON UNIVERSITY

GRADUATE SCHOOL OF ARTS AND SCIENCES

Dissertation

INNOVATION AND PRODUCTIVITY ANALYSIS WITH HETEROGENEOUS FIRMS

by

SHUHENG LIN

B.S.Comm.(Hons.), DePaul University, 2009
M.A., Boston University, 2012

Submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

2016

Approved by

First Reader ______________________________

Stefania Garetto, Ph.D.
Assistant Professor of Economics

Second Reader ______________________________

Jordi Jaumandreu, Ph.D.
Senior Academic Researcher

Third Reader ______________________________

Marc Rysman, Ph.D.
Professor of Economics

Acknowledgments

The teachers throughout my studies are dedicated and passionate about what they do and have made school intrinsically fun. As a result I have always enjoyed learning, so much so that when it came to the transition from consumption to production of research I struggled immensely. The completion of my Ph.D. would not have been possible without the relentless support from my dissertation committee, friends and family. No words of gratitude will ever be enough and I could only hope to pay forward their help in my career and life ahead.

Stefania Garetto has been giving me incredible help since my third year in the program. Meetings with Stefania always recentered me with the bigger picture and the underlying economics. Her ability to strip down a complicated problem for its core and to guide students through convoluted questions for a clear message will be something I aspire to emulate. What I have learned from Stefania extends beyond research, I could not have asked for a better female mentor.

I started working with Jordi Jaumandreu later in my dissertation phase but his advice and support have been instrumental. The methdological way he approaches a problem has kept me grounded when things get difficult. When discussing findings, Jordi never takes results for granted and always makes sure the underlying mechanism makes sense. His dedication and attitude towards research are infectious, and his selfless help has made the completion of this program possible for me.

Marc Rysman was brought onboard even later but his help was indispensible; Marc's sharp wit coupled with his rich tool box for solving problems have made an important impact. Marc is also one of the best teachers I have taken classes from; his lectures are so much fun and they have guided me towards a better understanding of economics and research in general.

Apart from my incredible dissertation committee, I have also benefited tremendously from the following professors' help: Claudia Olivetti, Kevin Lang, Andrew Newman, Sam Bazzi, Zhongjun Qu, Hiro Kaido, Luis Carvalho and Bart Lipman. Kadin Tseng and Katia Oleinik from the technology help desk at Boston University have provided great support on cluster-related problems.

The process of intellectual and personal growth during my study at Boston University is not whole without the confidants that I have gotten to know. Amrit Amirapu, Aparna Dutta, Anusha Nath, Francisca Perez, Marric Buessing, Fan Zhuo, Luis Huerta, Huailu Li, Yasemin Satir, Tanmay Shukla, Elisabeth Perlman, Nathaniel Young, Gene Schlack and Nathaniel Blalock have seen me through the thick of it all. Not a day passes by that I am not thankful for having these great friends in my life.

Last but not least, I wish to thank my family that has given up an immense amount so I could pursue interests of my own, as well as our family friends in Shanghai and Chicago who took me in as one of their own during my times there: I am nothing without them.

INNOVATION AND PRODUCTIVITY ANALYSIS WITH HETEROGENEOUS FIRMS

SHUHENG LIN

Boston University, Graduate School of Arts and Sciences, 2016

Major Professor: Stefania Garetto, Assistant Professor of Economics

ABSTRACT

This dissertation examines the relationship between productivity growth and research activities of heterogeneous firms, and the contribution of firm heterogeneity to business cycle fluctuations.

The first chapter uses a dynamic model to study firms' decisions on whether to conduct research in house, with external units or via both modes. Productivity is modeled to evolve endogenously according to Research and Development (R&D) modes, and the costs of starting and continuing research are random and mode specific. Model estimates from a panel of Chinese manufacturing firms show that in-house R&D is more effective and costs less to maintain, but smaller firms choose external R&D because of its lower startup cost. These estimates are consistent with the observed cross-sectional differences in firm size by research status, and can match the persistence and transition dynamics in R&D modes. Simulation exercises show that continuation cost reduction induces more changes in R&D decisions, but start up cost reduction leads to most of the aggregate productivity gain.

The second chapter investigates the impact of innovation on firm level prices. This impact depends on how innovation affects quality and efficiency and how the firm passes these changes onto prices. Estimation results of the empirical model with a panel of Spanish firms show that firms take advantage of process innovations to enlarge markups by not completely passing onto prices the decrease in cost. Product innovations could increase or decrease cost but they do not affect markups, thus we do not find prices to change systematically with them.

The third chapter examines the contribution of firm level shocks to output fluctuations for four OECD countries (US, Germany, Canada and the UK). Recent studies stemming from Gabaix

(2011) show that when few firms account for a disproportionately large share of production, shocks to these firms can propagate to generate business cycle fluctuations. However, we find that while firm size distribution is highly skewed in these four economies, the ability of the largest firms to transmit shocks is not universal and thus should not be taken for granted.

Contents

List of Tables

List of Figures

Chapter 1

The Dynamics of R&D Organization and Productivity Growth

1.1 Introduction

The organization of research and development (R&D) activities differs across firms and varies over time. Some firms choose to perform R&D in house, some contract with other firms or research institutions, and some use a combination of the two modes. While the existing studies have examined extensively the performance consequences of R&D organization, few consider what drives the variations in organization decisions for research and why a firm's chosen mode is not invariant over time.[1] Since R&D investments have uncertain returns but when successful they increase productivity, a better understanding of how firms organize research is important for assessing the implications of R&D for growth and welfare.

In this paper, I use a dynamic structural model to show that selections into R&D modes are driven by variation in startup and continuation costs and are further reinforced by mode-specific returns. Which type of cost is relevant depends on a firm's previous period research status; the respective costs are also random and the associated uncertainty dictates the transitions in and out of R&D modes. I then estimate this model with a panel of Chinese manufacturing firms to quantify the costs and benefits of the different research modes. With the model estimates, I analyze how productivity and cost uncertainty affect firm dynamics in R&D organization, and in turn, what the implications are for firm-level and aggregate productivity growth when cost uncertainty is reduced.

[1]There are a few notable exceptions. Aghion and Tirole (1994) use an incomplete contract framework to model the static management of R&D. Aw et al. (2011) and Peters et al. (2013) also use a dynamic structural model to investigate the costs of overall R&D activity, without considering how it is organized.

The task of quantifying research costs calls for a dynamic structural model because they are often unobserved, vary and paid across time periods. Firms not active in R&D face the cost of setting up research labs and recruiting talent in the case of internal R&D or the transaction costs associated with identifying the appropriate research unit to partner with in the case of external R&D. After these irreversible startup costs are paid, firms incur a per-period cost to remain active in research. These maintenance costs not only entail explicit expenditures such as payments to R&D personnel and capital, but also unobserved heterogeneous components that manifest in the efficiency of the contracting environment, the availability of credit, R&D services and talent.

To account for the complex nature of research costs, I follow the approach in Aw et al. (2011) and Peters et al. (2013) to model them as random draws observed to the firm, but not to the econometrician. However, instead of treating R&D as a single activity, I let research (startup and continuation) costs and benefits differ by modes in order to examine organization dynamics and the associated implications for productivity growth.

The model mechanism is straightforward. Each period, firms consider the status-dependent cost draws associated with the different R&D modes in hand, and choose the R&D mode that gives the highest expected payoffs by solving a simultaneous choice problem. This expectation is formed over future uncertainty in research cost and the returns on R&D organization; current internal and external R&D decisions are endogenous and they have different growth effects because productivity (a combination of unobserved technical efficiency and product quality) is an endogenous Markov process of R&D organization. The parameters in the returns and the cost distributions are then estimated from the data such that they can rationalize the observed R&D decisions and dynamics.

The data used for estimating the model are a three year panel of Chinese firms from the 2003 Enterprise Survey conducted jointly by the World Bank and the Chinese National Bureau of Statistics. Some empirical regularities arise from this panel of Chinese firms. Firms that do not perform R&D have the smallest sales, and among R&D performers firms that conduct research both in house and with external units are the largest, followed by internal only, and lastly by external R&D only performers. The transition matrix in R&D structure shows firms are persistent in R&D modes, but more so in internal than in external. Finally, firms not active in research are more likely to start with external R&D and firms with in-house R&D switch in and out of external activities.

The model estimates are consistent with the aforementioned empirical observations and provide insights into the trade-offs faced by firms. In house R&D is more effective, cheaper to maintain, but costs more to start than external R&D. This delivers selection into R&D structure that reinforces the cross-sectional differences in firms by research status observed in the data. The combination of the large startup but smaller continuation cost estimated for in house R&D rationalizes the observation where firms with internal knowledge capacity rarely quit research. The lower estimate for external startup cost and higher external continuation cost relative to that of internal R&D are in line with the more frequently observed entries and exits in external R&D activities. It is also consistent with the hypothesis that it requires less commitment to outsource research. Since contractual R&D does not commit firms to engineers and capital, internal R&D performers additionally switch in and out of external R&D. The reason could be the firm requires additional skills for one-off projects or it wants to take advantage of off-the-shelf technology from external sources.

The estimated model provides a laboratory for examining what happens to firm dynamics in R&D organization and productivity growth if research costs were reduced. I find that reduction in startup cost would induce entry but also quitting because of a reduction in the option value of investing. Reduction in continuation costs would increase persistence in research and induce entry. In either type of cost reduction, adjustments along external R&D dimension account for almost 70% of the productivity gain and they also affect a larger share of firms. This is in line with the findings in Agrawal et al. (2014), where they find that under a tax incentive program for firms in Canada, after-tax cost elasticity is higher for contractual R&D than R&D wage bills for in-house personnel.

Although continuation cost reduction induces more changes in R&D organization via entry and increases persistence, startup cost reduction leads to a higher aggregate productivity gain. For this sample of Chinese firms, there are still a sizable proportion of firms inactive in research and the startup cost is much more substantial than continuation cost. As a result, the quitting channel does not operate pervasively to erode productivity gain. While continuation cost reduction induces entry, it also encourages firms less productive in research to continue. These results from this simulation exercise suggest that growth implications of R&D cost reduction are context specific. They are determined by the degree of cost uncertainty associated with different research modes and the initial distribution of firms' research status and productivity. Failing to take these factors into account jointly could mislead R&D policies.

Lastly, to assess the importance of R&D organization, I compare the model with one where R&D is a single activity. When R&D organization is not considered, the startup cost is twice the external startup cost and only slightly lower than that of in-house R&D. This precludes some medium size firms from investing in research and some large firms from contracting with external units for temporary research needs. As a result, cost reductions lead to productivity growth that is almost 13% less when firms cannot choose how to organize R&D.

The rest of this paper is structured as follows. Section 1.2 discusses this paper's contribution to the relevant literature. Section 3.2 describes the data and provides descriptive statistics that motivate the dynamic model, which is then laid out in Section 1.4. The estimation strategy of the model is outlined in Section 1.5. The estimation results are summarized in Section 1.6. Using the estimates, I conduct R&D cost reduction simulation exercises in Section 1.7. Finally, Section 1.8 concludes.

1.2 Literature Review

This paper is most closely related to studies that examine the costs of R&D: one strand of which finds evidence of R&D sunk and fixed cost while the other estimates them as this paper does. By examining the lagged structure of R&D and export decisions, Aw et al. (2007) find that R&D activities are much less persistent than export for Taiwanese firms and differences in sunk cost is likely the explanation. Following a tax incentive scheme in Canada, Agrawal et al. (2014) find firms' differential responses in R&D wage bills versus contractual expenditures as evidence for the existence of R&D fixed costs. Aw et al. (2011) and Peters et al. (2013) estimate the actual level of R&D costs for Taiwanese and German firms. The former paper estimates a dynamic structural model of the firm's decision to export and conduct R&D, and the latter of the firm's decision to conduct R&D taking into account innovation outcomes. Both Aw et al. (2011) and Peters et al. (2013) find R&D costs to be significant, and that sunk cost is significantly more than fixed cost. This paper takes after these two studies in modeling the trade-offs faced by a firm when making R&D decisions, but allows research to be conducted in house and/or outsourced in order to study organization dynamics.[2]

This paper contributes to the literature that empirically investigates the effectiveness of internal versus external R&D and whether they are substitutes or complements. These studies often find internal R&D to be more effective than external R&D.[3] However, there is no consensus on whether R&D modes are complements or substitutes, e.g. whether internal R&D increases a firm's "absorptive capacity" to learn from external units.[4] Lokshin et al. (2008) find complementarity to exist only when firms have invested enough in internal R&D. Using a panel of Swedish firms, Bergman (2011) also finds internal and contractual R&D are complements only for firms with high R&D intensities. The conclusion in Audretsch et al. (1996) is similar in that they find complementarity (substitutability) between the two modes for high (low) tech industries. Cassiman and Veugelers (2006) find the degree of complementarity depends on how reliant firms are on universities and research centers. Breaking up external R&D to contractual versus collaborative, Schmiedeberg (2008) finds complementarity between in house and cooperative R&D but not contractual R&D.[5]

The complementarity of R&D modes remains context-specific, and even within a given context the effect is heterogeneous. The way R&D costs are modeled in the paper could potentially reconcile these conflicting findings. While I do not find significant interaction of internal and external R&D in the context of China (lack of complementarity), the realized random cost draws could put a firm in either situation. And the reason for which could be due to either correlations in the cost

[2]Another closely related study is Amoroso (2014); she also estimates the sunk and fixed costs of research, but the aim is to understand how these costs are different when firms also cooperate with external units. She finds sunk cost of R&D to be higher than that of innovation, and these costs are lower when firms collaborate with external units.

[3]Bonte (2003) is an exception where he finds external R&D to have a higher productivity effect than internal R&D using a panel of West German firms.

[4]Cohen and Levinthal (1989) is one of the first studies to use the term "absorptive capacity" to describe that tacit knowledge is more easily acquired if the firm also engages actively in research pertaining to relevant fields.

[5]For examples of studies that define technology sourcing as external R&D, see Hou and Mohnen (2013) and Hou (2013). Since the objective of this paper is to examine the firm's deliberate effort to conduct research, I do not consider additionally the role of licensing.

shocks or economies of scope.

This paper is also related to the literature that examine the determinants of R&D structure, one of which concerns the differences in innovation needs. Beneito (2006) finds in a panel of Spanish firms that internal R&D is most productive for significant innovation while contracted R&D is oriented more towards innovation of incremental nature. On the other hand, industry leaders conclude that focused innovations often take place internally and serendipitous ones externally.[6] In an incomplete contract framework, Aghion and Tirole (1994) predicts that research is more likely to be conducted in house if effort or capital inputs from the firm are more important for innovation success than those from the research unit. This paper takes an agnostic approach when mapping innovation needs and R&D modes; the motive for research is to enhance profit position, and firms use different R&D modes to achieve process and product innovation since productivity is modeled as a combination of technical efficiency and product quality.[7]

Another determinant of R&D modes is a firm's internal financing capability, and the theoretical model in Aghion and Tirole (1994) predicts that firms with "deep pockets" (or larger firms) are more likely to integrate research in house.[8] This paper also documents the observed selection of larger firms into doing research in house. However, it does not model the role of credit constraints; firms are capable of paying the lump sum set up and continuation costs as long as the marginal benefit exceeds the cost.

1.3 Empirical Evidence

This section first describes the data and then presents descriptive statistics that motivate the dynamic model. There exists a pecking order in firm heterogeneity with respect to research modes; none-R&D performers are the smallest and least productive, followed by firms that only conduct research with external units, then firms with in-house research teams, and lastly firms that do both

[6] Views of industry leaders are summarized from the 2008 Centennial Global Business Summit that took place in the Harvard Business School.

[7] See Cohen (2010) for a more comprehensive overview for determinants of overall R&D expenditure level.

[8] There is also a related strand of studies that investigates the role of credit constraint on the overall level of R&D investments, rather than on R&D modes. See Stiglitz and Weiss (1981) for the theoretical backdrop, Himmelberg and Petersen (1994) and Hottenrott and Peters (2012) for empirical evidence in the US and Germany.

are the largest. Firms are highly persistent in their research modes, but entry and exits in external R&D is much more frequently observed.

1.3.1 Data and Variable Description

The empirical analyses in this paper are based on the 2003 Enterprise Survey conducted jointly by the World Bank and the Chinese National Bureau of Statistics. The sample consists of manufacturing establishments representative of 9 sectors, located across 14 provinces and 1 municipality in China.[9] 27% of all establishments reported to be a member of a group company, but "each made its own financial decisions and had its own financial statements separate from those of the firm. It also had its own management and control over its payroll."[10] The rest of this paper uses establishments and firms interchangeably, and treats them as independent decision-makers.

Firms were surveyed in 2003 for extensive information on production, business and investment climate over the fiscal year ending in 2002, and 2 years of retrospective data were recorded for a subset of questions. In the end a three year panel for 1266 firms spanning the period 2000-2002 was constructed for the following variables: sales, capital, material inputs, and R&D activities. A firm is defined as an internal R&D performer if it had its own research team in house with R&D personnel, and an external R&D performer if it had contractual relationships with other firms, universities or research institutions.[11]

Apart from production and R&D variables that are used to estimate the model, I also take variables of innovation outcome, firms' research and organization environment from the cross-sectional part of the extensive questionnaire to examine how these variables correlate with R&D modes. This part of the empirical exercise is meant to complement the model by shedding light on

[9] The 9 sectors are food processing, biotech products, garment & leather, household electronics, metallurgical production, electronic equipment, transportation parts, chemical products and electronic parts. Their geographic locations were chosen to reflect where the bulk of manufacturing activities took place. China has 22 provinces, 5 autonomous regions, 4 municipalities and 2 special administrative regions.

[10] Requirement taken from the questionnaire manual. 34% of the establishments that were members of a group company reported to benefit from R&D programs of other member companies, which makes up about 9% of the total number of firms in the sample.

[11] Whether the nature of contractual relationship was "cooperative" is not known. In a separate section of the survey, firms additionally reported R&D expenditures on labor, capital and technology sourcing. However, there is no clear mapping on how to split these categories further into expenditures on in house versus external activities.

which R&D mode is associated with more innovation success and thus higher returns, and what the unobserved research costs in the model could represent.

1.3.2 Firm Level Characteristics by R&D Modes

Table 1.1 reports average firm level characteristics by the four mutually exclusive R&D modes (NoRD, ExtRD only, IntRD only and Both) to examine how firms differ across categories. Panel A reports basic measures of firm heterogeneities, Panel B summarizes innovation performance, and Panel C presents organization and research environment variables.

The observation counts in the last row show that 45.4% of the 1266 firms do not perform R&D, 8.6% conduct R&D with external units only, 26.6% in house only and the rest of the firms do both (19.3%). The distribution of firm heterogeneity as measured by average sales, number of employees, capital, capital intensity and productivity show firms that conduct R&D both in house and with external units are the largest, followed by internal R&D only and then external R&D only performers. The hierarchical order is robust to industry and location controls, the results for which are summarized in Table 1.10 of Appendix 1.9.5.

The pecking order suggests selection into R&D modes by firm heterogeneity and/or a mirrored pattern of returns to R&D modes. However, performance difference across R&D modes is not trivial to assess, since R&D organization decisions are endogenous to firm specific capabilities, economic and organization environment. Firms that find internal R&D to be most effective are more likely to conduct research in house, and firms that find arm's length R&D to suit their needs will choose accordingly. A simple comparison of the resulting innovation outcomes may mislead one about the effectiveness across R&D modes.

Panel B of Table 1.1 reports innovation outcomes that firms reported as dummy variables, with the exception of number of patents obtained. This limits one from comparing qualitative differences in research output across firms by R&D modes. Nonetheless, Panel B shows that the pecking order by R&D mode is also observed (albeit much less stark) in number of patents obtained and whether a firm introduced new product, new business line and new process. However, the difference between external only and internal only R&D mode is not statistically different from each other. Though it is very clear that firms that do both types of R&D unambiguously produce the most research output. Lastly, firms defined as none performers in this paper also have some research output.[12]

While the existing literature has extensively studied the determinants of overall R&D expenditure and innovation success, few explore the impact of organization and research environments on the chosen R&D modes. Panel C attempts to shed light on these variables by reporting summary statistics on whether a firm had a manager with a post graduate degree, whether it had a line of credit at the bank, the percentage of firms that are stateowned, whether the firm was located in a special economic zone and the whether R&D services were available. While these variables are statistically significant in explaining R&D organization, there is not a clear pattern on which of them are associated with a specific R&D mode.

Table 1.1: Average Firm Level Characteristics by R&D Modes

	NoRD	ExtRD	IntRD	Both
A. Firm Size and Productivity				
Sales (mil. RMB)	57.3	60.6	170.6	308.3
Number of employees	296	380	531	990
Capital (mil. RMB)	25.4	37.2	49.9	167.1
Capital labor ratio (mil. RMB/employee)	21.7	29.8	41.9	58.8
Levinsohn-Petrin productivity	2.06	2.17	2.34	2.40
B. Innovation Performance				
Introduced new product (%)	28.1	60.2	61.9	86.2
Introduced new business line (%)	13.5	30.2	31.4	46.6
Introduced new process (%)	27.1	55.9	51.3	75
Number of patents obtained	0.063	0.401	0.462	1.121
C. Organization and Research Environment				
GM had post-graduate degree	4.8	12.2	16.3	25.5
Had line of credit at bank	20.8	33.2	36.4	44.7
Stateowned	16	24.2	19.5	22.5
Located in special zone	19.7	28.8	35.2	45.4
Availability of R&D services	39.6	67.3	63.8	79.7
N (%)	575 (45.4)	109 (8.6)	337 (26.6)	245 (19.3)

All figures are averages computed from 2000-2002. 1USD≈8RMB in 2002.

[12] Peters et al. (2013) also find that R&D expenditure is not necessary for introducing new products or processes.

1.3.3 Transitions in R&D Modes

In addition to differences in firm performance by R&D modes, it is also observed in the data that firms did not always conduct research via a fix mode. Table 1.2 shows the transition matrix in R&D structure and provides some insights on how differences in costs across R&D mode could rationalize the observed pattern.

The diagonal of the matrix reports the percentage of firms remaining in the same status from $t-1$ to t, and it suggests R&D organization structure is highly persistent (the probability of staying put ranges between 0.882 and 0.928). The first row shows that in this particular sample, non R&D performers are more likely to start with external than in-house R&D in the next period (0.046 versus 0.021), and firms rarely transition to doing both types of R&D. The last two rows show that once an internal research team is put in place, the firm almost never quits research and additionally switches in and out of external R&D (11.4% of the internal R&D only performers take on external R&D additionally the next period, while 8.2% of the firms that perform both drop off their external R&D activities).

The persistence in R&D structure is indicative of hysteresis behavior that can be induced by sunk set up costs and relatively lower continuation cost required to maintain research. The higher frequencies of transitioning in and out of external compared with internal R&D suggest the former may be associated with a lower start up but higher continuation cost. This observation is consistent with the three potential roles external R&D plays. Firms with no prior R&D experience may start with external R&D first before committing to research via in-house R&D, or to take advantage of off-the-shelf technology for short-term performance. Internal R&D performers may additionally take on external R&D for one-off projects in order to avoid building up excess capacity, or to complement in-house capacity in the case of cooperative research.[13]

[13]It is beyond the scope of this paper to tease out the exact reasons behind internal R&D performers branching out to external R&D.

Table 1.2: Transition Matrix in R&D Modes

	noRD_t	extRD_t	intRD_t	both_t
noRD_{t-1}	0.928	0.046	0.021	0.005
extRD_{t-1}	0.071	0.888	0.005	0.036
intRD_{t-1}	0.004	0	0.882	0.114
both_{t-1}	0.002	0	0.082	0.915

All figures are averages computed from 2000-2002.

1.4 Model

The model developed in this section provides a framework for how heterogeneous firms in productivity and size make R&D investment decisions. A firm's problem each period is divided into two components: its static pricing and production decisions determine short-run profit, and its dynamic choices of R&D modes enter to affect future productivity and profitability. Rather than deciding on doing R&D or not, this model differs from its predecessors in that firms additionally consider whether to conduct research in house, with external units or via a combination thereof. Facing cost uncertainties associated with the different R&D modes, firms then choose the research organization that gives the highest expected future payoffs.

1.4.1 Firms' Static Production Decisions

Firms compete in a monopolistic competitive market. There are not strategic interactions among producers, and they each take demand for its variety j at time t, Q_{jt}, as given:

$$Q_{jt} = \Phi_t P_{jt}^{-\eta} \exp(\zeta_{jt}), \;\; \Phi_t \equiv Q_t P_t^{\eta-1}. \tag{1.1}$$

This demand is a function of the market demand index (Φ_t) consisting of the CES aggregate of differentiated varieties and the price index (P_t), firm-level demand shifters (ζ_{jt}), the elasticity of substitution across varieties that is assumed to be constant and greater than 1 (η), and the price set by the firm (P_{jt}).

The profit-maximizing price that satisfies demand follows from the firm's FOC, and it is a constant markup over its short-run marginal cost:

$$P_{jt} = \frac{\eta}{\eta - 1} MC_{jt}. \tag{1.2}$$

Firm j with Hicks-neutral technology (ψ_{jt}) produces with capital (K_{jt}) and variable inputs of labor and materials, and the production function is assumed to be constant-returns-to-scale in these variable factors. In turn, short-run marginal cost (MC_{jt}) depends on (ψ_{jt}, K_{jt}) and common input prices (X_t), whose exact expression depends on the functional form assumption on the production function:[14]

$$MC_{jt} = MC(K_{jt}, X_t)\exp(-\psi_{jt}). \tag{1.3}$$

Combining the equilibrium price and quantity yields the following per-period revenue and profit (in logs):[15]

$$r_{jt} = \ln \Phi_t - (\eta - 1)\ln\frac{\eta}{\eta - 1} - (\eta - 1)mc(k_{jt}, X_t) + (\eta - 1)\underbrace{\left(\psi_{jt} + \frac{1}{\eta - 1}\zeta_{jt}\right)}_{\omega_{jt}}, \tag{1.4}$$

$$\pi_{jt} = -\ln\eta + r_{jt}, \tag{1.5}$$

where ω_{jt} is the firm-specific productivity that is a combination of technology (ψ_{jt}) and demand shifter (ζ_{jt}) that can reflect product quality as in Peters et al. (2013). The technology component captures heterogeneities in revenue across firms relative to inputs, or a firm's ability to produce at a different unit cost from others. The quality component represents demand (dis)advantages that could be explained by product differentiation. Both components of ω_{jt} are known to the firm but unobserved to the econometrician.

In the productivity literature, ω_{jt} is often modeled as an exogenous Markov process, with lagged productivity ω_{jt-1} to capture the persistence in a firm's performance. In the innovation

[14] See Appendix 1.9.1 for its derivations.

[15] Equation 1.5 is derived by subtracting total variable cost (TVC $= MC_{jt}Q_{jt}$) from revenue, and simplified using (1.2) to arrive at: $\Pi_{jt} = (1 - \frac{\eta-1}{\eta})Q_{jt}P_{jt} \equiv \frac{1}{\eta}R_{jt}$.

literature, there is a long tradition of modeling R&D investments as an additional input to the production function (starting with Griliches, 1979). The object of empirical interest is then the marginal returns of "knowledge capital stock" constructed from R&D expenditures. Doraszelski and Jaumandreu (2013) endogenize the productivity process to depend on R&D efforts in a flexible manner, later adapted in Aw et al. (2011), Peters et al. (2013), Bøler et al. (2015) and now also in this paper.

Specifically, productivity evolves according to R&D organization decisions in a first-order Markov process:

$$\begin{aligned} \omega_{jt} &= g(\omega_{jt-1}, i_{jt-1}, e_{jt-1}) + \xi_{jt} \\ &= \alpha_0 + \alpha_1 \omega_{jt-1} + \alpha_e e_{jt-1} + \alpha_i i_{jt-1} + \alpha_{ie} i_{jt-1} * e_{jt-1} + \xi_{jt}. \end{aligned} \tag{1.6}$$

ω_{jt} shifts according to firm's R&D organization decisions in the previous period for whether it conducted research in house (i_{jt-1}) and/or externally (e_{jt-1}). ξ_{jt} is the idiosyncratic shock that captures uncertainty that is inherent in R&D investments, in addition to any surprises unknown to he firm at the time of production. Thus the transition mapping of productivity growth is probabilistic rather than deterministic. Finally, R&D decisions have permanent effects on performance via ω_{jt-1}. How much of the benefits persist over time depends on the coefficient on lagged productivity, relative contribution of the different R&D modes and the amount of uncertainty represented by ξ_{jt}.

As also apparent in (1.6), both efficiency and profitability can be shaped by R&D activities. This paper takes an agnostic view on whether firms use internal or external R&D for improving marginal cost or product attractiveness, for incremental or drastic innovations. The reasons are twofolds. Firstly, information on innovation outcome is only available as categorical variable over the three years in the panel. Secondly, even with continuous variables recorded at the right frequency, it is an econometric challenge to causally identify whether internal or external R&D is more productive for which kind of innovation for reasons outlined in Section 1.3.2.

In the end, the key static parameters of interest for estimation are those governing the returns from research modes in the productivity process. These parameters are common knowledge and the same across firms. However, the resulting impact of research each period is size (K_{jt}) and shock (ξ_{jt}) dependent; since productivity is Hicks neutral, a given percentage increase in ω_{jt} translates into a higher level of profit for larger firms. While R&D decisions are endogenous, I do not model the choice on capital in this paper and take them as given.

1.4.2 Firms' Dynamic Problem

With the law of motion governing productivity in eq (1.6), firms can form expectations of their future profits for a given R&D mode, and the actual research choice depends also on R&D costs that must be incurred to realize the expected productivity gain. Firms wishing to start R&D or restart after inactivity pay a lump-sum set up cost to form a research lab with equipment and recruited technical personnel, or the transaction cost to locate and then form a partnership with an external unit. Once a research lab is established or a research partner is identified, the costs to maintain research in the form of capital and wage expenditures continue to be different each period. These costs can also have unobserved components that represent variations in the efficiency of the contract and organization environment, availability of credit or external R&D services.

The heterogeneities in the aforementioned R&D costs are difficult to measure and often unobserved, and may also vary over time. I follow Aw et al. (2011) and Peters et al. (2013) to account for them in the model as independent draws from a joint CDF $C^{\mathcal{F}}$, where $\mathcal{F} = (F^i, F^e, f^i, f^e)$ is the set of the four random cost variables and $(F^i_{jt}, F^e_{jt}, f^i_{jt}, f^e_{jt})$ are the realized internal and external R&D startup (upper case F) and fixed costs (lower case f).[16]

Each period, firms compare the expected present discounted value of profits from the possible R&D choices with the realization of the cost draws to choose a research mode that gives the highest pay off. The corresponding Bellman equation with the vector of observed state variables $s_{jt} = \left(k_{jt}, \omega_{jt}, i_{jt-1}, e_{jt-1}\right)$ and the vector of unobserved state variables $\mathcal{F}_{jt}$ is as follows:

[16]Internal and external cost draws are independent of firm characteristics and of each other. One can alternatively model the costs as fix parameters and each R&D mode is associated with a stochastic shock and model these shocks to be correlated. This specification and preliminary results are outlined in Appendix 1.9.4.

$$V(s_{jt},\mathcal{F}_{jt}) = \Pi(s_{jt}) + \max_{i_{jt},e_{jt}}\{-i_{jt}C^i_{jt} - e_{jt}C^e_{jt} + \delta \sum_{\omega_{jt+1}} \int_{\mathcal{F}_{jt+1}} V(s_{jt+1},\mathcal{F}_{jt+1})dC^{\mathcal{F}}(\mathcal{F})dP(\omega_{jt+1}|\omega_{jt}, i_{jt}, e_{jt})\}, \tag{1.7}$$

where δ is the discount factor, $dC^{\mathcal{F}}$ is the aforementioned distribution of costs, and $dP(\omega_{jt+1}|\omega_{jt}, i_{jt}, e_{jt})$ is the Markov transition probability distribution that summarizes firms' beliefs about future states governed by eq (1.6).[17] Lastly, C^i_{jt}, C^e_{jt} represent the costs associated with internal and external R&D choices that depend on lagged decisions (i_{jt-1}, e_{jt-1}).

Firm j pays a continuation cost if it was active in research the previous period, and a set-up cost if it was not for each of the two R&D choices:

$$C^i_{jt} = (1 - i_{jt-1})F^i_{jt} + i_{jt-1} f^i_{jt},$$
$$C^e_{jt} = (1 - e_{jt-1})F^e_{jt} + e_{jt-1} f^e_{jt}.$$

The cost expressions indicate that stopping and starting R&D is costly because a firm will have to pay the start up cost again that is irreversible. The requirement of a startup cost induces hysteresis behavior; in face of a bad continuation draw (higher cost for the same percentage improvement in productivity), a firm may not quit research right away because future costs may be lower and by maintaining research through bad times it avoids paying the startup cost again. In addition to choosing whether to start or stop research, firms also need to consider the trade-offs among the R&D modes by comparing how the startup and fixed costs differ across them, in conjunction with how they contribute differentially to productivity gain via the transition matrix of productivity $dP(\omega_{jt+1}|\cdot)$.

With the choices of internal and external mode, a firm in the end faces a menu of four research statuses to choose from: none (n), external R&D only (e), internal R&D only (i), and both internal and external (ie). The actual realized cost draws that rationalize the firms' decisions are observed

[17] It is implied from the expression for $dP(\cdot)$ that the transition probability of is independent of the state variable k_{jt}.

by the firm but not the econometrician, thus the dynamic parameters of interest for estimations are those that characterize the distribution of costs.

Integrating the Bellman equation over these unobserved shocks, and re-writing it with the short-hands for choice-specific continuation values gives the following expression:

$$\bar{V}(s_{jt}) = \int \left(\Pi(s_{jt}) + \max_{i_{jt}, e_{jt}} \left\{ \begin{matrix} \delta E\bar{V}^n, \\ -C^e_{jt} + \delta E\bar{V}^e, \\ -C^i_{jt} + \delta E\bar{V}^i, \\ -C^i_{jt} - C^e_{jt} + \delta E\bar{V}^{ie} \end{matrix} \right\} \right) dC^{\mathcal{F}}(\mathcal{F}), \tag{1.8}$$

where $\bar{V}(s_{jt}) = \int V(s_{jt}, \mathcal{F}_{jt}) dC^{\mathcal{F}}(\mathcal{F})$ is the integrated value function, and EV^n, EV^e, EV^i, EV^{ie} are the choice-specific value functions corresponding to the four possible R&D statuses:

$$E\bar{V}^n \equiv \sum_{\omega_{jt+1}} \bar{V}(s_{jt+1}) dP(\omega_{jt+1} | \omega_{jt}, i_{jt} = 0, e_{jt} = 0), \tag{1.9}$$

$$E\bar{V}^e \equiv \sum_{\omega_{jt+1}} \bar{V}(s_{jt+1}) dP(\omega_{jt+1} | \omega_{jt}, i_{jt} = 0, e_{jt} = 1), \tag{1.10}$$

$$E\bar{V}^i \equiv \sum_{\omega_{jt+1}} \bar{V}(s_{jt+1}) dP(\omega_{jt+1} | \omega_{jt}, i_{jt} = 1, e_{jt} = 0), \tag{1.11}$$

$$E\bar{V}^{ie} \equiv \sum_{\omega_{jt+1}} \bar{V}(s_{jt+1}) dP(\omega_{jt+1} | \omega_{jt}, i_{jt} = 1, e_{jt} = 1), \tag{1.12}$$

A firm's organization decision for each of the four choices from the maximization problem must then satisfy the following inequality:

$$\begin{aligned} & n_{jt}\delta E\bar{V}^n + e_{jt}(\delta E\bar{V}^e - C^e) + i_{jt}(\delta E\bar{V}^i - C^i) + ie_{jt}(\delta E\bar{V}^{ie} - C^i - C^e) \\ & > \{(1 - n_{jt})\delta E\bar{V}^n \ \& \\ & (1 - e_{jt})(\delta E\bar{V}^e - C^e) \ \& \\ & (1 - i_{jt})(\delta E\bar{V}^i - C^i) \ \& \\ & (1 - ie_{jt})(\delta E\bar{V}^{ie} - C^i - C^e)\}. \end{aligned} \tag{1.13}$$

For a given state $(\omega_{jt}, K_{jt}, i_{jt-1}, e_{jt-1})$, the set of cost draws that are consistent with the inequality (1.13) are depicted in Figure 1.1. In each of the three panels, the plane is divided into four disjoint regions by solid lines. Each region represents one out of the four R&D modes (n, e, i, ie) and is sectioned off by the set of inequalities that specifies the corresponding action that gives the highest return. The specific expressions for the cutoffs are:[18]

$$\begin{aligned} a &= \delta(E\bar{V}^i - E\bar{V}^n),\ a' = \delta(E\bar{V}^e - E\bar{V}^n), \\ b &= \delta(E\bar{V}^{ie} - E\bar{V}^e),\ b' = \delta(E\bar{V}^{ie} - E\bar{V}^i), \\ c &= \delta(E\bar{V}^i - E\bar{V}^e),\ m = \delta(E\bar{V}^{ie} - E\bar{V}^n), \\ ab &= \delta(E\bar{V}^i - E\bar{V}^n),\ a'b' = \delta(E\bar{V}^e - E\bar{V}^n). \end{aligned}$$

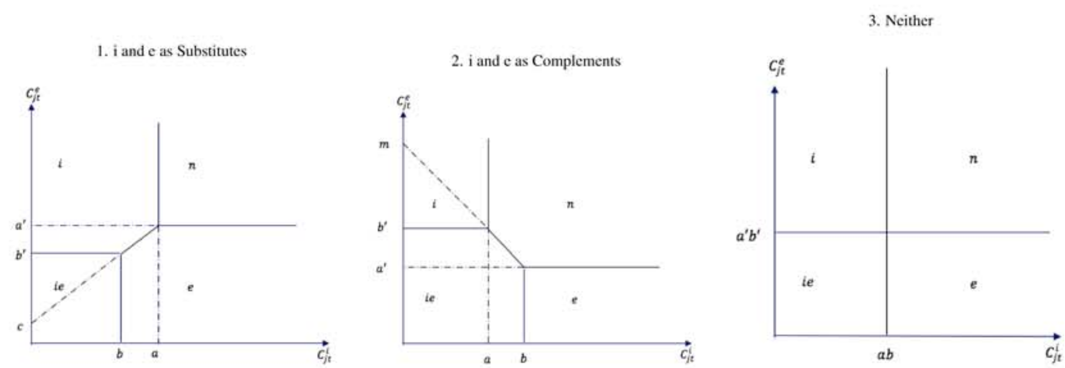


Figure 1.1: R&D Modes Defining Regions

There are three panels in Figure 1.1 because the ordering of the marginal benefits between different research modes (a, b and a', b') depends on whether the two R&D choices are substitutes, complements or neither. Panel 1 depicts the case when the two activities are substitutes; farming out research in addition to in-house research has the similar implication of diminishing returns to increasing research intensity. This scenario can be represented by the following restriction:

$$\delta(E\bar{V}^{ie} - E\bar{V}^n) < \delta(E\bar{V}^i - E\bar{V}^n) + \delta(E\bar{V}^e - E\bar{V}^n). \tag{1.14}$$

[18]The derivations are outlined in the Appendix.

Panel 2 depicts the case of complementarity with the opposite inequality sign in (1.14), or that the marginal benefit of doing one type of R&D is greater when the firm also does the other. This has the interpretation that in-house R&D may increase a firm's absorptive capacity to obtain tacit knowlege from external units, a concept first discussed in Cohen and Levinthal (1989). Last but not least, Panel 3 depicts the case when neither substitution nor complementarity is at play, and so the marginal benefit of doing both types of R&D is simply the sum of each activity.

In which scenario a firm ends up depends on the static returns parameters in the productivity process (1.6), the relative costs associated with each R&D mode, and the productivity and size of firm. The estimation of the model allows one to assess the importance of research costs, and how well this model uses them to explain firms' research decisions.

1.5 Estimation Strategy

The estimation procedure consists of two stages. The first stage is the estimation of the parameters in a firm's profit function: Φ_t, η, marginal cost function parameters in $\mathrm{mc}(k_{jt}, X_t)$, as well as the returns to R&D modes in the Markov process $g(\omega_{jt-1}, e_{jt-1}, i_{jt-1}, ie_{jt-1})$. These static parameters form the basis of a firm's expectation of its productivity evolution and discounted profit streams, which are then embedded in the dynamic estimation of the firm's problem to obtain the R&D cost distribution parameters using Bayesian Markov Chain Monte Carlo (MCMC).

1.5.1 Identification Strategy and Estimation Procedure of the Static Parameters

To estimate the parameters in the revenue and profit function, it remains to specify the marginal cost function.[19] For a given production function, the corresponding short-run marginal cost (1.3) can be estimated as a function of factor prices (x_t) and capital:

$$mc_{jt} = \beta_0 + \beta_x x_t - \beta_k k_{jt} - \psi_{jt}. \tag{1.15}$$

[19] In this section, all lower case letters denote the variables in logs.

Substituting this expression into revenue and rearranging the terms give the following equation for revenue:

$$r_{jt} = \text{const.} + \beta_t t + (\eta - 1)(\beta_k k_{jt} + \omega_{jt}) + \nu_{jt}, \tag{1.16}$$

where the constant term absorbs (η, β_0), and $\beta_t t$ captures the economy-wide market index and factor prices.[20]

However, the revenue equation (1.16) is not directly estimable because productivity is unobserved. In earlier papers of the productivity estimation literature, such as Olley and Pakes (1996), Levinsohn and Petrin (2003a) and Ackerberg et al. (2006), productivity is assumed to be an exogenous AR(1) process, whereas Doraszelski and Jaumandreu (2013) endogenize the process and allow it to evolve according to a firm's R&D expenditures. Aw et al. (2011) adapt the endogenous specification and estimate productivity from the revenue equation in the first stage. With the estimated productivity series, they obtain the coefficients for the Markov process in the second stage.[21] This paper follows the same estimation strategy and is outlined as follows.

I assume a firm's choices on material depend on their state variables k_{jt} and ω_{jt}:

$$m_{jt} = m(k_{jt}, \omega_{jt}). \tag{1.17}$$

As long as the demand function for material is monotonically increasing in productivity and capital, one can invert (1.17) and use a function of capital and materials observed in the data to proxy for productivity: $\omega_{jt} = m^{-1}(k_{jt}, m_{jt})$. I assume this inverted function $m^{-1}(\cdot)$ to be a polynomial of degree two in its arguments.[22] Substituting out ω with the polynomial gives an estimable revenue equation with observables:

$$r_{it} = \text{const.} + \beta_t t + (\eta - 1)(\beta_k k_{jt} + m^{-1}(k_{jt}, \omega_{jt})) + \nu_{it}. \tag{1.18}$$

[20] The exact expressions are: const. $= -(\eta-1)\beta_0 - (\eta-1)\ln\frac{\eta}{\eta-1}$ and $\beta_t t = \ln \Phi_t + \beta_x x_t$.

[21] Ackerberg et al. (2006) also use a two-step estimation procedure but with an exogenous AR(1) process.

[22] It's been argued that this approach of estimating productivity is subject to misspecification. In Appendix 1.9.5.2, I lay out an alternative estimation strategy and its corresponding results.

However, one will not be able to separately identify η, β_k and the coefficient for capital in the polynomial. Instead, an estimate for the function $h(\cdot)$ can be obtained:

$$r_{jt} = \text{const.} + \beta_t t + \underbrace{\beta_1 k_{jt} + \beta_2 k_{jt}^2 + \beta_3 m_{jt} + \beta_4 m_{jt}^2 + \beta_5 k_{jt} m_{jt}}_{h(k_{jt}, m_{jt})} + \nu_{jt}. \tag{1.19}$$

We know from (1.18) that $\widehat{h_{jt}} \equiv (\widehat{\eta} - 1)(\beta_k k_{jt} + \omega_{jt})$, or $\omega_{jt} = \frac{\widehat{h_{jt}}}{\widehat{\eta}-1} - \beta_k k_{jt}$. Substituting the expression into the productivity process (1.6) gives:

$$\begin{aligned} \widehat{h_{jt}} &= (\eta - 1)\alpha_o + (\widehat{\eta} - 1)\beta_k k_{jt} + (\widehat{\eta} - 1)\alpha_1(\widehat{h_{jt-1}} - (\widehat{\eta} - 1)\beta_k k_{jt-1}) \\ &\quad + (\widehat{\eta} - 1)\alpha_e e_{jt-1} + (\widehat{\eta} - 1)\alpha_i i_{jt-1} + (\widehat{\eta} - 1)\alpha_{ie} ie_{jt-1} + (\widehat{\eta} - 1)\xi_{jt}, \end{aligned} \tag{1.20}$$

which can be estimated using NLS with $\widehat{\eta}$ that is estimated from the profit function:

$$\pi_{jt} = \ln \eta + r_{jt} + e_{jt}. \tag{1.21}$$

The identification of the parameters in the productivity process relies on two main assumptions. The first is that shocks at time t ξ_{jt} are orthogonal to the decisions made in the previous period. The second assumption is that the estimate for $h(\cdot)$ is unbiased. More specifically, if there exists unobservables that are correlated jointly with revenue and R&D organization structure in the first stage (1.19), then the parameters in (1.19) are biased because unobserved productivity is a function of R&D modes. If the estimate for $h(\cdot)$ is biased then subsequently so are the estimates in (1.20).

The rich questionnaire of this dataset allows me to incorporate firm-specific variables that may be correlated with R&D decisions and at the same time affect revenue. Incorporating the corresponding dimensions of firm heterogeneities increases the precision of the estimates, but at the expense of a substantially longer estimation time of the dynamic component of the model due to an increase in the number of state variables. In Appendix 1.9.5.3 Table 1.12, I show estimation results for including these controls chosen by economic reasoning based on the innovation literature. The qualitative features of the model are still preserved, and in the rest of this paper I focus on results from the baseline case that contains no additional controls.

1.5.2 Estimation Procedure of the Dynamic Parameters

Knowing the parameters for the revenue function and the productivity process, a firm can compute its profit accordingly, form expectations of its productivity evolution, and arrive at the continuation values of different R&D modes to compare them with the actual cost draw. These cost draws are observable only to the firms, and the dynamic parameters of interest are the parameters of the cost distributions. How variations in the data identify these parameters can be explained as follows. R&D set up costs are identified from the probability of entry across firms with similar profit streams but differ in their previous period status, whereas continuation costs are identified from the relative frequencies (probability) of exit in R&D modes.[23] In what follows, I outline the estimation strategy in detail.

I assume the independent cost variables (F^i, f^i, F^e, f^e) follow the following exponential distribution ($x > 0$, $y > 0$):

$$F^i \sim G(x) = 1 - \exp\left(-x/\gamma^F\right),\ f^i \sim G(x) = 1 - \exp\left(-x/\gamma^f\right),$$
$$F^e \sim L(y) = 1 - \exp\left(-y/\lambda^F\right),\ f^e \sim L(y) = 1 - \exp\left(-y/\lambda^f\right)$$

where $\Theta = (\gamma^F, \gamma^f, \lambda^F, \lambda^f)$ is the set of distribution means for internal and external sunk and fixed costs. To estimate Θ with the dynamic model, I construct the likelihood function for the observed patterns of firms' R&D decisions:

$$\mathcal{L}(\Theta) = \prod_j^J \prod_t^T P(i_{jt}, e_{jt} | s_{jt}, \Theta). \tag{1.22}$$

$P(i, e|s, \Theta) = P_{ie}^{ie} P_i^{i(1-e)} P_e^{(1-i)e} P_n^{(1-i)(1-e)}$ is the probability of the firm's observed R&D decision conditioning on its productivity, capital level, and R&D modes in the previous period. $\prod_t^T P(i, e|s, \Theta)$ is then each firm's contribution to the likelihood. For a particular vector of Θ, these conditional

[23]One may argue from the transition matrix shown in Table 1.2 that there is almost no exit in internal R&D in the data, but the upper bounds for internal sunk and fixed costs are still identified despite this observation.

choice probabilities are evaluated over the regions specified in Figure 1.1 with cutoffs that are determined by the value functions.[24] But since the value functions are themselves functions of the cost parameters, an iterative estimation procedure is invoked. One would start with guessed values for the vector Θ, compute the value functions from iterating the Bellman equation (1.8), maximize $\mathcal{L}$ with respect to Θ and iterate this process until convergence (Rust, 1987). In this paper, I use instead Bayesian Markov Chain Monte Carlo (MCMC) as Aw et al. (2011) do in their paper.

The intuitions behind Bayesian MCMC is as follows. Rather than assuming Θ as fixed to obtain their point estimates for which the observed data are most probable to occur, Bayesian inference treats these parameters as random variables and characterize the estimates and associated uncertainties with their posterior distributions instead. The posterior is computed by multiplying the likelihood and the prior distribution of the parameters. However, since the distribution of the likelihood function is unknown, one can use MCMC to sample parameter values from the posterior and then use these values to approximate the posterior means. In the end, Bayesian MCMC in this context is especially valuable as it allows one to get around the problem of maximizing a non-smooth or non-linear likelihood with respect to the parameters, as will be the case in this paper due to the joint-decision problem. The detailed steps for the value function computation and the construction of the likelihood function are spelled out in Appendix 1.9.2, whereas a more detailed discussion of the Bayesian MCMC methodology can be found in Appendix 1.9.3.

1.6 Estimation Results

1.6.1 Static Parameters

In addition to the full model, I also estimate a version of it that does not distinguish R&D modes for comparison. Recall the static parameters of interest are those in the revenue function and the Markov productivity process (eq. 1.19 & 1.20). Stage 1 results and the estimate for the elasticity of substitution from the relationship between revenue and profit are presented in Table 1.3. These results are the same for both models. The variables in the polynomial of capital and materials are

[24] The expressions for each of the probabilities are outlined in Appendix 1.9.2.

significant in explaining log revenue. The year dummies capture variations in aggregate demand and common input prices. Their coefficients are insignificant, meaning that the index levels are not that different from those in the base year 2000. The coefficient on profit is the elasticity of substitution among varieties; it is significant and implies a markup of 26%.

With the first stage results, one can construct an expression that contains productivity, substituting into the endogenous Markov productivity process to estimate the parameters using nonlinear least square. The results for the full model and the model with a single R&D choice are presented in Column (1) and (2) of Table 1.4. In the full model, in-house and external R&D are expected to increase ω by 2.6% and 2.1% respectively. The benefits of doing both types of R&D is significant and the highest of all at 3.9%. The interaction of these two modes is negative but not statistically different from zero. In the model where research organization is not taken into consideration, the return to conducting research either via in house staff or with external units is 3%.

1.6.2 Dynamic Parameters

The dynamic estimates that correspond to these two models are summarized in Table 1.5. Column 1 presents the means of the exponential distributions, whereas Column 2 presents the average realized cost draws paid by firms from 100 simulations. The average research cost as a share of revenue and profit ratios are recorded in Columns (3) and (4), also computed from 100 simulations. These averages are weighted by a firm's size measured as revenue and profit relative to the economy output.

The top panel of Table 1.5 presents the results for the full model are shown. The unconditional (conditional) costs show that internal R&D costs almost 2.5 (1.5) times more than external to start, but external R&D costs almost 2.5 (1.5) times more to maintain. These estimates elucidate the trade-offs between the two research options, and are consistent with the observed transition in R&D modes. Despite the higher returns of in-house research, smaller firms select into external R&D due to its lower start up cost. Once a firm overcomes the set-up cost to establish an in-house research team, it almost never quits. However, these internal R&D performers may additionally switch in

Table 1.3: Static Estimates I

Estimation results for stage 1 and elasticity of substitution. Regression of log revenue on year dummies and a polynomial in log materials and log capital in Column (1) and on profit in Column (2).

	(1)	(2)
log profit		4.871***
		(0.143)
log materials	0.479***	
	(0.097)	
log materials sq.	0.034***	
	(0.005)	
log capital	0.223***	
	(0.075)	
log capital sq.	0.025***	
	(0.005)	
log materials * log capital	-0.052***	
	(0.009)	
year 2001	-0.012	
	(0.038)	
year 2002	0.008	
	(0.035)	
constant	3.610***	
	(0.516)	
N	3798	3797
R^2	0.841	0.234

Robust standard error corrected for heteroskedasticity in parentheses, $p < .1^{*}$, $p < .05^{**}$, $p < .01^{***}$.

Table 1.4: Static Estimates II

Estimation results for the Markov process of productivity.
Column (2) consists of results for the model with
single R&D category.

	(1)	(2)
log capital	0.070***	0.070***
	(0.015)	(0.015)
α_0	0.120***	0.117**
	(0.018)	(0.017)
α_1	0.945***	0.946***
	(0.011)	(0.011)
α_e	0.021**	
	(0.10)	
α_i	0.026***	
	(0.007)	
α_{ie}	-0.008	
	(0.516)	
α_{rd}		0.030***
		(0.006)
N	2531	2405

Bootstrapped standard errors in parentheses,
$p < .1^{*}$, $p < .05^{**}$, $p < .01^{***}$. α_i not
statistically different from α_e.

Table 1.5: Dynamic Parameter Estimates

parameters		(1) distribution mean (mil.RMB)	(2) realized mean (mil. RMB)	(3) $\sum \frac{R}{\sum R}\frac{C}{R}$	(4) $\sum \frac{\Pi}{\sum \Pi}\frac{C}{\Pi}$
Full Model with R&D Organization					
Internal	γ^F	688	44.67	0.36	1.56
		(47.78)	(8.46)		
	γ^f	3.14	3.13	0.018	0.078
		(0.13)	(0.13)		
External	λ^F	274	33.38	0.13	0.58
		(11.68)	(4.16)		
	λ^f	8.89	7.54	0.030	0.133
		(0.31)	(0.45)		
Model without R&D Organization					
Startup Cost		516.58	62	0.56	2.478
		(125.35)	(10.82)		
Continuation Cost		4.5	4.4	0.027	0.119
		(0.301)	(0.18)		

In 2002, 1USD≈8RMB. SE in brackets.

and out of external R&D, to avoid having to build up in-house capacity for one-off projects.[25]

For either mode, startup cost is many times more than continuation cost (roughly 15 times over in the case of conditional internal cost and 7 times in the case of external), which rationalizes the persistence in R&D activities observed in the data. Forward looking firms will want to stick it out in face of bad continuation draws in order to avoid having to pay for the start up cost again. Lastly, this startup/fix ratio is also much higher for internal R&D. Coupled with the finding that external R&D costs more to maintain, the higher startup/fix ratio for internal R&D implies that in-house R&D is a continuing process of investment in knowledge capacity while external R&D is relatively short-lived.[26]

[25]One can think of these one-off projects as projects that require external expertise or projects that require additional staff that the firm is not able to accomodate.

[26]These ratios are much higher than those found in the referenced papers, and it could be attributed to the fact that I define R&D very differently here. Firms must have active R&D personnel in-house and have declared to have contractual relationship with external partners. In both Aw et al. (2011) and Peters et al. (2013), where R&D is not split into categories, a firm is active in research as long as it has positive R&D expenditures. In Amoroso (2014), a firm is active if it has R&D and if it has product or process innovations introduced.

The summary of these costs as ratios of a firm's revenue and profit in Columns (3) & (4) present a similar pattern. The commitment to establish an internal R&D program is substantial, amounting to almost 1.5 times per-period profit flow. This gives way to the possibility that credit-constraint is one of the reasons that firms do not choose internal R&D in reality. Lastly, the continuation costs as a share of revenue come out to be very close to R&D intensity measured in the data.

The bottom panel of Table 1.5 presents the results for the model without considerations for R&D organization. The findings are qualitatively similar to those in the existing literature. Startup cost is substantially higher than continuation cost. The actual magnitude varies from study to study, however. How R&D actvities are defined, what their returns are, and the transitions in and out of research all play important roles in determining their levels. The conditional costs from the average of the 100 simulations are comparable to those in the full model; the startup cost is lower than the sum of internal and external startup cost, and the continuation cost is slightly lower than the average of the internal and external continuation costs.

To examine in-sample performance of the estimated parameters for the full model, I compare how well the model predicts firms' actions and how well the predicted series of productivity evolution match those in the data. More specifically, I use the dynamic parameters estimates and simulate 100 data sets of 3 periods taking the initial period R&D status as given, then compute the transition matrix to report the averages over the 100 simulations in Table 1.6. The overall fit is reasonable compared with those in Table 1.2. However, the model over-predicts the propensity to undertake either mode of research for none-R&D performers. External R&D performers also quit or additionally take on internal R&D much more often than what is observed in the data, and firms that use both R&D modes quit external more frequently than in reality.[27]

In addition to examining the transition matrix, I compare the percentage of correct predictions for each R&D modes. Averaging across the 100 simulations, the model predicts internal and external R&D modes correctly 91.4% and 79.4% of the time. These percentages are lower when I break up the modes further; they are 78%, 55.5%, 75.5% & 65.3% for none, external only, internal only and both. Lastly, I compare the kernel density of data versus simulated productivity level

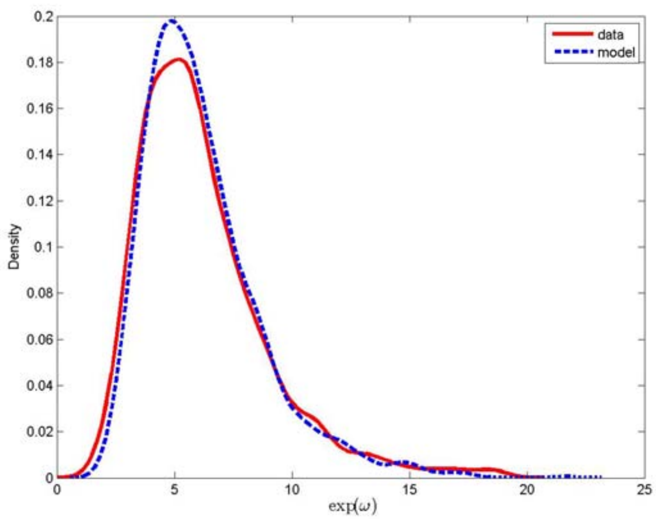
[27]These discrepancies hint at the possibility that the external fixed cost may still too high and the sunk cost too low.

from period 3 in Figure 1.2. The Kolmogorov-Smirnov test does not reject the null that these two samples are from the same distribution, serving as additional evidence that the model fit is good.

Table 1.6: Transition Matrix from Simulated Data using Model Estimates

	noRD_t	extRD_t	intRD_t	both_t
Model with R&D Organization				
noRD_{t-1}	0.871	0.068	0.055	0.006
extRD_{t-1}	0.174	0.740	0.009	0.077
intRD_{t-1}	0.015	0	0.875	0.110
both_{t-1}	0	0	0.119	0.871
N (%)	40%	9.7%	29.6%	20.7%
Model without R&D Organization				
noRD_{t-1}	0.903	0.097		
extRD_{t-1}	0.030	0.970		
N (%)	42%	58%		

Figure 1.2: Kernel Density of Data versus Simualted Productivity Level



1.7 Simulation Results

In this section, I perform counter-factual exercises to examine how R&D cost reductions impact organization dynamics, as well as how the resulting dynamics translate into productivity growth. Specifically, I take the sample of firms' initial year productivity and R&D organization as given, then simulate cost shocks and solve for the firms' decisions forward under a 50% reduction in the cost parameters.

The key findings from the simulation exercises can be summarized as follows. Both startup and continuation cost reductions stimulate R&D activities, but startup cost reduction leads to the most productivity growth. Specifically, adjustments along the external margin account for most of the changes in research status and almost 70% of the productivity gain. Relative to the model without R&D organization, allowing firms to organize their research leads to a 13% more increase in aggregate productivity when there is cost reduction. The reason is under a model with single R&D choice, fewer firms are able to start research via a cheaper mode and larger firms are not able to additionally contract with external units. In the rest of this section, I explain these findings with simulation results in more details.[28]

Table 1.7 reports the changes in the transition matrix, after a 50% reductions in startup and continuation R&D costs. More specifically, the changes are presented as the proportion of firms that start and continue R&D relative to the proportions before reduction (Table 1.6). They are averages over 100 simulations and each simulation consists of three time periods, the same as that of the panel. I also report the one period growth in the aggregate productivity weighted by revenue share after cost reduction in the last row:

$$\Delta\widetilde{\omega_t} \equiv \widetilde{\omega_t} - \widetilde{\omega_{t-1}} = \sum \frac{R_{jt}}{\sum R_{jt}}\omega_{jt} - \sum \frac{R_{jt-1}}{\sum R_{jt-1}}\omega_{jt-1}.$$

Focusing on the results for reduction in startup cost first, Row 1 of Columns (1)-(4) shows that 5.6% of the none-R&D performers take advantage of the startup cost reduction to start research,

[28]Like Peters et al. (2013), I also do not make any welfare analysis on whether the costs of the subsidy program outweighs the benefits. There is no spillover effects of innovation in the model, thus firms internalize all R&D benefits.

with internal only and external only research modes absorbing comparable fraction of entrants. The fact that not more firms are selecting into internal R&D to take advantage of the lower continuation cost implies that internal startup cost is still too high.

For R&D firms (Rows 2-4), there are asymmetric responses in how they switch modes in response to startup cost reduction. Some external only R&D performers quit research (the category n_t sees an increase of 1.4%); others take advantage of cost reductions to start in house R&D (3.6%), but some do so at the expense of dropping external activity (1%). Internal R&D firms rarely quit in-house research but they adjust by either additionally taking on external R&D or dropping it. The quitting effect is attributed to the narrowing of the "inaction band," it refers to the set of random costs that renders none performers unwilling to enter and performers unwilling to quit.

Nonetheless, the quitting channel under startup cost reduction did not operate pervasively enough to counteract the gain from the entrants for this sample of firms in the aggregate, especially in the case of internal R&D. In the end, the one period productivity growth from reducing both internal and external R&D startup costs is 1.36%, which is a 16% increase compared to before reduction where $\Delta\widetilde{\omega_{jt}}$ is 1.17%. Moreover, 70% of the gain comes from adjustment along the external margin.[29] In the model without R&D organization, one period productivity growth is just 3% more than before cost reduction.[30] This additionally shows the importance of considering firm organization in research.

In contrast to startup cost reduction, no firms will quit as a result of continuation cost reduction. Some firms may additionally find entry attractive because the marginal benefit of research is now higher than the startup cost. Columns (5)-(8) in Table 1.7 show results consistent with expectation. Among firms that were not previously active in research, external R&D also attracts the most entrants, a proportion that is comparable to that under sunk cost reduction (2.6%). Reduction in internal continuation cost, on the other hand, attracts almost no entrants since the startup cost is still too high. A sizable percentage of internal R&D only firms additionally take on external R&D. Firms that do both remain active in both modes of research, and more than 50% of the firms that

[29]This is computed from examining the gain when only external startup cost is reduced.

[30]Changes in transition matrix for the model with no R&D organization is reported in Appendix 1.9.6.

would have dropped external activities now continue.

While the net adjustments in R&D status appear to be more favorable in the case of continuation cost reduction, aggregate productivity growth is 1.16% and even slightly lower compared to before cost reduction. To understand why productivity gain is smaller under continuation cost, I examine more closely the affected firms. Table 1.8 presents productivity growth from firms that changed status as a result of the cost reduction, their revenue share in the economy and the number of these firms as the share of the total number of firms. Panel A features firms that have either started or managed to keep research activities after cost reduction, while B features firms that quit as a result of startup cost reduction.

Table 1.8 shows that sunk cost reduction induces productivity gain from 5.5% of firms but they make up 11.6% of total revenue. Firms that quit external R&D because of the option value effect counteracts more than half of the productivity gain, but since there are only a few of them the negative effect did not amount to any significance in the end. Continuation cost reduction induces productivity gain from 6.9% of firms yet they make up only 1.6% of total revenue. This suggests that while continuation reduction induced some entry, it also allowed firms less productive in research to continue.

Table 1.7: Changes in Transition Matrix from 50% Reductions in R&D Startup and Continuation Costs, Averaging over 100 Simulations

	Reduction in Sunk				Reduction in Continuation			
	n_t	e_t	i_t	ie_t	n_t	e_t	i_t	ie_t
n_{t-1}	-0.056	0.026	0.023	0.006	-0.036	0.026	0.008	0.003
e_{t-1}	0.014	-0.060	0.010	0.036	-0.126	0.123	-0.006	0.009
i_{t-1}	0.001	0	-0.050	0.049	-0.016	0	-0.013	0.028
ie_{t-1}	0	0	0.018	-0.018	-0.009	0	-0.078	0.087
N (%)	-0.040	0.008	0.002	0.031	-0.058	0.037	-0.027	0.049
$\Delta\widetilde{\omega}_t$	0.0136				0.0116			

Table 1.8: Summary of Productivity Gain from Affected Firms

	Start Up Cost Reduction		Continuation Cost Reduction	
Panel A				
	Internal	External	Internal	External
Productivity Growth (%)	2.73	1.96	3.01	1.97
Revenue Share (%)	2.47	9.12	0.27	1.31
Share of Firms (%)	1.96	3.56	1.39	5.52
Panel B				
	Internal	External	Internal	External
Productivity Growth (%)	-3.05	-1.54	N.A.	N.A.
Revenue Share (%)	~ 0	0.13	N.A.	N.A.
Share of Firms (%)	~ 0	0.62	N.A.	N.A.

1.8 Conclusion

This paper estimates a dynamic structural model of firms' R&D organization decisions with a panel of Chinese firms to quantify the benefits and costs of R&D organization. The model is driven by the following empirical observations: firms conduct research in house, with external units or via a combination of both modes; there exists a pecking order in firms' selection into research modes by size; once an R&D mode is chosen, firms additionally transition in and out of research. To account for these features, productivity is modeled to evolve endogenously according to firms' internal and external R&D decisions, the costs of starting and continuing research are random and differ by R&D modes. The existence of fixed costs select the largest firms into research and this selection is then reinforced by the different returns to R&D, and research cost uncertainty explains the transitions in and out of R&D modes observed in the data.

Model estimates show that in-house R&D is more effective, costs less to maintain but substantially more to start than external R&D. These estimates elucidate the trade-offs between the two research modes, and are consistent with the observed transition in R&D modes. Despite the higher returns of in-house research smaller firms select into external R&D due to its lower startup cost. Once a firm overcomes the startup cost to establish an in-house research team, it almost never quits. These findings reveal the intuitions for the trade-offs firms face in organizing their R&D: internal

R&D requires a high level of initial commitment but low fixed cost as opposed to external R&D; internal R&D is a continuing process of investment in knowledge capacity while external R&D is relatively short-lived.

The counter-factual exercises of reducing cost uncertainty shows that productivity growth is 2.5 times higher when firms can choose how to organize R&D, and the ability to adjust along external R&D contributes to almost 70% of the difference. Comparing startup versus continuation cost reduction, I find that while continuation cost reduction leads to the most changes in research status, startup cost reduction leads to the most productivity gain. In investigating this finding further reveals that growth implications of cost reductions is determined by the degree of uncertainty associated with different research modes and the initial distribution of firms' research status and productivity. Thus failing to take these factors into account jointly could mislead R&D policies.

1.9 Appendix

1.9.1 Derivation of the Marginal Cost Function

Assuming firms take capital as given and the factor markets are competitive, then the firm's input decisions are solutions to the following problem:

$$\min_{L_{jt},M_{jt}} W_{jt}L_{jt} + P_{Mjt}M_{jt} \text{ subject to } F(K_j, L_{jt}, M_{jt}) \geq Q_{jt}/\exp(\psi_{jt}). \tag{1.23}$$

The first order conditions are

$$\begin{aligned} W_{jt} &= \lambda F_L \\ P_{Mjt} &= \lambda F_M \\ F(K_j, L_{jt}, M_{jt}) - Q_{jt}/\exp(\psi_{jt}) &= 0, \end{aligned}$$

and they can be used to solve for the 3 unknowns:

$$\begin{aligned} \lambda &= C_\lambda(Q_{jt}/\exp(\psi_{jt}), W_{jt}, P_{Mjt}, K_j), \\ M^*_{jt} &= M(Q_{jt}/\exp(\psi_{jt}), W_{jt}, P_{Mjt}, K_j), \\ L^*_{jt} &= L(Q_{jt}/\exp(\psi_{jt}), W_{jt}, P_{Mjt}, K_j). \end{aligned}$$

The total cost is

$$TC_{jt} = r_{jt}K_j + C(K_j, W_{jt}, P_{Mjt}, K_j, Q_{jt}/\exp(\psi_{jt})),$$

and the marginal cost is

$$MC_{jt} = \frac{\partial C}{\partial Q_{jt}}(K_{jt}, W_{jt}, P_{Mjt}, Q_{jt}/\exp(\psi_{jt}))\exp(-\psi_{jt}). \tag{1.24}$$

By Shephard's lemma, optimal input M_{jt} is the partial derivative of the total cost with respect to

P_{Mjt}:

$$M_{jt} = \frac{\partial C}{\partial P_{Mjt}}(K_{jt}, W_{jt}, P_{Mjt}, Q_{jt}/\exp(\psi_{jt})), \tag{1.25}$$

which is then be inverted and used to replace $Q_{jt}/\exp(\psi_{jt})$ in (1.24). Finally, we arrive at the new marginal cost function:

$$MC_{jt} = MC(K, M_{jt}, W_{jt}, P_{Mjt})\exp(-\psi_{jt}), \tag{1.26}$$

that gives us the empirical counterpart as in (1.3).

1.9.2 Value Function Iteration and Likelihood Function

The actual cost draws are only observable to the firm, and I work with the integrated value function instead over the random costs to arrive at the following expression:[31]

$$\begin{aligned} EV = \Pi + P_n\delta EV^n + P_e(\delta EV^e - E\left[C^e|e\right]) \\ + P_i(\delta EV^i - E\left[C^i|i\right]) + P_{ie}(\delta EV^{ie} - E\left[C^i + C^e|ie\right]), \end{aligned}$$

where P_n, P_e, P_i, P_{ie} stand for the probabilities of conducting R&D via a specific mode given the cost distribution, and $E\left[C|\cdot\right]$ are the conditional expectations of the cost draws.[32] With the following assumed distributions for the cost parameters (for $c^i > 0$, $c^e > 0$):

$$\begin{aligned} C^i \sim G(c^i) = 1 - \exp\left(-c^i/\gamma\right), \\ C^e \sim L(c^e) = 1 - \exp\left(-c^e/\lambda\right). \end{aligned}$$

The means of the distribution for sunk and fixed costs are γ^F, γ^f for internal R&D, and λ^F, λ^f for external R&D.

[31]To simplify notations I have suppressed the subindex for firm and time.

[32]Whether C^i (C^e) takes F^iorf^i (F^eorf^e) depends on previous period status. Abusing the notation a little, I use C^i, C^e as random variables.

The expressions for probabilities are as follows:

$$\begin{aligned}
P_n = P(\delta EV^n > \delta EV^e - C^e &\text{ and}\\
\delta EV^n > \delta EV^i - C^i &\text{ and}\\
\delta EV^n > \delta EV^{ie} - C^i - C^e),&\\
P_i = P(\delta EV^i - C^i > \delta EV^n &\text{ and}\\
\delta EV^i - C^i > \delta EV^e - C^e &\text{ and}\\
\delta EV^i - C^i > \delta EV^{ie} - C^i - C^e),&\\
P_e = P(\delta EV^e - C^e > \delta EV^n &\text{ and}\\
\delta EV^e - C^e > \delta EV^i - C^i &\text{ and}\\
\delta EV^e - C^e > \delta EV^{ie} - C^i - C^e),&\\
P_{ie} = P(\delta EV^{ie} - C^i - C^e > \delta EV^n &\text{ and}\\
\delta EV^{ie} - C^i - C^e > \delta EV^i - C^i &\text{ and}\\
\delta EV^{ie} - C^i - C^e > \delta EV^e - C^e).&
\end{aligned}$$

These conditions can be simplified when there is complementarity, substitutability, or neither between the two activities. When there is substitutability, the probabilities are as follows:

$$\begin{aligned}
P_n &= P(C^i > a,\, C^e > a')\\
&= [1 - G(a)][1 - L(a')],\\
P_e &= P(C^i > b,\, C^e < b') + P(b' < C^e < a',\, C^e - C^i < -c)\\
&= [1 - G(b)] * L(b') + \frac{k}{\lambda}[1 - G(c)][H(a') - H(b')],\\
P_i &= P(C^i < a,\, C^e > a') + P(b' < C^e < a',\, C^e - C^i > -c)\\
&= G(a) * [1 - L(a')] + L(a') - L(b') - \frac{k}{\lambda}[1 - G(c)][H(a') - H(b')],\\
P_{ie} &= P(C^i < b, C^e < b') = G(b)L(b').
\end{aligned}$$

where[33]

$$a = \delta(EV^i - EV^n),\ a' = \delta(EV^e - EV^n),$$

$$b = \delta(EV^{ie} - EV^e),\ b' = \delta(EV^{ie} - EV^i),$$

$$c = \delta(EV^i - EV^e),\ k = \frac{\gamma\lambda}{\gamma + \lambda},\ H(z) = 1 - \exp(-z/k),$$

and the probabilities for internal and external are derived by dividing integration over the non-rectangular region into two for tractability.

The corresponding conditional expectations multiplied by the probabilities are then:

$$\begin{aligned}
P_e E[C^e|e] &= P_e\{E[C^e|C^i > b,\, C^e < b'] + E[C^e|b' < C^e < a',\, C^e - C^i < -c]\} \\
&= \int_b^\infty \int_0^{b'} c^e l(c^e) g(c^i) dc^e dc^i + \int_{b'}^{a'} \int_{c^e+c}^\infty c^e l(c^e) g(c^i) dc^e dc^i \\
&= [1 - G(b)][L(b')(b' + \lambda) - b'] \\
&+ \frac{k}{\lambda}[1 - G(c)]\{[1 - H(b')](b' + k) - [1 - H(a')](a' + k)\}, \\
P_i E[C^i|i] &= P_i\{E[C^i|C^i < a,\, C^e > a'] + E[C^i|b' < C^e < a',\, C^e - C^i > -c]\} \\
&= \int_{a'}^\infty \int_0^a c^i g(c^i) l(c^e) dc^i dc^e + \int_{b'}^{a'} \int_0^{c^e+c} c^i g(c^i) l(c^e) dc^i dc^e \\
&= [1 - L(a')][G(a)(a + \gamma) - a] \\
&- \frac{k}{\lambda}[1 - G(c)]\{[1 - H(b')](b' + k) - [1 - H(a')](a' + k)\} \\
&- \frac{c + \gamma}{\lambda} k[1 - G(c)][H(a') - H(b')] + \gamma[L(a') - L(b')] \\
P_{ie} E[C^i + C^e|ie] &= P_{ie}\{E[C^i|C^i < b,\, C^e < b'] + E[C^e|C^i < b,\, C^e < b']\} \\
&= L(b')[G(b)(b + \gamma) - b] + G(b)[L(b')(b' + \lambda) - b'].
\end{aligned}$$

[33]In case of complementarity, the diagonal cut off is $m = \delta(EV^{ie} - EV^n)$

In the case of complementarity, the probabilities and conditional expectations are:

$$
\begin{aligned}
P_n &= P(C^i > b,\, C^e > a') + P(a' < C^e < b',\, C^e + C^i > m) \\
&= [1 - G(b)][1 - L(a')] + \frac{k'}{\lambda}[1 - G(m)][H'(b') - H'(a')], \\
P_e &= P(C^i > b,\, C^e < a') \\
&= [1 - G(b)]L(a'), \\
P_i &= P(C^i < a,\, C^e > b') \\
&= G(a)[1 - L(b')], \\
P_{ie} &= P(C^i < b,\, C^e < a') + P(a' < C^e < b',\, C^e + C^i < m) \\
&= G(b)L(a') + L(b') - L(a') - \frac{k'}{\lambda}[1 - G(m)][H'(b') - H'(a')], \\
k' &= \frac{\gamma\lambda}{\gamma - \lambda},\ H'(z) = 1 - \exp(z/k')
\end{aligned}
$$

$$
\begin{aligned}
P_e E[C^e|e] &= P_e\{E[C^e|C^i > b,\, C^e < a']\} \\
&= \int_b^\infty \int_0^{a'} c^e l(c^e) g(c^i) dc^e dc^i = [1 - G(b)][L(a')(a' + \lambda) - a'], \\
P_i E[C^i|i] &= P_i\{E[C^i|C^i < a,\, C^e > b']\} \\
&= \int_{b'}^\infty \int_0^{a} c^i g(c^i) l(c^e) dc^i dc^e = [1 - H(b')][G(a)(a + \gamma) - a], \\
P_{ie} E[C^i + C^e|ie] &= P_{ie}\{E[C^i + C^e|C^i < b,\, C^e < a'] + E[C^i + C^e|a' < C^e < b',\, C^e + C^i < m]\} \\
&= \int_0^{a'} \int_0^{b} c^i g(c^i) l(c^e) dc^i dc^e + \int_0^{b} \int_0^{a'} c^e l(c^e) g(c^i) dc^e dc^i \\
&= \int_{b'}^{a'} \int_0^{m - C^e} c^i g(c^i) l(c^e) dc^i dc^e + \int_{b'}^{a'} \int_0^{m - C^e} c^e g(c^i) l(c^e) dc^i dc^e \\
&= G(b)[L(a')(a' + \lambda) - a'] + L(b')[G(b)(b + \gamma) - b] \\
&+ (\gamma + \lambda)[L(b') - L(a')] + a'[1 - L(a')] - b'[1 - L(b')] \\
&- \frac{(m + \gamma)}{\lambda} k'[1 - G(m)][H'(b') - H'(a')].
\end{aligned}
$$

Under exponential shocks but restricting substitutability or complementarity, the integrated value function is instead:

$$
\begin{aligned}
EV = \Pi &+ [1 - G(p)][1 - H(q)]\delta EV^n + G(p)[1 - H(q)]\delta EV^i \\
&+ [1 - G(p)]H(q)\delta EV^e + G(p)H(q)\delta EV^{ie} \\
&+ p[1 - G(p)] + q[1 - H(q)] - \gamma G(p) - \lambda H(q),
\end{aligned}
$$

where

$$
\begin{aligned}
p &= \delta EV^i - \delta EV^n, \\
q &= \delta EV^e - \delta EV^n.
\end{aligned}
$$

One of the features of the exponential distribution is that its truncation has a nice form that allows one to simplify the expressions for the value functions. For a given set of cost parameters, one can numerically solve for V that is made up of V^n, V^e, V^i, V^{ie}, each is 100 (ω) by 10 (k) grids. With the value functions, one can proceed to construct the joint likelihood function.

1.9.3 Bayesian MCMC Estimation

Rather than assuming θ is fixed and using MLE to obtain their point estimates:

$$p(\theta|y) \propto p(y|\theta),$$

I use the Bayesian approach; treat θ as a random variable to maximize the posterior distribution instead:

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)} \propto p(y|\theta)p(\theta).$$

With a diffuse prior for θ, I update with the data to form posterior beliefs about the parameters using the Metropolis-Hastings algorithm.

The basic idea of the M-H algorithm is to take simulated draws that are slightly dependent and are approximately from a posterior distribution. The hope is when the chain runs long enough it converges to the stationary distribution. I implement the algorithm as follows:

1. Diffuse priors: $P(\theta) \sim N(0, 1000)$ or $N(0, 100)$; 30 sets of random starting values θ_0.

2. For simulations $i = 1$ to 10000, draw a candidate θ_* from $J(\theta_i|\theta_{i-1})$ and accept with probability $r = \min\left(\frac{q(\theta_*)L(X|\theta_*)}{q(\theta_{t-1})L(X|\theta_{i-1})}, 1\right)$. If not accepted, set θ_i as θ_{i-1}.

3. The jumping distribution is a random walk and symmetric: $J(\theta_i|\theta_{i-1}) = \theta_{i-1} + rand_i * step$, where $rand_i \sim N(0, 1)$ and $step$ is the parameter-specific step size.

1.9.4 Alternative Model Specifictaion

Instead of assuming independent cost draws, one can estimate the cost variables as fixed parameters and have the cost shocks as correlated draws from a bivariate normal distribution $\Phi(\varepsilon^i, \varepsilon^e)$ with mean zero, correlation ρ and the covariance matrix:

$$\Sigma = \begin{bmatrix} \sigma_i^2 & \rho\sigma_i\sigma_e \\ \rho\sigma_i\sigma_e & \sigma_e^2 \end{bmatrix}$$

The new Bellman equation (counter part to (1.8)) is then

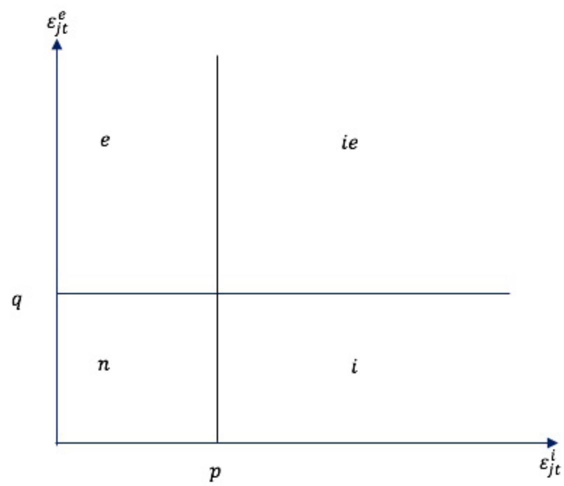
$$\bar{V}(s_{jt}) = \int \left(\Pi(s_{jt}) + \max_{i_{jt}, e_{jt}} \left\{ \begin{matrix} \delta E\bar{V}^n, \\ -C^e + \varepsilon^e + \delta E\bar{V}^e, \\ -C^i + \varepsilon^i + \delta E\bar{V}^i, \\ -C^i - C^e + \varepsilon^e + \varepsilon^i + \delta E\bar{V}^{ie} \end{matrix} \right\} \right) d\Phi(\epsilon), \tag{1.27}$$

and the integrated value function is:

$$\begin{aligned} EV = \Pi + P_n\delta EV^n + P_e(\delta EV^e - C^e + E\left[\varepsilon^e|e\right]) \\ + P_i(\delta EV^i - C^i + E\left[\varepsilon^i|i\right]) + P_{ie}(\delta EV^{ie} - C^i - C^e + E\left[\varepsilon^i + \varepsilon^e|ie\right]), \end{aligned}$$

It is important to note that ε^i, ε^e also depend on a firm's previous period status, so each is a vector of two variables, one for start up cost shock and the other for maintenance. This implies that the there are a total of four different variances as well as correlations in total.

Under this specification, I have restricted the the interaction term between internal and external R&D in the Markov process (α_{ie}) to be zero, and instead let the correlation coefficient in the cost shocks pick up correlations between the cost draws that could alternatively be interpreted as economies of scope. The updated figure as a result of the restriction of $\alpha_{ie} = 0$ is now:



where

$$p = \delta E\bar{V}^n - (\delta E\bar{V}^i - C^i)$$

$$q = \delta E\bar{V}^n - (\delta E\bar{V}^e - C^e)$$

This specification is attractive since one can take advantage of the Cholesky decomposition to uncorrelate the cost draws. This makes the problem tractable as the probabilities and the conditional expectations will be much simpler as a result. More specifically, call the transformed variable $\epsilon = C\varepsilon$, one needs a matrix C such that the transformed ε has mean 0 and variance I_2, which means the candidate C must be such that $C\Sigma C = I_2$. The Cholesky decomposition of Σ is $\Sigma = U'U$:

$$\begin{pmatrix} \sigma_i & 0 \\ \rho\sigma_e & \sigma_e(1-\rho^2)^{1/2} \end{pmatrix}\begin{pmatrix} \sigma_i & \rho\sigma_e \\ 0 & \sigma_e(1-\rho^2)^{1/2} \end{pmatrix}$$

And the natural candidate for C is U'^{-1}.

To proceed with the estimation, I first pre-multiply the cutoffs p, q with U'^{-1} so that $(\varepsilon^i, \varepsilon^e)$ are now independent standard normal bivariate variables. I evaluate the conditional expectations correspondingly and then rescale the expectation by pre-multiplying it with U'. One can then proceed to solve for the value function by iteration, and then estimate the parameters using Baysian MCMC.

In the end, the set of parameters under this alternative model consists of the four costs and the variance of each of the costs (Panel A of Table 1.9), and the correlation between them (Panel B of Table 1.9).

Table 1.9: Dynamic Parameter Estimates under Alternative Specification

parameters		(1) mean (mil.RMB)	(2) SE
Internal	γ^F	655.1	(0.792)
	γ^f	1.67	(0.005)
	$\sigma(\varepsilon_{\gamma^F})$	6.69	
	$\sigma(\varepsilon_{\gamma^f})$	7.94	
External	λ^F	71.7	(0.718)
	λ^f	5.46	(0.094)
	$\sigma(\varepsilon_{\lambda^F})$	20.62	
	$\sigma(\varepsilon_{\lambda^f})$	19.58	
Correlation			
		ε_{λ^F}	ε_{λ^f}
	ε_{γ^F}	-0.2336	0.0065
	ε_{γ^f}	0.1183	-0.0501

In 2002, 1USD≈8RMB. SE in parentheses.

The estimates for internal and external startup and continuation costs are close to those under the baseline model, but the shocks associated with external R&D activities have much higher variance. I next examine the correlations among the cost shocks. There is a negative correlation between internal and external startup costs, this supports the observation in the data that firms rarely start both types of research at once. The correlations between internal continuation (startup) and external startup (continuation) costs are positive, but close to zero for the latter combination. This may serve as preliminary evidence that firms with an in-house team may have an incentive to start external R&D and vice versa. Lastly, the correlation between internal and external continuation costs is negative but close to zero. This is in line with the empirical finding that the two activities serve as substitutes for firms that are active in research.

1.9.5 Robustness Checks

1.9.5.1 Pecking Order by R&D Mode

Table 1.10: The Pecking Order by R&D Mode is Robust

	lnsales	lnK	Productivity	New Proc	New Prod	Patents
$\mathbb{1}(e_{jt-2} = 1)$	0.641***	0.423**	0.360**	0.562***	0.626***	0.445**
	(0.184)	(0.201)	(0.150)	(0.144)	(0.145)	(0.181)
$\mathbb{1}(i_{jt-2} = 1)$	1.092***	0.943***	0.717***	0.561***	0.744***	0.520***
	(0.128)	(0.131)	(0.0968)	(0.0931)	(0.0943)	(0.119)
$\mathbb{1}(ie_{jt-2} = 1)$	1.608***	1.517***	0.977***	1.207***	1.457***	0.987***
	(0.166)	(0.169)	(0.124)	(0.116)	(0.125)	(0.130)
Constant	7.072***	13.48***	7.034***	-0.925***	-0.896***	-1.894***
	(0.321)	(0.298)	(0.222)	(0.233)	(0.227)	(0.309)
N	1266	1266	1266	1260	1260	1266
R^2(Pseudo R^2)	0.382	0.353	0.320	(0.127)	(0.194)	(0.130)

Coefficients on year, industry, city, exporter and state-owned status omitted. Robust standard error clustered by firms. Depvars: Productivity is Levihnson & Petrin type of estimator with an exogenous first order Markov process; New Proc, New Prod, and Patents are indicator variables for whether firms introduced new process, product or held patents.In all regressions R&D modes are jointly significant; in the last three columns e_{jt-2} is not statistically different from i_{jt-2}. This last finding suggests both R&D structures contribute significantly to innovation success, regardless of whether it is a new product or a new process.

1.9.5.2 Alternative Estimation Strategy for the Productivity Process

One can alternatively estimate the productivity process in a one-step procedure. This is possible with either a nonparametric or parametric control function for ω_{jt}. In what follows, I adopt the strategy under the assumption that the production function is Cobb-Douglas.

When the production function is Cobb-Douglas, the exact expression for ω_{jt}, or the inverted $h^{-1}(\cdot)$ is:[34]

$$\omega_{jt} \equiv h^{-1}(k_{jt}, m_{jt}) \equiv \kappa' + \delta_t' - \beta_k k_{jt} + m_{jt}. \tag{1.28}$$

Lagging (1.28) one period and substituting it for the Markov process for ω_{jt-1} as specified in (1.6) gives:

$$\omega_{jt} = \alpha_0 + \alpha_1(\kappa' + \delta_t' - \beta_k k_{jt} + m_{jt}) + \alpha_i i_{jt-1} + \alpha_e e_{jt-1} + \alpha_{ie} i_{jt-1} * e_{jt-1} + \xi_{jt} \tag{1.29}$$

The updated revenue is now[35]

$$\begin{aligned} r_{jt} = {} & \text{const.} + \beta_t t + (\eta - 1)\beta_k k_{jt} + (\eta - 1)\alpha_1 \left[\beta_k k_{jt-1} + m_{jt-1}\right] \\ & + (\eta - 1)\alpha_i i_{jt-1} + (\eta - 1)\alpha_e e_{jt-1} + (\eta - 1)\alpha_{ie} i_{jt-1} * e_{jt-1} + (\eta - 1)\xi_{jt}, \end{aligned} \tag{1.30}$$

and the parameters in the Markov process can be estimated in one-step using NLS.

The results are summarized in 1.11. For comparison purposes, I also include the results from the two-step estimation procedure in Column (1). The qualitative feature of the estimates are preserved compared with the case of two-step estimation procedure. Internal R&D is more effective than external R&D and the interaction of the two modes is not significant. However, the persistence parameter (α_1) is reduced significantly. Notice also that the constant term in the Markov process cannot be separately estimated from that in the revenue function under the one-step procedure.

---

[34] $\kappa' = \frac{\eta}{\eta-1} \ln \frac{\eta}{\eta-1} - \beta_l \ln \beta_l + \frac{\beta_l(\eta-1)-\eta}{\eta-1} \ln \beta_m$, $\delta_t' = -\frac{1}{\eta-1} \ln \Phi_t + \beta_l w_t - \frac{(\eta-1)\beta_l - \eta}{\eta-1} p_t^m$.

[35] $\kappa = -(\eta - 1) \ln \frac{\eta}{\eta-1} + (\eta - 1)\beta_l \ln \beta_l + (\eta - 1)\beta_m \ln \beta_m$, $\delta_t = \ln \Phi_t - (\eta - 1)\beta_l w_t - (\eta - 1)\beta_m p_t^m$.

Table 1.11: Estimation Results of the Markov Process for Productivity

	Baseline	(2)	(3)
η			4.871**
			(0.143)
β_k	0.070***	0.051***	
	(0.015)	(0.004)	
α_0	0.120***		
	(0.018)		
α_1	0.945***	0.168***	
	(0.011)	(0.004)	
α_e	0.021**	0.064***	
	(0.10)	(0.019)	
α_i	0.026***	0.072***	
	(0.007)	(0.012)	
α_{ie}	-0.008	-0.019	
	(0.516)	(0.023)	
Constant	0.120***	2.649***	
	(0.018)	(0.137)	
Obs	2531	2531	2406

Bootstrapped standard errors in parametheses in Columns (1) & (2).
Column (3) contains the estimation results of the profit function.

1.9.5.3 Static and Dynamic Estimates with Controls for Heterogeneity

Table 1.12: Dynamic Parameter Estimates with Controls for Heterogeneity

parameters		(1) distribution mean (mil.RMB)	(2) realized mean (mil. RMB)	(3) $\sum \frac{R}{\sum R} \frac{C}{R}$	(4) $\sum \frac{\Pi}{\sum \Pi} \frac{C}{\Pi}$
Full Model					
Internal	γ^F	334.3	22.9	0.17	1.37
		(31.986)	(4.796)		
	γ^f	1.62	1.59	0.010	0.066
		(0.081)	(0.066)		
External	λ^F	142.46	16.8	0.079	0.428
		(21.549)	(2.440)		
	λ^f	3.98	2.44	0.018	0.112
		(0.085)	(0.206)		
Model without R&D Organization					
Startup Cost		142.46	16.8	0.079	0.428
Continuation Cost		3.98	2.44	0.018	0.112

In 2002, 1USD≈8RMB. SE in brackets.

1.9.6 Additional Analysis

Table 1.13: Transition and Aggregate Productivity from 50% Reductions in R&D Costs, Computed from 100 Simulations

	Reduction in Startup Cost			Reduction in Cont. Cost		
	(1) RD_t	(2) $\bar{\omega}$	(3) $\bar{k}$	(4) RD_t	(5) $\bar{\omega}$	(6) $\bar{k}$
n_{t-1}	+0.040	2.690	16.4	+0.026	2.634	16
RD_{t-1}	-0.004	-	-	+0.028	-	-
N (%)	+0.030	-	-	+0.050	-	-

Chapter 2

Innovation and Prices *(with Jordi Jaumandreu)*

2.1 Introduction

This paper investigates the impact of process and product innovations on the output prices set by firms. It uses prices reported over an extended period of time by a sample of manufacturing firms. We assume prices are set with reference to (short run) marginal cost and hence they have to reflect variations in efficiency of production, both across firms and over time. Productivity, in turn, shifts with the introduction of process and product innovations by firms.

Models almost universally assume firms set prices with reference to marginal cost, although also recognize that optimal markups over marginal costs can vary with a host of factors such as the type and intensity of competition, firm behavior, the state of the market, and dynamic considerations.[1] Marginal cost, both in the short and long run, depends on firm-level productivity, and hence evolves according to the firm's innovation activities and, in particular, the introduction of process and product innovations. This level of productivity is usually considered heterogeneous and unobserved, something to be measured together with the impact that process and product innovations have on it.[2] Process innovations, aimed at reducing cost, are expected to shift the marginal cost function downwards. The introduction of product innovations (improved or new goods) are expected to change cost in more heterogeneous ways. Quality upgrades of the goods may sometimes

[1]This is a consequence of profit maximization, which implies that marginal revenue should equal marginal cost. Hall and Hitch (1939) is a departure from this view, in what Ellison (2006) characterizes as an early contribution to behavioral industrial organization.

[2]The literature on estimating heterogeneous unobserved productivity with structural methods starts with Olley and Pakes (1996); see also Levinsohn and Petrin (2003b) and Ackerberg et al. (2006). Latest models allow productivity to be endogenously determined. Peters et al. (2013), for example, model firms' marginal cost as depending on process and product innovation in the way that we are going to adopt in this paper. Aw et al. (2011) and Doraszelski and Jaumandreu (2013) are papers in which endogenous efficiency depends on R&D expenditure.

imply less unobserved productivity, at least temporarily.[3,4]

The final impact of innovation on prices is then something that depends on two very different mechanisms. The first is how innovations affect marginal cost as described above. The second is how firms translate the variations in marginal costs to prices, and specifically how the changes in cost due to innovations are passed onto prices. Although theory offers some guidance on how these processes are expected to work, it is something mainly to be investigated empirically.

Without changes in prices, markups would be continuously eroded by the increase in costs and subjected to more discrete and heterogeneous changes due to the impact of process and product innovations on costs. With prices fixed we expect process innovations to enlarge margins and product innovations to either increase or decrease them; there two additional important mechanisms by which innovation can affect margins.

On the one hand, innovations are likely to enlarge the demand for the product of the firm and maybe modify the price sensitivity of the clients with respect to the good. This seems particularly likely for product innovations, but it could be also true for process innovations.

On the other hand, innovation induced changes on demand in principle give firms incentives to continuously adapt prices to their profit maximizing levels. But a huge literature on price changes has given many reasons for which firm-level prices may show inertia[5] and shown it to be empirically relevant.[6] This literature has considered fixed and variable costs of price adjustment and

[3]For example, the new production may require new labor skills and material qualities not fully accounted for in the observable part of cost. And, when “learning by doing” is important, firms may lack experience in producing the new good and unobserved productivity goes down as a result.

[4]In addition, as most firms are in practice multiproduct, product innovations often only modify the product mix and their impact on cost are expected to be generally milder.

[5]Economists have discussed the rigidity of prices and its consequences at least since Berle and Means (1932). Models of adjustment costs in prices at the firm level were advanced, for example, in the works of Barro (1972) and Sheshinski and Weiss (1977, 1992) with lumpy costs generating inaction, and in Rotemberg (1982) with strictly convex costs generating partial adjustment. Carlton (1985) discusses the issue of price rigidity from the point of view of industrial organization in the first IO Handbook. A recent contribution to the theory of rigid prices (in the presence of collusion) is Athey and Bagwell (2008).

[6]Early works described the micro price-setting of particular industries and cases, e.g. Cechetti (1986), Carlton (1986), Slade (1991), Lach and Tsiddon (1992) and Kashyap (1995). Slade (1998) and Aguirregabiria (1999) are full structural dynamic models that estimate firm-level fixed and variable adjustment costs. Newer evidence has incorporated interindustry studies and surveys on price setting. Reviews of the extensive accumulated evidence can be found in Álvarez et al. (2006) for Europe and Klenow and Malin (2010) for the US. Two recent studies using respectively industry and firm-level data are Goldberg and Hellerstein (2011) and Eichenbaum et al. (2011). The latest is part of a huge literature that combines detailed micro-evidence with the macro discussion of the neutrality of money. A recent contribution of this kind is Midrigan (2011).

developed models for both discrete and level changes in different situations. The costs of changing prices could refer to menu costs, costs of convincing customers the necessity of the new prices, and perhaps dynamic considerations on future demand.

No literature, at least not to our knowledge, has explored the role that process and product innovation may play on the costs of adjusting prices but it is clear that it can be very important. When process innovations reduce marginal costs they also reduce the cost of updating the margin, because it can be redressed without changing price or changing it by a smaller amount. Implementing product innovations may imply important differences in the cost of changing prices (because, for example, the different client's information requirements when the product is new). In other words, process and product innovations become state variables of the dynamic problem of prices.

All the above makes a very strong simplification to consider that margins are constant, as they were in a permanent long-run equilibrium. But still more important is to recognize that ignoring the markup changes due to innovations is likely to induce an important bias in the estimated impacts of process and product innovations on prices and costs and other variables that involve price or price and cost. Intuitively, we may be leaving in the residuals changes in the margins that are correlated with process and product innovations. For example, this paper finds that when process innovations decrease costs there is an incomplete pass through to prices, and firms enlarge their margins as a result. This may explain the insignificant impact of process innovation on prices in an OLS framework, and can also explain the weak effects that researchers get for the impact of process innovation on "revenue" productivity.[7]

The model that we develop in this paper takes advantage of the fact that we (imperfectly) observe both margins and prices.[8] It specifies the (log) margin over marginal cost as the result of a dynamic decision of the firm. We show that, given the rest of relevant variables, choosing an optimal price is the same as choosing an optimal markup. So in the model the firm directly decides

[7] In principle longer firm-level series should be able to deal satisfactorily with these problems, but typical panel data sets are still just characterized by few time observations for each individual.

[8] On the one hand, we have firm-level price indices, on the other we estimate the markup as the scale parameter times the the price-average cost margin: $\nu\frac{R}{C} = \nu\frac{P}{AC}$, where R is revenue, C are variable costs, P is price and AC is average cost. This is an alternative to estimate markup in the older IO literature recently revisited by De Loecker and Warzynski (2012).

if it wants to update the value of the margin and by how much. To do this, the firm weighs the profits against the costs of changing the margin.[9] The cost of adjusting the margin is specified as a function of its level, the state of the demand for the product, whether the firm has introduced a process or product innovation and other state variables. The implication of the model for reduced-form estimation is that the margin is likely to follow a first order Markov process that also takes into consideration the roles of all the state variables, observed and unobserved.

Then we nest this margin in a structural equation that gives account of (log) price as the result of both the evolving markup and the marginal cost of the firm. Marginal cost is modeled as the result of observable variables and unobserved productivity, the latter specified as an endogenous Markov process which depends on the introduction of process and product innovations. We estimate the model as a nonlinear simultaneous system with two dependent variables (margin and price) and cross-equation restrictions on the parameters. We estimate two versions of the model. In one, there are no adjustment costs of prices and the firm sets the markup according to the elasticity of demand (as impacted by cyclical changes and innovations). In the other, the markups follow the law of motion derived from the dynamic model. In this version we only have implicit demand "equivalent" elasticities, or elasticities that would determine the same markup that we see under optimal dynamic pricing.

Our basic model considers the short-run marginal cost proportional to average variable cost (that is, the short-run parameter of elasticity of scale is constant) and hence takes the choice of the firm as the choice of its price-average cost margin up to a constant. It can be argued that the observed price-average cost margin ignores some relevant components of marginal cost, such as the adjustment costs of labor. Conversely, it could be argued that the relevant marginal cost for the firm to set the markup over should not include short-run cyclical components.[10] One important advantage of our approach is that we can test for the relevant marginal cost by including in our

[9] The costs of changing the margin are the costs of changing the price when this change is needed plus some specific costs. For example, an increased margin may augment the cost of dealing with the retributions of profits among managers and workers.

[10] It has been argued that firm set prices with an optimal markup on "reference" marginal costs, whose time variation is less frequent than the very short-run marginal costs. E.g., Eichenbaum et al. (2011) provides empirical evidence for this price setting behavior of a major US retailer.

specifications the relevant corrections to the observed price-average cost margin. Our conclusion is that labor adjustment costs are a relevant part of marginal cost.

There are at least three reasons for the interest in the details of the impact of innovation on prices. First of all, how technological change via firm-level innovation transmits to prices is relevant for understanding the working of many economic forces. For example, prices play a crucial role in the allocation of resources and thus growth. When a process innovation induces a price decrease, market demand for the good is expected to increase. As a result, the demand for inputs required for production is expected to increase by more than what is needed to compensate for any input displacement provoked by technology. However, the magnitude and how complete, rapid and heterogeneous is this mechanism is something to be assessed empirically.

Second, the analysis of what is in firm-level prices is important for addressing many key questions of productivity measurement. The quantities used in productivity analysis are, almost without exception, obtained by deflating firm revenue by a price index. When an industry-wide index is used, it has been argued that it leaves out the difference between the firm-level and the general indices and one may get biased estimates as a result.[11] In fact, revenue conditional on inputs and their prices may not contain any information on productivity.[12] This clearly contrasts with the host of estimates of productivity that find productivity growth using revenue deflated by a non firm-level price index. An explanation for these results is that productivity is not instantaneously passed onto prices. When this is the case, revenue conditional on inputs and their prices is no longer invariant to productivity.[13]

Lastly, how well prices reflect productivity improvements is important for the measurement of welfare. On the one hand, if prices do not accurately reflect technical efficiency improvements then the analysis of the increases in consumer surplus coming from innovation becomes more

[11] This problem was firstly pointed out by Klette and Griliches (1996). Later, it has been at least addressed by **?**, Loecker (2011) and Katayama et al. (2009), who propose different ways to deal with it.

[12] Let's consider for simplicity the case of constant returns to scale. Production function is $Q = F(x)e^{\omega}$, where ω stands for idiosyncratic productivity and F shows CRTS. Cost is $C = c(w)Qe^{-\omega}$, where w represents prices, and marginal cost $MC = c(w)e^{-\omega}$.Under a markup $\frac{\eta}{\eta-1}$, revenue is $R = PQ = \frac{\eta}{\eta-1}c(w)e^{-\omega}F(x)e^{\omega} = \frac{\eta}{\eta-1}c(w)F(x)$, so conditional on inputs and their prices it does not depend on productivity.

[13] Suppose that at a given moment price is $P = \frac{\eta}{\eta-1}c(w)e^{-\theta\omega}$, where $\theta < 1$ reflects the partial transmission of productivity to prices. Now revenue is $R = \frac{\eta}{\eta-1}c(w)F(x)e^{(1-\theta)\omega}$.

difficult. By the same token we can have prices that do not properly reflect product novelty. There is an ongoing literature that studies how the methods applied by Statistical Offices with respect to new goods may impact welfare measurements.[14] One relevant dimension of this problem is to understand how firms price product improvements in practice.

To estimate our model we use the prices reported over an extended period of time (17 years) by a sample of Spanish manufacturing firms. We use data on ten (unbalanced panel) industry samples, which in total amount to more than 2,300 manufacturing firms and 17,000 observations, corresponding to the period 1990-2006. We have firm-level price indices for each firm, constructed from the reported yearly output price increases in the main market, we observe the timing of the process and product innovations introduced by firms, and we have the relevant information to construct margins, output and input use.

Preliminary GMM estimates show a good fit of the most general model. Margins show inertia but tend to change procyclically with the state of the market. Importantly, firms tend to take advantage of process innovations to enlarge margins by not passing onto prices the decrease in cost. Product innovations, instead, do not affect margins. Our model cannot detect if this is the effect of product innovations not affecting the elasticity of demand or the firm's reluctance to modify margins when new products are introduced. On the other hand, process innovations unambiguously increase productivity and thus decrease cost. Product innovations are on average associated with productivity increases but by amounts that are less than those associated with process innovations. Consequently, the reductions in cost as a result are not as big and, for some industries, product innovations increase cost. Compared with the simplest model, our dynamic specification gives sharper cost reducing effects of process innovations and more complex effects of product innovations.

The paper is organized as follows. In Section 2, we lay out the econometric model under static and dynamic pricing. In Section 3, we focus on the method of estimation for the model of dynamic pricing. Section 4 discusses the data and Section 5 presents the estimation results. Finally Section 6 concludes.

[14]See, for example, Pakes (2003).

2.2 Model

In this section, we first lay out the econometric model that gives rise to a general system of equations that describes the structural link of prices, productivity and innovation. The identification of these equations is then discussed, but their exact specifications depend on whether there exists static or dynamic pricing or not. Thus the rest of the section is devoted to deriving the system of estimable equations under either scenario.

2.2.1 The Structural Link of Prices, Productivity and Innovation

We assume that firm j operates in a monopolistically competitive market and sets the price of its product with a markup on (short-run) marginal cost.[15] How large is the firm demand at given price may differ according to exogenous and endogenous shifters that we discuss later in detail.[16] Let's write demand for the moment as $Q_{jt} = Q(P^*_{jt}, Z_{jt})$, where Q_{jt} and P^*_{jt} represent quantity and (unobserved) price and Z_{jt} is a vector of shifters. We assume that the researcher observes price with an error (see below).

The firm knows its cost function. We assume that the firm production function is

$$Q_{jt} = F(K_{jt}, L_{jt}, M_{jt}) \exp(\omega_{jt}),$$

where K_{jt}, L_{jt}, and M_{jt} stand for capital, labor and materials, and ω_{jt} represents the Hicks neutral level of efficiency of firm j. Following the literature we call ω_{jt} productivity, and it is observed by the firm but unobservable for the econometrician. The firm minimizes cost of labor and materials given capital, $W_{jt}L_{jt} + P_{Mjt}M_{jt}$, where W_{jt} and P_{Mjt} are the prices of labor and materials. Variable cost minimization implies the existence of the function $C_{jt} = C(K_{jt}, W_{jt}, P_{Mjt}, Q_{jt}/\exp(\omega_{jt}))$ and

[15] By monopolistic competition we understand that each firm faces a downward-sloping demand for its product, makes no profit in the long run equilibrium, and a price change by one firm has only a negligible effect on the demand of any other firm (Tirole, 1989).

[16] Later we are going to consider the following shifters: an index of industry prices P_{It}, the state of the demand for the product of the firm D_{jt} and the introduction of process and product innovations z_{jt} and d_{jt}.

marginal cost can be written as

$$MC_{jt} = MC(X_{jt}) \exp(-\omega_{jt}), \tag{1}$$

where $X_{jt} = \{K_{jt}, M_{jt}, W_{jt}, P_{Mjt}\}$ is a vector of observable variables (see Appendix A). Cost minimization given capital implies $MC_{jt} = \frac{1}{\varepsilon_{Ljt}+\varepsilon_{Mjt}} AC_{jt}$, where ε_{Ljt} and ε_{Mjt} are the output elasticities of labor and materials. We assume that the short-run parameter of scale $\nu = \varepsilon_{Ljt} + \varepsilon_{Mjt}$ is constant and less than unity.

The firm sets price by choosing the optimal markup over marginal cost $\mu_{jt} = \frac{P^*_{jt}}{MC_{jt}}$. Under our assumptions, profits at time t can be written as a function of the markup given capital, productivity, input prices and shifters:[17]

$$\begin{aligned} \pi_{jt} &= R_{jt} - C_{jt} = \left(\frac{\mu_{jt}}{\nu} - 1\right) C(K_{jt}, W_{jt}, P_{Mjt}, Q(\mu_{jt}, Z_{jt}, K_{jt}, W_{jt}, P_{M_{jt}}, \omega_{jt})/\exp(\omega_{jt})) \\ &= \pi(\mu_{jt}, K_{jt}, W_{jt}, P_{Mjt}, Z_{jt}, \omega jt). \end{aligned}$$

The problem of choosing the optimal price given the rest of the variables can be therefore set, without loss of generality, as a problem of deciding the optimal markup. We will later find this fact very useful for the specification of the dynamic model. We are going to assume that observed price meets the relationship

$$P_{jt} = P^*_{jt} \exp(e_{jt}) = \mu_{jt} MC_{it} \exp(e_{jt}), \tag{2}$$

where e_{jt} is a measurement error orthogonal to all information available when the firm takes the decisions.[18]

[17] The markup $\mu_{jt} = \frac{P^*_{jt}}{MC_{jt}} = \nu \frac{P(Q_{jt}, Z_{jt}) Q_{jt}}{C(K_{jt}, W_{jt}, P_{Mjt}, Q_{jt}/\exp(\omega_{jt}))}$ is a monotonic function of Q_{jt} given the rest of variables, and the inverse function $Q_{it} = Q(\mu_{jt}, Z_{jt}, K_{jt}, W_{jt}, P_{M_{jt}}, \omega_{jt})$ exists. Monotonocity holds because the derivative $\frac{\partial \mu_{jt}}{\partial Q_{jt}} = -\frac{\mu_{jt}}{Q_{jt}}(\frac{1}{\eta_{jt}} + \frac{1}{\nu} - 1)$, where η_{jt} stands for the (absolute value of) elasticity of demand, is negative for any positive markup.

[18] An important part of the literature on prices has shown that prices tend to follow a framework that can be seen as "reference" prices responding to "reference" costs. Although the underlying rule is quite inertial, prices tend to vary frequently in the short run. For a recent example see Eichenbaum et al. (2011) on prices from a large US supermarket retailer. In our paper, by means of the disturbance e_{jt} we could extend our model to additionally allow for sundry factors that are unexpected at the time of setting the price. In this extension the demand should be modified to be a function of P_{jt} rather than P^*_{jt}.

Our interest lies in endogenous productivity, or productivity that grows as the result of the firms' innovation efforts.[19] We assume, as in the subsequent literature to **?**, that ω_{jt} follows a first order Markov process and we model productivity as depending on past productivity and the shifts induced by the introduction of process and product innovations.[20] That is,

$$\omega_{jt} = g(\omega_{jt-1}, z_{jt}, d_{jt}) + \xi_{jt}, \tag{3}$$

where z_{jt} and d_{jt} are indicators of the process and product innovations introduced by firm j, g is an unknown function aimed at picking up both the path dependence of productivity as well as the unknown impact of the innovations, and ξ_{jt} is a random innovation mean-independent of all the arguments of $g(\cdot)$. Here we specify z_{jt} and d_{jt} to affect productivity contemporaneously but in fact this is an empirical issue and we are going to experiment with these variables lagged as well.

Inserting (3) in (1) and (1) in (2) and taking logs we have our structural model linking prices to innovation:

$$p_{jt} = \ln \mu_{jt} + mc(X_{jt}) - g(\omega_{jt-1}, z_{jt}, d_{jt}) - \xi_{jt} + e_{jt}. \tag{4}$$

This equation shows that we expect prices to structurally evolve according to productivity and innovation. But also that we will only be able to pick up this relationship as long as we adequately control for endogenous markups and the observable part of marginal cost.

Estimation of equation (4) presents two main difficulties from the point of view of estimation: neither lagged productivity ω_{jt-1} nor the markup μ_{jt} are directly observable. Now we detail how we deal with these two problems.

Unobserved productivity

Unobservable lagged productivity ω_{jt-1} is interrelated with the observables z_{jt} and d_{jt}. In addition ω_{jt-1} is likely to be correlated with most of the explanatory variables. In order to estimate the

[19]The endogenous productivity literature models productivity shifting with the R&D activities of firms. See, for example, Doraszelski and Jaumandreu (2013); Aw et al. (2011); Peters et al. (2013) and Boler, Moxnes and Ullveit-Moe (2015).

[20]This is also the specification used in Peters et al. (2013).

model we will employ an Olley and Pakes (1996) type of procedure, replacing the unobservable ω_{jt-1} by an expression in terms of observables. We discuss several alternatives in Appendix B and the implementation below. Let us write in general $\omega_{jt-1} = f(Z_{jt-1})$, noticing that the set Z_{jt-1} changes with the different alternatives. The estimable model can now be written as

$$p_{jt} = \ln \mu_{jt} + mc(X_{jt}) - g(f(Z_{jt-1}), z_{jt}, d_{jt}) - \xi_{jt} + e_{jt}.$$

Markup estimation

On the other hand, we cannot address in general the estimation of the markup inside this equation without incurring serious problems of identification. For example, if d_{jt} and z_{jt} affect the markup then these variables enter two parts of the equation. As a result, the productivity impact of these variables cannot be separated from their role in determining the margin. However, under our assumptions, we have an observed variable that differs from μ_{jt} only by a constant and an uncorrelated error. That is, we know from equation (2) that

$$\ln \frac{R_{jt}}{C_{jt}} \equiv \ln \frac{P_{jt}}{AC_{jt}} = -\ln \nu + \ln \mu_{jt} + e_{jt}. \tag{5}$$

Let's write $\ln \mu_{jt} = h(Y_{jt}) + \varepsilon_{jt}$ where $h(\cdot)$ is the expectation of (log) markup conditional on the set of observed variables Y_{jt}. This implies that we can estimate the system

$$\begin{aligned} r_{jt} - c_{jt} &= -\ln \nu + h(Y_{jt}) + \varepsilon_{jt} + e_{jt} \\ p_{jt} &= h(Y_{jt}) + mc(X_{jt}) - g(f(Z_{jt-1}), z_{jt}, d_{jt}) - \xi_{jt} + \varepsilon_{jt} + e_{jt}, \end{aligned}$$

where r_{jt} and c_{jt} are the logs of revenue and variable cost. We can impose the equality of $h(\cdot)$ across equations.[21] We will later see that set of variables Y_{jt} is going to be a subset of the union of the sets X_{jt} and Z_{jt} plus an exogenous variable describing the state of demand for the products of the firm.

[21]The constant implicit in $h(\cdot)$, however, cannot be estimated separately from any other constant in the equations.

2.2.2 Identification

We have developed the model with a minimum of functional form or distributional assumptions. Is the model identified at this level of generality? To answer this question notice first that the set X_{jt} contains only one endogenous variable, the input quantity M_{jt}, and the set Z_{jt-1} none. M_{jt} is the only variable that is determined once the innovation ξ_{jt} is known. All the rest of the variables are likely to be correlated with productivity but, once the predictable part of productivity has been specified in terms of observables, it only remains unobserved the value of the innovation ξ_{jt}. This innovation is only revealed after the values of these variables have been chosen. This is true for the innovation variables because they are the (partly random) outcome of past R&D investments. And this is true for the input prices under the assumption that they are determined in competitive markets and hence given for the firm.

Do we have instruments for M_{jt}? The answer is yes, but the available instruments depend on the alternative Z_{jt-1} sets. Appendix B identifies two possible alternatives to substitute for ω_{jt-1}: the inverted demand for an input (materials, say) and the inverted production function. If we are replacing ω_{jt-1} by the inverted conditional demand for materials, a natural instrument for M_{jt} is L_{jt-1}. Variable L_{jt-1} is excluded from the inverted conditional demand, so it doesn't appear in the equation, and according to the model is correlated with M_{jt-1} and hence M_{jt}. If we are using the inverted production function, two natural instruments are the lagged input prices W_{jt-1} and P_{Mjt-1}. With this kind of substitution there are no input prices in the expression that replaces ω_{jt-1} and lagged input prices are, however, related to M_{jt}. Since M_{jt} is correlated to M_{jt-1}, it is also correlated with the lagged prices that determine M_{jt-1} . The model is hence identified in very general conditions. In practice we use a particular parametric specification that makes identification straightforward.

2.2.3 Static and Dynamic Pricing

We consider different hypothesis for the choice of μ_{jt}. First, we assume that the choice of the markup is a static problem. Up to our knowledge all empirical exercises on productivity estimation have used explicitly or implicitly this assumption. Then, we consider that the choice of the optimal markup as a dynamic problem due to the presence of adjustment cost of prices.

Static Pricing

If the firm maximizes current profits the solution to the optimal μ_{jt} has the form $\mu_{jt} = \frac{\eta_{jt}}{\eta_{jt}-1}$, where η_{jt} is the elasticity of demand.[22] The model can be written as

$$
\begin{aligned}
r_{jt} - c_{jt} &= -\ln \nu + \ln \frac{\eta(Y_{jt})}{\eta(Y_{jt}) - 1} + \varepsilon_{jt} + e_{jt} \\
p_{jt} &= \ln \frac{\eta(Y_{jt})}{\eta(Y_{jt}) - 1} + mc(X_{jt}) - g(f(Z_{jt-1}), z_{jt}, d_{jt}) - \xi_{jt} + \varepsilon_{jt} + e_{jt}.
\end{aligned} \tag{6}
$$

The firm sets the markup according to the elasticity of demand. This is a model that can be taken to the data by specifying the elasticity as a function of the relevant variables. For example, in the empirical exercise we consider that $Y_{jt} = \{D_{jt}, z_{jt}, d_{jt}\}$, where in addition of the innovation dummies we include the indicator D_{jt} of the state of the firm market.

Dynamic Pricing

There is a large micro and macro literature that considers that setting prices is subject to adjustment costs and hence the resulting prices should be envisaged as the optimal solution of a dynamic problem. Stemming from this tradition we consider costs of adjusting the markup μ_{jt} and specify it as the result of a dynamic optimization problem. The adjustment costs of markup include the costs of adjusting prices but are somewhat different, since the markup can change even if the price does not change. In what follows we derive the law of motion for markup under the implications of dynamic pricing.

Now we need to be specific about the demand shifters Z_{jt}. Demand for the product of the firm depends on its price, prices in the industry, the state of the market, product innovations and (maybe)

[22] $\frac{\partial \pi_{jt}}{\partial \mu_{jt}} = \frac{1}{\nu} C_{jt} + (\frac{\mu_{jt}}{\nu} - 1) \frac{\partial C_{jt}}{\partial Q_{jt}} \frac{\partial Q_{jt}}{\partial \mu_{jt}} = \frac{1}{\nu} C_{jt} - (\frac{\mu_{jt}}{\nu} - 1) \frac{\partial C_{jt}}{\partial Q_{jt}} \frac{Q_{jt}}{\mu_{jt}(\frac{1}{\eta_{jt}} + \frac{1}{\nu} - 1)} = 1 - \frac{(\frac{\mu_{jt}}{\nu} - 1)}{\mu_{jt}(\frac{1}{\eta_{jt}} + \frac{1}{\nu} - 1)} = 0$ implies $\mu_{jt} = \frac{\eta_{jt}}{\eta_{jt}-1}$.

process innovations. Both the state of the market and innovations can shift the demand and/or affect its elasticity with respect to price. Let's write

$$Q_{jt} = Q(P^*_{jt}, P_{It}, D_{jt}, z_{jt}, d_{jt}),$$

where P_{It} is the industry price and D_{jt} is the state of the market.

Capital K_{jt} is a state variable (whose dynamic choice we leave implicit for simplicity) and the prices of the inputs are the industry wide prices $W_{It} = \{W_t, P_{Mt}\}$ subject to a vector of firm level disturbances ζ_{jt}. Cost can hence be written as

$$C(K_{jt}, W_{It}, \zeta_{jt}, Q_{jt}/\exp(\omega_{jt})),$$

where W_{It} represent industry input prices and ζ_{jt} the idiosyncratic shocks in the input prices.

Adjustment costs in markup consist of costs that vary with markup in the previous period μ_{jt-1}, the state of the market D_{jt}, and the introduction of process and product innovations z_{jt-1} and d_{jt-1}. The introduction of a process innovation is likely to reduce cost and thus enlarge the margin, lessening the markup adjustment costs. The introduction of a new product may facilitate or hinder the change of the markup, perhaps depending on consumers' reception of the price change. Let's write the adjustment costs as

$$A(\mu_{jt}, \mu_{jt-1}, D_{jt}, z_{jt}, d_{jt}).$$

Replacing ω_{jt} as before by the law of motion $g(\omega_{jt-1}, z_{jt}, d_{jt}) + \xi_{jt}$ it is easy to see that static profits can then be written as

$$\pi_{jt} = \pi(\mu_{jt}, K_{jt}, P_{It}, W_{It}, \zeta_{jt}, D_{jt}, z_{jt}, d_{jt}, \omega_{jt-1}, \xi_{jt}).$$

Collecting the state variables in the vector $S_{jt} = (\mu_{jt-1}, K_{jt}, P_{It}, W_{It}, \zeta_{jt}, D_{jt}, z_{jt}, d_{jt}, \omega_{jt-1}, \xi_{jt})$

the Bellman equation for the dynamic problem is

$$\begin{aligned} V(S_{jt}) &= \max_{\mu_{jt}}[\pi(\mu_{jt}, K_{jt}, P_{It}, W_{It}, \zeta_{jt}, D_{jt}, z_{jt}, d_{jt}, \omega_{jt-1}, \xi_{jt}) - A(\mu_{jt}, \mu_{jt-1}, D_{jt}, z_{jt}, d_{jt})] \\ &\quad +\beta E_t[V(S_{jt+1})|S_{jt}, \mu_{jt}], \end{aligned}$$

which gives rise to a policy function for the markup that has the form

$$\mu_{jt} = \widetilde{h}_t(\mu_{jt-1}, K_{jt}, D_{jt}, z_{jt}, d_{jt}, \omega_{jt-1}) + \widetilde{\varepsilon}_{jt},$$

where we have collapsed the industry prices in the time subindex and collected together the errors in the disturbance $\widetilde{\varepsilon}_{jt}$.

The latest equation gives us the reduced form for the markup that corresponds to the setting of the price as a dynamic problem. Past markup shows up here because the adjustment costs make it a state variable. The firm chooses when and by how much to change the markup according to the state of the market and the introduction of innovations. Notice that the capital of the firm, which affects the current costs, and the impact of the evolution of future industry output and input prices (represented by the time dependence) as well as lagged productivity are also part of the equation.

Recall from equation (5) that $r_{jt-1} - c_{jt-1} + \ln \nu = \ln \mu_{jt-1} + e_{jt-1}$. Integrating the expectation function over the lagged unobservable e_{jt-1}, and replacing ω_{jt-1} by $f(Z_{jt-1})$ the system of equations becomes:[23]

$$\begin{aligned} r_{jt} - c_{jt} &= -\ln \nu + h_t(r_{jt-1} - c_{jt-1} + \ln \nu, K_{jt}, D_{jt}, z_{jt}, d_{jt}, f(Z_{jt-1})) + \varepsilon_{jt} + e_{jt} \\ p_{jt} &= h_t(r_{jt-1} - c_{jt-1} + \ln \nu, K_{jt}, D_{jt}, z_{jt}, d_{jt}, f(Z_{jt-1})) \\ &\quad + mc(X_{jt}) - g(f(Z_{jt-1}), z_{jt}, d_{jt}) - \xi_{jt} + \varepsilon_{jt} + e_{jt}. \end{aligned} \tag{7}$$

The systems of equations (6) and (7) only differ in the $h(\cdot)$ function that we are using to estimate the conditional expectation of the margin. In the first case we are estimating the current elasticity

[23] We slightly abuse the notation by writing the function $h_t(\cdot)$ and the error ε_{jt} using the same symbols as those in equation (6).

of demand, in the second case the elasticity of demand is not the relevant pricing rule although we will be getting implicitly an "equivalent" elasticity derived from the firm choice.

Here the first equation of the system doesn't introduce any particular problem, because all variables are by assumption uncorrelated with the composite error $\varepsilon_{jt} + e_{jt}$. The presence of the unobservable ω_{jt-1}, replaced by $f(Z_{jt-1})$, implies the same restrictions on usable instruments that we have previously discussed in detail. As the function $h_t(\cdot)$ shows up in the second equation we well have to add any extra instrument used in the first equation to the set used in the second.

The marginal cost relevant for the firm may additionally include the adjustment costs of labor. These costs are only partially included in the observed average cost (the part that has an accounting counterpart, as the overtime done by the workers). A dynamic model where the costs of adjusting labor are explicitly considered gives rise to expressions for these adjustment costs based on observables (Doraszelski and Jaumandreu, 2013, 2015, see). Here we develop a correction for the average cost when there are labor adjustment costs that is based on the ratio of materials bill to wage bill (see Appendix C). Assuming that one part of the relevant costs is already included in the observed variable costs we have $MC_{jt} = \frac{AC^*_{jt}}{\nu}(1 + s_{Ljt}\Delta_{jt})^{1-\theta}(1 + s_{Ljt}\Delta_{jt})^{\theta} = \frac{AC_{jt}}{\nu}(1 + s_{Ljt}\Delta_{jt})^{\theta}$, where AC^*_{jt} are latent average costs and AC_{jt} observed average costs. According to this correction, the ratio of revenue over variable costs should be written as

$$r_{jt} - c_{jt} = \theta \ln(1 + s_{Ljt}\Delta_{jt}) - \ln \nu + \ln \mu_{jt} + e_{jt},$$

where s_{Ljt} is the share of labor in variable costs and Δ_{jt} is the the gap between the wage and the shadow price of labor under adjustment costs. When applying this correction, everything is as our variable $r_{jt} - c_{jt}$ becomes $r_{jt} - c_{jt} - \theta \ln(1 + w_{Ljt}\Delta_{jt})$, where θ is a parameter to be estimated. We estimate the system with and without the adjustment cost correction, and the results favor the specification with adjustment costs.

In what follows, we detail the estimation of the static and dynamic pricing model under a particular parametric specification for the production function. This specification makes the model particularly simple and the identification straightforward.

2.3 Estimation

To take equations in (6) and (7) to the data, three components remain to be specified: the observable component of marginal cost, the Markov process that governs the endogenous productivity process and the markup.

Let us start by the marginal cost function. We consider for simplicity the Cobb-Douglas production function

$$q_{jt} = \beta_0 + \beta_k k_{jt} + \beta_l l_{jt} + \beta_m m_{jt} + \omega_{jt},$$

so the short-run marginal cost function in terms of materials is

$$mc_{jt} = \kappa - \beta_k k_{jt} + (1 - \beta_l - \beta m) m_{jt} + (1 - \beta_l) p_{Mjt} + \beta_l w_{jt} - \omega_{jt}, \tag{8}$$

where $\kappa = -\beta_0 - \ln(\beta_l + \beta_m) - \beta_l \ln \beta_l + \beta_l \ln \beta_m$.

With regards to the productivity process we are going to use the inhomogeneous Markov process $\omega_{jt} = \beta_t + g(\omega_{jt-1}, z_{jt}, d_{jt}) + \xi_{jt}$. Our preferred alternative is using the lagged inverted production function to substitute for ω_{jt-1}. Let us use f_{jt-1} as a shorthand for $q_{jt-1} - \beta_k k_{jt-1} - \beta_l l_{jt} - \beta_m m_{jt-1}$. Specifyng $g(\cdot)$ as a complete polynomial of order three in its arguments we have:[24]

$$\begin{aligned}\omega_{jt} = \beta_t + g(f_{jt-1}, z_{jt}, d_{jt}) + \xi_{jt} &= \beta_t + \gamma_1 f_{jt-1} + \gamma_2 f_{jt-1}^2 + \gamma_3 f_{jt-1}^3 + \gamma_4 z_{jt} + \gamma_5 d_{jt} \\ &+ \gamma_6 f_{jt-1} \cdot z_{jt} + \gamma_7 f_{jt-1}^2 \cdot z_{jt} \\ &+ \gamma_8 f_{jt-1} \cdot d_{jt} + \gamma_9 f_{jt-1}^2 \cdot d_{jt} \\ &+ \gamma_{10} f_{jt-1} \cdot z_{jt} \cdot d_{jt} + \gamma_{11} f_{jt-1}^2 \cdot z_{jt} \cdot d_{jt} + \xi_{jt}.\end{aligned} \tag{9}$$

It remains to specify the function $h(\cdot)$. In the static case we use

$$h(\cdot) = \ln \frac{\eta_{jt}}{\eta_{jt} - 1} = \ln \frac{1 + \exp(y_{jt}\lambda)}{\exp(y_{jt}\lambda)} = \ln(1 + \exp(y_{jt}\lambda)) - y_{jt}\lambda, \tag{10}$$

[24] We implicitly collapse the constant of the unknown function in the constant of the equation.

where $y_{jt} = \{md_{jt}, d_{jt}, z_{jt}\}$. The function restricts elasticity to be greater than one while allows for its cyclical fluctuation according to the firm-level state of the market indicator md_{jt} or *market dynamism.*

In the dynamic case we specify the law of motion for markup as a function of a polynomial in the the lagged markup, capital, the firm-level state of the market indicator md_{jt} or *market dynamism,* product and process innovations, and lagged productivity replaced by f_{jt-1}:[25]

$$h_t(\cdot) = \beta_t' + \lambda_1(r_{jt-1} - c_{jt-1}) + \lambda_2(r_{jt-1} - c_{jt-1})^2 + \lambda_3(r_{jt-1} - c_{jt-1})^3$$
$$+ \lambda_4 k_{jt} + \lambda_5 md_{jt} + \lambda_6 z_{jt} + \lambda_7 d_{jt} + \lambda_8 f_{jt-1}. \qquad (10')$$

Lastly, recall that the relevant markup $(r_{jt} - c_{jt})$ may change depending on whether we additionally consider there are adjustment costs in labor or not

In the estimations that we present below we plug expressions (8), (9) and (10) into (6) for the case of static pricing, and expressions (8), (9) and (10′) into (7) for the case of dynamic pricing to estimate each system of equations by two-step nonlinear GMM. Write the residuals $\nu_{1jt} = \varepsilon_{jt} + e_{jt}$ and $\nu_{2jt} = -\xi_{jt} + \varepsilon_{jt} + e_{jt}$ as a function of the variables x_{j_t} and the vector θ of parameters to estimate. Stacked for each firm j, the GMM problem is

$$\min_{\theta} \begin{bmatrix} \frac{1}{N}{}_j A_1(z_j)\nu_{1j}(x_j, \theta,) \\ \frac{1}{N}{}_j A_2(z_j)\nu_{2j}(x_j, \theta) \end{bmatrix}' \widehat{W} \begin{bmatrix} \frac{1}{N}{}_j A_1(z_j)\nu_{1j}(x_j, \theta) \\ \frac{1}{N}{}_j A_2(z_j)\nu_{2j}(x_j, \theta) \end{bmatrix}$$

where $A_1(\cdot)$ is an $L_1 \times T_j$ and $A_2(\cdot)$ an $L_2 \times T_j$ matrix of functions of exogenous variables z_j (a vector partially overlapped with x_j); $\nu_{1j}(\cdot)$ and $\nu_{2j}(\cdot)$ are the $T_j \times 1$ vectors of residuals, and N is the number of firms. $L = L_1 + L_2$ denotes the total number of moments that we use and T_j the number of observations for firm j. For the first step we use the consistent weighting matrix

[25] This particular specification legitimates to get rid of the constant $\ln \nu$.

$$\widehat{W} = \begin{bmatrix} (\frac{1}{N}{}_j A_1(z_j)A_1(z_j)')^{-1} & 0 \\ 0 & (\frac{1}{N}{}_j A_2(z_j)A_2(z_j)')^{-1} \end{bmatrix},$$

and for the second the optimal weighting matrix. Lastly, notice that the subscript on $A(\cdot)$ implies that we use different set of instruments for each of the two equations in the system. We have estimated the equations under the dynamic pricing model and will now talk about the instruments we used.

There are two equations in the system specified in (7). The set of exogenous variables for the first equation includes: constant, time dummies (15), a polynomial of order three in lagged markup, capacity utilization rate and a polynomial of order three in its lag, current and lagged prices of capital, lagged variable input prices (price of materials and wage), lagged inputs (capital, labor and materials) and the two dummies of lagged innovation.

The set of exogenous variables for the second equation includes those used for the first equation and the following: a polynomial of order three in lagged inputs (capital, labor and materials), the interaction of the two innovation dummies as well as their product with each one of the lagged inputs.

Contemporaneous capital is a legitimate instrument. However, we avoid it because of potential measurement error. For the same reason we have also omitted some of the polynomial terms in lagged inputs and lagged markup for some industries. In the end, for the system of equations we end up with a minimum of 70 instruments that expands to an upwards of 75 for certain industries.

We have to estimate 4 parameters that enter nonlinearly: $\beta_k, \beta_l, \beta_m$, the parameters of the marginal cost/production functions, and θ, the parameter on the labor adjustment cost in the modeling of the markups. We have 50 or 51 other parameters that enter linearly (7 or 8 coefficients from the law of motion for markup,[26] 32 coefficients from the two sets of constant and time dummies that correspond to each equation, and 11 polynomial coefficients in the productivity process), for whose estimation we apply the procedure of "concentrating out." This gives us a minimum of 16 over-identifying restrictions that allow to test for the specification.

[26]For some industries we have excluded capital in the markup specification.

2.4 Data

We estimate the model with data on ten (unbalanced panel) industry samples, which in total amount to more than 2,300 manufacturing firms and 17,000 observations, corresponding to the period 1990-2006. All variables come from the survey ESEE (Encuesta Sobre Estrategias Empresariales), a firm-level panel survey of Spanish manufacturing starting in 1990. At the beginning of this survey, firms with fewer than 200 workers were sampled randomly by industry and size strata, retaining 5%, while firms with more than 200 workers were all requested to participate, and the positive answers represented more or less a self-selected 70%. To preserve representation, samples of newly created firms were added to the initial sample every subsequent year. At the same time, there are exits from the sample, coming from both death and attrition.

The survey then provides a random sample of Spanish manufacturing with the largest firms oversampled. Information on the firms include, in addition to the usual output and input quantity measures, the firm-level variations for the price of the output and the price of the inputs, the introduction of technological (process and product) innovations, and the indicator of the state of the market *market dynamism*.[27]

Tables 1 and 2 report the size of the samples and some descriptive statistics. Table 1 focuses on prices. The evolution of prices over the whole period is moderate, but heterogeneous enough to motivate the exploration of the role of innovation in their evolution. An OLS regression of the growth of prices on the dummies of process and product innovation, controlling by time dummies, turns out to be hardly informative: almost no effect is statistically significant. Table 2 focuses on innovation. The introduction of process and product innovations is more heterogeneous. Firms tend to introduce more process than product innovations, at the approximate paces of one innovation every three and four years respectively. Innovation is, as expected, especially important in the more technology intensive sectors 3, 4, 5 and 6.

[27]Firms are asked to assess the current and future situation (contraction, stability, or expansion) of up to 5 separate markets in which they operate. The index of market dynamism is computed as a weighted average of the responses and proxies for the demand shifter.

2.5 Estimation Results

We have carried out the estimation of the system of equations (7), i.e. under dynamic pricing, with very sensible results that allow us to examine the impacts of innovation on prices. In the near future, we would like to also compare results with those under static pricing to evaluate which model is a better fit for the data. For the rest of this section, we first present these estimates (Table 3) and then derive the implications for the productivity growth and price variations via also markup due to the impact of innovation (Table 4).

The results of the preliminary estimates are summarized in Table 3. Columns (1)-(4) report the key parameters from the markup equation, Columns (5)-(7) contain the results for the parameters in the marginal cost/production function, and the last two columns present over-identifying test statistics and p-values. As noted in Table 3, to reach these results we use instead contemporaneous innovations in three industries (3, 8 and 9). In all but four industries (1, 2, 5 and 6) we are not able to include capital in the margin without worsening the results. Overall, the specification test passed in 8/10 industries.

The reported coefficients from the markup equation are those of market dynamism, innovation and adjustment cost. These results suggest the state of the market has a significant positive effect on the markup in six industries. Markup also increases in process innovation in seven industries. The impact of product innovation, however, is non-significant in all of them. Lastly, the labor adjustment cost parameter in column (4) is positive and of a reasonable size in all industries but one.

The parameters in the marginal cost/production functions in Columns (5) to (7) are in general good (although a few sectors show insignificant capital elasticities). Returns to scale are very close to constant in all industries but one.

The last part of our analysis is to comprehensively summarize the relationship among innovation, markup and prices using the estimates. Recall that innovation affects productivity and thus marginal costs. Furthermore, there exists adjustment costs in the markup under dynamic pricing, there exists adjustment costs in the markup, therefore cost changes induced by innovation may not be perfectly passed onto prices. To assess these two separate components, we first estimate the distribution of productivity ω_{jt} and summarize them by industry and innovation category in Columns (1)-(4) of Table 4. We then combine these productivity changes with changes in the markup due to innovation to summarize the findings on final predicted price changes in Columns (5)-(8) of the same Table.

To first get at the relationship between innovation and productivity, we first compute the individual values of productivity growth rates, given by $\Delta\omega_{jt} = \omega_{jt} - \omega_{jt-1}$. Then we split the industry samples into observations corresponding to the introduction of a process innovation only, product innovation only, both kind of innovations and, finally, observations corresponding to moments without any kind of innovation. Finally, we calculate weighted averages of the rates of growth, using firms sales shares lagged two periods as weights and replicating the observations for the small firms (less than 200 workers) as indicated by the known starting representativeness in the survey of each kind of firm.[28]

The corresponding results that capture the relationship between innovation and productivity are reported in Columns (1) to (4) of Table 4. They show that process innovation is associated with the most productivity increase as expected. However, as discussed before we do not expect a particular relationship between the increases in productivity when there is no innovation and when there is product innovation. This remains largely an empirical question. We find that for these Spanish manufacturing firms, absent any innovation there exist exogenous improvements in productivity in 7 industries. When there is only product innovation productivity increases are less intense in four industries and there is even a productivity decrease in two industries. It is then perhaps not surprising that for firms that have both process *and* product innovation, productivity increases are large in some industries, neutralized to near zero in some others and negative in the rest.

To arrive at the final predicted price changes, we need to also take into account the behavior of the markup and the results are summarized in Columns (5)-(8) of Table 4. They show that absent innovation, prices decrease by a more modest amount of 0.6 percentage point a year across industries on average. The introduction of process innovations changes this price decrease to 1.9

[28]Since the proportions were approximately 5% and 70% this implies replicating the smallest firms 14 times.

percentage points a year, which is quite remarkable if we take into account that firms take advantage of process innovations to enlarge the markup of 1.2 percentage points. Product innovation sometimes decreases price, and other times increases it, giving an average of almost zero change. Lastly, product innovations implemented together with process innovations tend to raise prices by an average of 0.4 percentage points.

2.6 Concluding Remarks

This paper examines the impact of process and product innovations on prices set by firms. This impact is modeled to be the result of two related processes: 1) the way in which innovation affects productivity and hence marginal cost and 2) the extent to which firms pass changes in costs onto prices via markup when there are adjustment costs. From this dynamic pricing problem with adjustment costs in markup, we derive a set of equations that is then estimated with yearly prices and innovations reported by an unbalanced panel of Spanish manufacturing firms.

Our estimation results show that markup is highly persistent and mostly pro-cyclical. Process innovations increase productivity and thus decrease marginal costs as expected. However, since cost reductions are only imperfectly passed through to prices, firms are also able to take advantage of process innovations by increasing their markups. In contrast, the effect of product innovations on prices is ambiguous; after the launch of improved or new products, markups are not significantly affected and productivity could either increase or decrease. The overall findings in the paper contribute to the scarce literature on innovation and firm-level prices.

Table 2.1: Sample Size and Price Descriptive Statistics, 1991-2006

Industry	Sample Size		Price Indices		OLS of Price Growth on Innovation (Time Dums. Incl.)			
	Obs.	Firms	Log (s.d.)	% Growth (s.d.)	Constant (s.d.)	Process (s.d.)	Product (s.d.)	Standard Error
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1. Metal and Metal Products	2365	313	0.106	0.017	0.008	-0.004	0.003	0.050
			(0.185)	(0.052)	0.006	0.003	0.003	
2. Non-metallic Minerals	1270	163	0.066	0.012	-0.003	-0.002	0.002	0.057
			(0.199)	(0.058)	0.011	0.004	0.004	
3. Chemical Products	2168	299	0.045	0.008	0.014	-0.003	0.001	0.053
			(0.197)	(0.055)	0.005	0.002	0.003	
4. Agric. and Ind. Machinery	1411	178	0.122	0.015	0.023	-0.002	0.003	0.026
			(0.150)	(0.026)	0.004	0.002	0.002	
5. Electrical Goods	1505	209	0.051	0.008	0.015	-0.004	-0.004	0.045
			(0.186)	(0.046)	0.007	0.003	0.004	
6. Transport Equipment	1206	161	0.053	0.008	0.011	-0.007	0.005	0.030
			(0.136)	(0.031)	0.010	0.002	0.002	
7. Food, Drink and Tobacco	2455	327	0.160	0.021	0.037	-0.001	0.002	0.053
			(0.188)	(0.054)	0.005	0.003	0.003	
8. Textile, Leather and Shoes	2368	335	0.119	0.015	0.015	-0.001	0.002	0.042
			(0.171)	(0.042)	0.005	0.002	0.003	
9. Timber and Furniture	1445	207	0.128	0.020	0.022	0.001	0.010	0.030
			(0.141)	(0.031)	0.005	0.002	0.002	
10. Paper and Printing Products	1414	183	0.124	0.017	0.018	-0.007	0.005	0.069
			(0.232)	(0.074)	0.009	0.004	0.004	

Table 2.2: Descriptive Statistics on the Introduction of Innovations, 1991-2006

Industry	Prop. of Obs.		Obs. with Proc.		Obs. with Prod.		Firms with Proc.		Firms with Prod.	
	Proc. (s.d.)	Prod. (s.d.)	Stable (%)	Occas. (%)	Stable (%)	Occas. (%)	Stable (%)	Occas. (%)	Stable (%)	Occas. (%)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
1. Metal and Metal Products	0.373	0.184	151	732	66	369	27	196	12	126
	(0.484)	(0.387)	(6.4)	(31.0)	(2.8)	(15.6)	(8.6)	(62.6)	(3.8)	(40.3)
2. Non-metallic Minerals	0.265	0.172	31	306	18	201	7	96	4	72
	(0.442)	(0.378)	(2.4)	(24.1)	(1.4)	(15.8)	(4.3)	(58.9)	(2.5)	(44.2)
3. Chemical Products	0.403	0.345	121	752	152	597	29	197	32	175
	(0.490)	(0.476)	(5.6)	(34.7)	(7.0)	(27.5)	(9.7)	(65.9)	(10.7)	(58.5)
4. Agric. and Ind. Machinery	0.332	0.354	85	384	79	420	17	96	14	89
	(0.471)	(0.478)	(6.0)	(27.2)	(5.6)	(29.8)	(9.6)	(53.9)	(7.9)	(50.0)
5. Electrical Goods	0.375	0.365	68	496	164	385	16	131	33	94
	(0.484)	(0.481)	(4.5)	(33.0)	(10.9)	(25.6)	(7.7)	(62.7)	(15.8)	(45.0)
6. Transport Equipment	0.464	0.313	149	411	105	273	32	97	21	82
	(0.499)	(0.464)	(12.4)	(34.1)	(8.7)	(22.6)	(19.9)	(60.2)	(13.0)	(50.9)
7. Food, Drink and Tobacco	0.305	0.223	148	602	141	407	31	182	28	149
	(0.461)	(0.416)	(6.0)	(24.5)	(5.7)	(16.6)	(9.5)	(55.7)	(8.6)	(45.6)
8. Textile, Leather and Shoes	0.242	0.230	71	502	122	422	17	158	21	117
	(0.482)	(0.421)	(3.0)	(16.5)	(5.2)	(17.8)	(5.1)	(47.2)	(6.3)	(34.9)
9. Timber and Furniture	0.285	0.257	71	341	41	331	13	110	11	96
	(0.451)	(0.437)	(4.9)	(23.6)	(2.8)	(22.9)	(6.3)	(53.1)	(5.3)	(46.4)
10. Paper and Printing Products	0.293	0.141	37	378	17	182	7	118	5	60
	(0.455)	(0.348)	(2.6)	(26.7)	(1.2)	(12.9)	(3.8)	(64.5)	(2.7)	(32.8)

Table 2.3: GMM Estimation of the System (Markup and Price Equation)[a]

Industry	Effects on Markup (SE)			Adj. Cost	Elasticities (SE)			Overidentifying Restrictions test	
	Mkt Dyn.	Proc.	Prod.	(SE)	K	L	M	$\chi^2(df)$	p val.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
1. Metals and Metal Products	0.010	0.017	-0.003	0.618	0.029	0.232	0.786	21.503	0.368
	(0.011)	(0.007)	(0.007)	(0.270)	(0.026)	(0.019)	(0.044)	(20)	
2. Non-metallic Minerals	0.045	-0.001	0.007	0.389	0.056	0.340	0.808	20.922	0.230
	(0.009)	(0.011)	(0.011)	(0.302)	(0.064)	(0.017)	(0.074)	(17)	
3. Chemical Prod.	0.007	0.013	-0.003	0.909	-0.025	0.177	0.888	33.251	0.032
	(0.007)	(0.006)	(0.006)	(0.243)	(0.028)	(0.013)	(0.030)	(20)	
4. Agric. & Ind. Machinery	0.021	0.016	-0.004	0.350	0.053	0.281	0.770	22.197	0.223
	(0.008)	(0.006)	(0.008)	(0.128)	(0.022)	(0.019)	(0.033)	(19)	
5. Electrical & Electronic Prod.	0.014	0.011	0.003	0.420	0.126	0.340	0.544	21.621	0.361
	(0.009)	(0.007)	(0.007)	(0.104)	(0.030)	(0.020)	(0.028)	(20)	
6. Transport Equipment	0.004	0.010	0.006	0.649	-0.030	0.262	0.807	13.927	0.788
	(0.010)	(0.009)	(0.007)	(0.166)	(0.049)	(0.035)	(0.055)	(19)	
7. Food, Drink & Tobacco	-0.011	0.013	0.000	1.214	0.115	0.248	0.690	14.801	0.610
	(0.011)	(0.006)	(0.007)	(0.426)	(0.038)	(0.024)	(0.059)	(19)	
8. Textile, Leather & Shoes	0.022	0.009	0.006	0.459	0.125	0.316	0.553	33.244	0.016
	(0.009)	(0.008)	(0.006)	(0.093)	(0.021)	(0.019)	(0.019)	(18)	
9. Timber and Furniture	0.041	0.016	-0.001	-0.370	0.077	0.209	0.734	26.436	0.067
	(0.010)	(0.007)	(0.007)	(0.187)	(0.031)	(0.015)	(0.036)	(17)	
10. Paper and Printing Prod.	0.019	0.010	-0.011	0.235	0.038	0.258	0.747	10.785	0.823
	(0.009)	(0.006)	(0.008)	(0.279)	(0.046)	(0.027)	(0.059)	(16)	

[a] In industries 4,7 and 10, capital not included in the markup equation. In industries 3, 8 and 9, process and product innovations are contemporaneous.

Table 2.4: Productivity and Price Effects of Process and Product Innovation

	Average productivity change				Average predicted price change			
Industry	No inn.	Proc. only	Prod. only	Both	No inn.	Proc. only	Prod. only	Both
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1. Metals and metal products	0.010	0.039	0.006	0.043	-0.010	-0.022	-0.006	-0.026
2. Non-metallic minerals	0.014	0.013	-0.048	0.002	-0.014	-0.013	0.048	-0.002
3. Chemical products	0.022	0.024	0.016	0.013	-0.022	-0.011	-0.016	-0.000
4. Agric. and ind. Machinery	-0.005	0.018	0.012	0.005	0.005	-0.002	-0.012	0.011
5. Electrical and electronic products	0.004	0.083	0.029	-0.012	-0.004	-0.072	-0.029	0.023
6. Transport equipment	0.011	0.043	0.006	0.017	-0.011	-0.033	-0.006	-0.007
7. Food, drink and tobacco	0.008	0.012	0.002	-0.005	-0.008	-0.001	-0.002	0.018
8. Textile, leather and shoes	-0.010	0.034	-0.026	0.002	0.010	-0.025	0.026	0.009
9. Timber and furniture	-0.004	0.022	-0.001	-0.002	0.004	-0.006	0.001	0.018
10. Paper and printing products	0.011	0.030	0.017	0.015	-0.011	-0.020	-0.017	-0.005

Based on the estimates of Table 2.3.

2.7 Appendix

Appendix A

Variable cost minimization implies the cost function

$$C_{jt} = C(K_{jt}, W_{jt}, P_{Mjt}, Q_{jt}/\exp(\omega_{jt})),$$

where W_{jt} and P_{Mjt} stand for the prices of the variable inputs L_{jt} and M_{jt} respectively. Marginal cost is

$$MC_{jt} = \frac{\partial C_{jt}}{\partial Q_{jt}} = \frac{\partial C}{\partial (Q_{jt}/\exp(\omega_{jt}))}(K_{jt}, W_{jt}, P_{Mjt}, Q_{jt}/\exp(\omega_{jt}))\exp(-\omega_{jt}).$$

On the other hand, by Shephard's lemma, optimal materials choice conditional on output is

$$M_{jt} = \frac{\partial C_{jt}}{\partial P_{Mjt}} = C_{P_M}(K_{jt}, W_{jt}, P_{Mjt}, Q_{jt}/\exp(\omega_{jt})).$$

Inverting the latest equation for $Q_{jt}/\exp(\omega_{jt})$, and using the resulting expression to replace $Q_{jt}/\exp(\omega_{jt})$ in marginal cost, we get

$$MC_{jt} = MC(K_{jt}, M_{jt}, W_{jt}, P_{Mjt})\exp(-\omega_{jt}) = MC(X_{jt})\exp(-\omega_{jt}),$$

where $X_{jt} = \{K_{jt}, M_{jt}, W_{jt}, P_{Mjt}\}$ is a vector of observable variables. Variable M_{jt} could be alternatively replaced by L_{jt}. In practice we prefer to keep the specification with M_{jt} because materials are less likely to be subject to adjustment costs.

Appendix B

The choice of which expression to use is not trivial. The inclusion of the disturbance e_{jt} implies that we only know the price decided by the firm up to this error: $P_{jt} = P^*_{jt}\exp(e_{jt})$. As in the structural estimation of production functions, a natural choice for the expression to be used would be an inverted input demand relationship. For example, profit maximization implies the demand for materials

$$M_{jt} = M(K_{jt}, W_{jt}/P^*_{jt}(1 - \frac{1}{\eta_{jt}}), P_{Mjt}/P^*_{jt}(1 - \frac{1}{\eta_{jt}}), \omega_{jt}).$$

But, as this equation makes clear, demands for labor and materials include the output price that when specified in terms of observed price will include the error e_{jt}. Having a variable measured with error in the expression to use in the $g(\cdot)$ function, even if it is uncorrelated with everything, is a drawback for the identification of the model.[29]

If we assume that we observe Q_{jt}, however, we have more than one solution.[30] A first possibility is inverting the (lagged) conditional demand for materials (see Appendix A), that is

$$\omega_{jt-1} = \ln Q_{jt-1} - \ln C^{-1}_{P_M}(K_{jt-1}, M_{jt-1}, W_{jt-1}, P_{Mjt-1}).$$

Another possibility is directly inverting the production function

$$\omega_{jt-1} = \ln Q_{jt-1} - \ln F(K_{jt-1}, L_{jt-1}, M_{jt-1}).$$

[29]Scholars estimating production functions tend to avoid this complication by assuming that all prices are the same for all firms and can be replaced by dummies. In this case prices disappear from the demand for materials, that can be written $M_{jt} = M_t(K_{jt}, \omega_{jt})$. See, for example, Levinsohn and Petrin (2003b) or Loecker (2011).

[30]Quantity is obtained by dividing revenue by the observed price. If revenue and price are affected by the same e error this ratio gives the correct quantity.

Appendix C

Write the Bellman equation using the explicit profits in terms of labor and materials and adding the cost of adjusting employment $AL(L_{jt}, L_{jt-1})$

$$\begin{aligned} V(S_{jt}, L_{jt-1}) &= \max_{\mu_{jt}}[\left(\frac{\mu_{jt}}{v} - 1\right)(W_{jt}L_{jt} + P_{Mjt}M_{jt}) - A(\mu_{jt}, \mu_{jt-1}, D_{jt}, z_{jt}, d_{jt}) - AL(L_{jt}, L_{jt-1})] \\ &\quad +\beta E_t[V(S_{jt+1})|S_{jt}, L_{jt}]. \end{aligned}$$

The first order condition for labor can be written as

$$\left(\frac{1}{v}\frac{\partial \mu_{jt}}{\partial Q_{jt}}(W_{jt}L_{jt} + P_{Mjt}M_{jtjt}) - \frac{\partial A_{jt}}{\partial \mu_{jt}}\frac{\partial \mu_{jt}}{\partial Q_{jt}} - \beta E_t[\frac{\partial A_{jt+1}}{\partial \mu_{jt}}|S_{jt}, L_{jt}]\frac{\partial \mu_{jt}}{\partial Q_{jt}}\right)\frac{\partial Q_{jt}}{\partial L_{jt}}$$

$$+\left(\frac{\mu_{jt}}{v} - 1\right)W_{jt} - \frac{\partial AL_{jt}}{\partial L_{jt}} - \beta E_t[\frac{\partial AL_{jt+1}}{\partial L_{jt}}|S_{jt}, L_{jt}] = 0,$$

where we use the shorthands $A_{jt}, A_{jt+1}, AL_{jt}$ and AL_{jt+1} for the values of the adjustment cost functions. From the first order condition for output it is easy to check that the first parenthesis can be replaced by $-\left(\frac{\mu_{jt}}{v} - 1\right)MC_{jt}$ so that we obtain

$$\begin{aligned} MC_{jt}\frac{\partial Q_{jt}}{\partial L_{jt}} &= W_{jt}\left(1 - \frac{1}{W_{jt}\left(\frac{\mu_{jt}}{v} - 1\right)}\frac{\partial AL_{jt}}{\partial L_{jt}} - \frac{1}{W_{jt}\left(\frac{\mu_{jt}}{v} - 1\right)}\beta E_t[\frac{\partial AL_{jt+1}}{\partial L_{jt}}|S_{jt}, L_{jt}]\right) \\ &= W_{jt}(1 + \Delta_{jt}), \end{aligned}$$

where Δ_{jt} represents the gap between the wage and the shadow price of labor under adjustment costs.

We can similarly obtain the following first order condition for materials:

$$MC_{jt}\frac{\partial Q_{jt}}{\partial M_{jt}} = P_{Mjt}.$$

Adding the two first order conditions we can see that, under adjustment costs of labor, the relationship between marginal and average variable cost becomes

$$\begin{aligned} \nu MC_{jt} &= \frac{W_{jt}L_{jt} + P_{Mjt}M_{jtjt}}{Q_{jt}}(1 + s_{Ljt}\Delta_{jt}) \\ &= AVC_{jt}(1 + s_{Ljt}\Delta_{jt}), \end{aligned}$$

where s_{Ljt} is the share of the wage bill in variable costs.

To estimate Δ_{jt} we use the fact that with the Cobb-Douglas specification of the production function the ratio of first order conditions gives

$$\frac{P_{Mjt}M_{jt}}{W_{jt}L_{jt}} = \frac{\beta_M}{\beta_L}(1 + \Delta_{jt}).$$

We assume that, for each firm, the average of the gaps tends to cancel over time and that $\widehat{\left(\frac{\beta_M}{\beta_L}\right)} = \frac{1}{T_j}{}_j$ $\frac{P_{Mjt}M_{jt}}{W_{jt}L_{jt}}$. We then estimate Δ_{jt} as

$$\widehat{\Delta}_{jt} = \frac{\frac{P_{Mjt}M_{jt}}{W_{jt}L_{jt}}}{\widehat{\left(\frac{\beta_M}{\beta_L}\right)}} - 1,$$

and $\ln(1 + s_{Ljt}\widehat{\Delta}_{jt})$ is constructed using the observed cost shares.

Chapter 3

Do Firm Level Shocks Generate Aggregate Fluctuations?

(with Maria Francisca Perez Veyl)

3.1 Introduction

Business cycle fluctuations are often thought to have caused by aggregate shocks, since uncorrelated sector or firm level shocks average out in the aggregate due to the law of large numbers. However, a number of recent studies show the diversification of idiosyncratic shocks breaks down when sectoral linkages or firm size distributions are highly skewed. These studies provide the insight that in an economy where few sectors serve as major input suppliers or few firms account for a disproportionate share of production, shocks to these sectors or firms can propagate to generate aggregate fluctuations.

While the sector based story has a long theoretical tradition accompanied by empirical evidence,[1] the relevance of firm level shocks to aggregate fluctuations remains to be acknowledged and quantified. Gabaix (2011) was the first study to formally show that in an economy with fat-tailed size distribution of firms, idiosyncratic shocks to firms diversify at a milder rate that leads

[1]Sector-based stories date back to the seminal work of Long and Plosser (1983) where they show input-output linkages can propagate independent production shocks in each sector over time and across sectors with a six-sector dynamic stochastic general equilibrium model. This study was later challenged by Dupor (1999) because if the number of sectors is large enough, positive shocks in one sector can still be neutralized by negative shocks in others. Though around the same time Horvath (1998, 2000) argue the cancellation of sector-specific shocks is affected by the existence of factor demand linkages; whether shocks diversify depends on not the number of sectors but whether there are a handful of them that are important inputs in the production processes of many others. Two recent studies, Carvalho (2008); Acemoglu et al. (2012), build on Horvath's insight and show that if there are only few input supplier sectors then the convergence of aggregate volatility to zero slows down considerably. A productivity shock in one major sector has a "cascading" effect in the economy and as a result affects not only the direct downstream sectors but also the series of them that are interconnected with each other. In all aforementioned studies, the authors provide also empirical exercises that support their theoretical models.

to nontrivial effects on aggregate fluctuations. Further, output volatility originated from micro-shocks is an increasing function of the Herfindahl index of firms' sales shares, or in other words, of how concentrated the economy is.[2] The theoretical framework is then accompanied by an empirical analysis where the author shows shocks to largest firms in the US statistically explains GDP growth over the period 1952-2008.

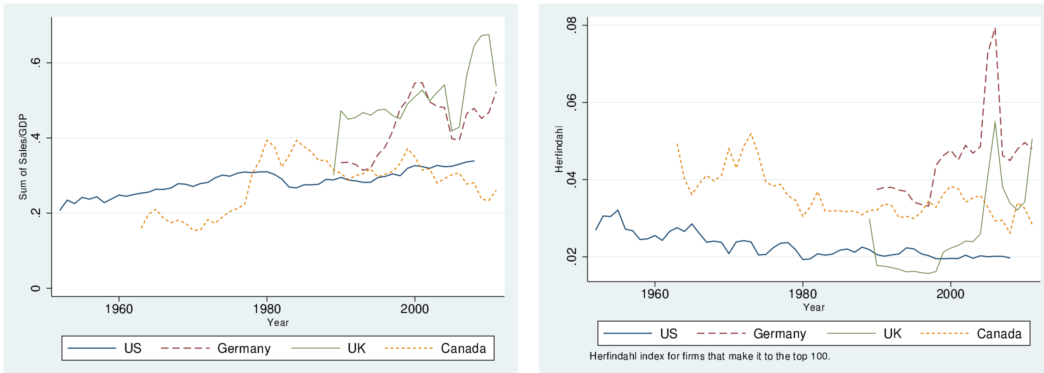
In light of Gabaix's theory, countries that are more concentrated than the US should observe stronger effects of firm level shocks on aggregate fluctuations. However, since the theory is relatively new and comprehensive panel data are not readily available for many countries, studies investigating the effects of idiosyncratic shocks outside the US are scarce.

This paper contributes to the existing literature by examining the relevance of idiosyncratic shocks to output fluctuations for three OECD countries (Germany, Canada and the UK) in addition to the US. Figure 3.1a presents the relative size of the top 100 firms in terms of their sales to GDP ratio in these four economies, and Figure 3.1b the Herfindahl index of firms' sales share among the largest 100 firms. Comparing with the US, the top 100 firms in other countries not only account for a larger share of their corresponding GDP, they are also much more concentrated. These descriptive statistics suggest large firms in these countries could be an important source of variability, and the effects of idiosyncratic shocks on the aggregate should be stronger as a result.

To assess the empirical relevance of idiosyncratic shocks for output fluctuations in these countries, we regress the growth of GDP per capita on the "granular residual" and its different lags. The granular residual, proposed by Gabaix (2011), is a transparent statistic that summarizes the importance of firm idiosyncratic shocks using the weighted sum of firm level shocks, where the weights are calculated as the firms' sales to GDP ratio to reflect their relative importance. This empirical strategy allows us to work with a broader sample of countries as we would not require input-output matrices to measure the importance of firm level productivity shocks for the economy, only annual firms' sales and employment, and countries' GDP per capita are necessary.[3]

[2]The results in Gabaix (2011) only require economies with sufficiently large herfindahls. The central case of firms' size being Zipf-distributed is justified because of its tractability and clean exposition. Also, this distribution generates very high herfindahls.

[3]Despite using an empirical strategy that is lenient on data requirements, we still face some data limitations along the time-series dimension. In the end, we are able to construct the granular residual using the top 100 firms for at least



(a) Top 100 Firms' Sales to GDP Ratio **(b)** Herfindahl Index of Top 100 Firms

Figure 3.1: Relative Size and Degree of Concentration of Top 100 Firms by Country

Our results show that idiosyncratic shocks to large firms are of little relevance in the UK or Canada but explain roughly 1/3 of the output fluctuations in the US and Germany. Further investigation indicates the top ranking firms are indeed the most important contributors to granular effects, but because diversification is at work still in the UK and Canada these firms are not able to explain GDP growth. Our study suggests that while firm size distribution is found to be highly skewed in many economies,[4] the ability of the largest firms to transmit shocks may not be universal and thus should not be taken for granted.

The studies closest to ours are di Giovanni et al. (2014) and Foerster et al. (2011); they both investigate the importance of firm level shocks but focus on one country each and differ in methodology. While the first finds support for the hypothesis that microeconomic shocks drive aggregate fluctuations, like ours the second paper casts doubts on the validity of this hypothesis. Using detailed French firm level data, di Giovanni et al. (2014) find firm-specific shocks to be almost twice as important as the combined effect of sectoral and macroeconomics shocks in driving aggregate sales growth. Further, they find that most of the effect is not coming directly from shocks to individual firms, but input-output linkages proxied by firm-to-firm covariance. Foerster et al. (2011) uses factor methods to decompose US industrial production and explore the plausibility of the

20 years for all of our countries.

[4]di Giovanni and Levchenko (2013) find in a set of 43 countries, firm size distributions are highly skewed and on average close to Zipf.

granular- and network-origins hypotheses of aggregate fluctuations. They find no support for the granular hypothesis, and aggregate volatility is better explained by co-variability among sectors. Furthermore, co-variability is explained by common factors, very little of which is the result of the transmission via input-output linkages.

The rest of the paper is organized as follows. Section 3.2 describes the empirical specification used to estimate the importance of firm level idiosyncratic shocks. In Section 3.3 we present the regression results for the countries in our sample, and in Section 3.4 we deconstruct the differences in the findings for the US and Germany versus those for the UK and Canada. Finally our conclusions are summarized in Section 3.5.

3.2 Empirical Strategy

In this section, we first summarize the theoretical framework from which the granular residual is derived and how to construct it. We then specify the empirical strategy and describe the data.

3.2.1 Granular Residual

In an economy with input-output linkages and endogenous input response, aggregate productivity growth is a weighted sum of firms' Hicks-neutral productivity growth, g_{z_i}:

$$g_{TFP} = \sum_i \frac{\text{sales of firm } i}{GDP} g_{z_i}, \tag{3.1}$$

where the weights $\sum_i \frac{\text{sales of firm } i}{GDP}$ capture the propagation effect from firm level shocks to the rest of the economy. Domar (1961) and Hulten (1978) show that these are the appropriate weights to measure the total effect of firm level productivity changes on aggregate productivity. The intuition is that an increase in productivity in firm i will increase output of other firms that use firm i's good as an intermediate input, which in turn increases output again, and so on.

Without disturbances, growth in GDP is proportional to the growth in TFP:

$$g_{GDP} = \mu \, g_{TFP}, \tag{3.2}$$

where $\mu \geq 1$ is the factor usage intensity that is a combination of the elasticity of substitution of labor and output elasticities with respect to production inputs.[5]

Combining (3.1) and (3.2), the impact of idiosyncratic shocks on aggregate output is captured by the following relationship:

$$g_{GDP} = \mu \sum_i \frac{\text{sales of firm } i}{GDP} g_{z_i}, \tag{3.3}$$

where $\Gamma^* = \sum_i \frac{\text{sales of firm } i}{GDP} g_{z_i}$ is Gabaix's "granular residual." Equation (3.3) provides the basis for the regression framework to test the granular hypothesis: granular effects are said to be present when the weighted sum of idiosyncratic shocks to large firms statistically explains GDP fluctuations as measured by growth in GDP per capita.

The only detailed calculation of this empirical strategy is the construction of firm level shocks. To avoid data availability issues, we also estimate firm level productivity using labor productivity of firm i:[6]

$$z_i = \ln \frac{\text{sales of firm } i \text{ in year } t}{\text{number of employees in firm } i \text{ in year } t}. \tag{3.4}$$

To compute the shocks to firms' productivity growth, we model firms' labor productivity growth, g, as depending on a set of firm's characteristics, X_{ijt}, and an idiosyncratic shock, ε_{ijt}:

$$g_{ijt} = \beta' X_{ijt} + \varepsilon_{ijt}. \tag{3.5}$$

For simplicity, we use year and industry dummies to proxy for firm's characteristics:

$$g_{it} = c + d_t + \varepsilon_{it} \tag{3.6}$$

$$g_{ijt} = c + d_t + \text{IND}_{jt} + \varepsilon_{ijt}, \tag{3.7}$$

[5]Consider an economy with the production technology $Y = A_t L_t^{\alpha}$ and consumer preference $U_t(C_t, L_t) = \log(C_t - \frac{L_t^{1+1/\phi}}{1+1/\phi})$. With competitive markets for output and labor, the equilibrium is characterized by $N_t = (\alpha A_t)^{\frac{1}{1+1/\phi-\alpha}}$ and $Y_t = \alpha A_t^{\frac{1+1/\phi}{1+1/\phi-\alpha}}$. In this economy, output growth is proportional to TFP growth $\hat{Y} = \mu \hat{A}$, where $\mu = \frac{1+1/\phi}{1+1/\phi-\alpha}$ is the factor usage.

[6]We implement the Olley and Pakes (1996) method to estimate the firm level Solow residual and contrast it with our measure of labor productivity. We find that the two methods yield similar and highly correlated estimates for the productivity growth in the US, the average correlation coefficient is 0.95.

and calculate firm level shocks as the demeaned labor productivity growth rates, where the mean is computed over all firms of the year (the case when we use only year dummies), or of the year and industry (when both year and industry dummies are used in the regressions).[7]

Using both estimations for the idiosyncratic shocks, we construct two versions of the "granular residual":

$$\Gamma_{t,v1} = \sum_{i=1}^{K=100} \frac{S_{i,t-1}}{Y_{t-1}} \hat{\varepsilon}_{it}; \quad \Gamma_{t,v2} = \sum_{i=1}^{K=100} \frac{S_{i,t-1}}{Y_{t-1}} \hat{\varepsilon}_{ijt}. \tag{3.8}$$

Since we are interested in the effect of the largest firm we only work with the largest top $K = 100$ firms ranked by their sales to output ratio in the previous period.[8]

3.2.2 Econometric Specification

To test for the effect of firm level idiosyncratic shocks on aggregate fluctuations we regress our measure of idiosyncratic shocks, Γ, on growth of GDP per capita using the following specification:

$$g_{Y_t} = \alpha + \sum_{i=0}^{2} \mu_i \Gamma_{t-i} + u_t. \tag{3.9}$$

The adjusted R-squared from regressing GDP growth g_{Y_t} on Γ_t and its different lags allows one to assess the extent to which idiosyncratic shocks explain the variability of GDP growth. If we recall equation (3.3), the coefficient on the granular residual will provide an estimation of the factor usage in these different countries.[9]

[7]The challenge with correctly identifying ε_{it} remains apparent. It is hard to identify ε_{it} in the data because aggregate shocks could cause firm i's volatility or reflect it. We do not directly address the reflection problem but perform robustness checks to also control for the common factors so to prove the explanatory power of the granular residual is not coming from aggregate shocks (e.g. oil, monetary, fiscal policy shocks and etc). The reflection problem is studied in (Manski, 2007). It is similar to "the problem of interpreting the almost simultaneous movements of a person and his reflection in a mirror." Manski points out unless one knows about optics and human behavior, he would not be able to tell whether the mirror image cause the person's movements or reflect them. We attempted to apply the few solution methods that have been proposed in the literature but to no avail. The reason is these solutions all involve exogenous variation that is hard to find in our context.

[8]To verify the empirical methodology we simulated a simple economy, with exogenous production and without linkages. With the true idiosyncratic shocks, we constructed the granular residual and regressed output on this measure. The empirical strategy tends to bias the granular residual downwards in its magnitude and volatility, decreasing its explanatory power. Further, it seems to capture well the idiosyncratic components of a firm's productivity growth in a granular world without inducing any spurious bias. Interestingly, the results also show that the explanatory power of the model decreases when the idiosyncratic shocks to large firms are less volatile.

[9]Gabaix (2011) takes $\mu = 2.6$ as the benchmark to compare the regression coefficient with for the US.

Table 3.1 provides a summary of country characteristics in terms of the variables used in the regression model. The average and the volatility of per capita GDP growth across the countries are very similar (columns 1 and 2), but firm level productivity growth is much more volatile in non-US countries (column 4). Column 5 shows the correlation between growth rates across firms is small for all of the countries, suggesting the measure we use for firm level shock does capture idiosyncratic variation to the firms. Finally, the average and standard deviation of the granular residual are smaller for version 2 than for 1 (columns 6 and 7 for version 1 and 8 and 9 for version 2). This shows when we demean by year-industry, we further get rid of industry-specific shocks that could confound our results.

Table 3.1: Country Characteristics

	(1) μ_{g_Y}	(2) σ_{g_Y}	(3) μ_g	(4) σ_g	(5) ρ_{g_i,g_j}	(6) $\mu_{\Gamma_{v1}}$	(7) $\sigma_{\Gamma_{v1}}$	(8) $\mu_{\Gamma_{v2}}$	(9) $\sigma_{\Gamma_{v2}}$
US	0.020	0.019	0.017	0.110	0.023	0.0003	0.0037	0.0003	0.0032
Canada	0.017	0.021	0.017	0.208	0.016	0.0006	0.0055	-0.0001	0.0044
Germany	0.019	0.019	0.015	0.179	0.020	0.0034	0.0092	-0.0001	0.0072
UK	0.019	0.022	0.011	0.258	0.010	0.0031	0.0144	-0.0016	0.0081

μ_{g_Y} is the average annual per capita GDP growth; σ_{g_Y} is the standard deviation of annual per capita GDP growth; μ_g is the average annual firm-level productivity growth; σ_g is the standard deviation of annual firm-level productivity growth; ρ_{g_i,g_j} is the average annual sample correlation of firm evel productivity growth; μ_Γ is l the average annual granular residual with year-demeaning; σ_Γ is the standard deviation of the annual granular residual; $\mu_{\Gamma^{sic2}}$ is the standard deviation of the annual granular residual with industry demeaning; $\sigma_{\Gamma^{sic2}}$ is the standard deviation of annual granular residual with industry demeaning.

3.2.3 Data

Firm level data for the US and Canada, UK and Germany are from Compustat North America, Computstat Global, respectively. The length of coverage varies and it spans from 1950 to present and from 1987 to present for NA and international companies. We keep only firms incorporated *and* headquartered in their home country so to exclude foreign firms to the best of our ability. The oil industry is excluded due to difficulty in teasing out real firm level shocks from the aggregate commodity price shocks.[10] Figure 3.4 summarizes the number of firms, with valid data required by our regression specification, by country and year for each country.

The advantage of using Compustat is the comparability of information reported across countries, but the coverage on the number of firms is limited in Germany and Canada for some of the years. Further, since the analysis is done over the top 100 firms in the economy we lose additional observations because for some years there are less than 100 firms. However, since sales of the top 50 and 100 firms as a percentage of GDP track each other closely for Germany and Canada, we restrict our samples to years with at least 50 rather than 100 firms. This means for years less than 100 (but more than 50) firms we use all firms in our empirical exercises.

Macroeconomic data (GDP, GDP per capita and GDP deflators) are taken from the World Bank's Development Indicators database. GDP deflators are used to convert sales into real terms.[11] Some of the sales figures in Compustat are denominated in non-local currencies, and we look to the respective country's central bank website to obtain the exchange rates.[12]

3.3 Estimation Results

In this section, we present the estimation results for the US and Canada, UK and Germany as specified in regression (3.9) for different lags for the granular residual.

3.3.1 Impact of Idiosyncratic Firm Level Shocks

Table 3.2 presents the regression results for the US, Germany, the UK and Canada. Taking the US as the benchmark, we see in columns (1) & (2) of Panel A that the granular residual explains 24% (35%) of the fluctuations in GDP growth using 1 (2) lag(s) when demeaning by year. When we control for the contemporaneous granular residual and its two lags, their coefficients are significant and hovering around 2.6—the theoretical value of μ Gabaix (2011) uses for comparison. The

[10] We also exclude financial firms as in Gabaix (2011), "because the nature of their sales are not in line with the meaning of 'gross output' in the paper."

[11] Ideally one would like to use industry production indexes but they are not readily available so we use GDP deflators instead.

[12] This could introduce potential measurement errors since the exchange rates were matched to the fiscal years, rather than the period covered by a firm's financial statements.

presence of granular effects is confirmed in these results.

When we control for industry-year specific shocks by demeaning at the 2-digit industry level, we observe an increase in the granular effects for the US in Panel B columns (1) & (2) of the same table. The resulting firm idiosyncratic shocks are closer to the true ε_{it} when industry specific shocks are controlled for, hence if the granular hypothesis holds we should expect to explain more of GDP fluctuations.[13]

The results for the other countries are mixed. For Germany, we find the GR to explain GDP fluctuations even better than for the US (35% (36%) when demeaning by year, and 24% (34%) when demeaning by year-industry). However, regardless of demeaning the GR by year or industry, we do not find granular effects in Canada or the UK. The adjusted R^2 is essentially zero and the coefficients are insignificant and often negative. Even though all countries meet the sufficient conditions, we find idiosyncratic shocks translate to aggregate fluctuations only in the US and Germany.

The natural progression at this point is to implement factor methods to examine whether any residual common shocks give rise to our findings, and to use principal component analysis to investigate which firms explain the aggregate fluctuations the most. However, since the identities of the largest firms change from year to year, we cannot use conventional methods to examine whether our findings are driven by common factors or specific firms. Thus in the next section we deconstruct our findings utilizing still the granular residual measure demeaned at the industry level, since this level of demeaning yields the most robust proxy for idiosyncratic shock as shown in Table 3.1.

[13]We also experimented with demeaning at the 3-digit and 4-digit level but the granular effects are weakened as a result. Increasing the level of disaggregation should improve the explanatory power further theoretically but at the same time inducing attenuation bias empirically because the mean would then be estimated with fewer firms. These two forces work against each other and complicate the interpretations of the regression results, hence we focus on year and year-industry at the 2-digit level demeaning in this paper as mentioned in the Introduction.

Table 3.2: Explanatory Power of the Granular Residual

Panel A: GR with Year Demeaning

	(1) USA	(2)	(3) Germany	(4)	(5) UK	(6)	(7) Canada	(8)
Γ_t	1.83**	2.51***	1.19**	0.93**	0.08	-0.00	-0.22	-0.37
	(0.69)	(0.69)	(0.42)	(0.44)	(0.32)	(0.35)	(0.58)	(0.60)
Γ_{t-1}	2.58***	2.88***	1.00**	1.15**	0.31	0.35	0.38	0.55
	(0.71)	(0.67)	(0.43)	(0.43)	(0.32)	(0.34)	(0.58)	(0.59)
Γ_{t-2}		2.13***		-0.23		-0.14		0.02
		(0.71)		(0.44)		(0.34)		(0.59)
Intercept	0.02***	0.02***	0.01	0.01	0.01***	0.02***	0.02***	0.02***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)
N	56	55	21	20	22	21	48	47
R^2	0.27	0.38	0.42	0.46	0.05	0.07	0.01	0.03
adj. R^2	0.24	0.35	0.35	0.36	-0.05	-0.09	-0.03	-0.04

Panel B: GR with 2-digit Industry Demeaning

	(1) USA	(2)	(3) Germany	(4)	(5) UK	(6)	(7) Canada	(8)
Γ_t	2.79***	3.70***	-0.04	0.00	0.21	0.00	-0.29	-0.57
	(0.76)	(0.78)	(0.61)	(0.55)	(0.61)	(0.68)	(0.70)	(0.72)
Γ_{t-1}	3.34***	3.92***	1.60**	1.95***	0.71	0.71	0.73	0.78
	(0.74)	(0.71)	(0.60)	(0.58)	(0.59)	(0.64)	(0.70)	(0.70)
Γ_{t-2}		2.07***		0.48		-0.24		0.32
		(0.75)		(0.55)		(0.61)		(0.70)
Intercept	0.02***	0.02***	0.01***	0.01***	0.02***	0.02***	0.02***	0.02***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)
N	56	55	21	20	22	21	48	47
R^2	0.37	0.48	0.31	0.44	0.07	0.10	0.03	0.05
adj. R^2	0.35	0.45	0.24	0.34	-0.03	-0.06	-0.02	-0.02

* for $p < .1$, ** for $p < .05$, and *** for $p < .01$. Standard errors in parenthesis.
Depvar: Per capita GDP growth is regressed on the granular residuals calculated over the top 100 firms.

3.4 Deconstructing the Empirical Results

We first examine the relative importance of top ranking firms amongst themselves in explaining GDP fluctuations. Since the ranking of these top firms changes every year, we drop observations that have the same ranking to construct the new GR. The results are presented in Figure 3.2 where we drop a top ranked "composite firm" one at a time and plot the percentage reduction in adjusted-R^2s against its average weights over the sample period. The figure shows lower ranks bunched together around zero in all four countries, suggesting their relative importance is small. However, the impact of the largest firms is prominent in all countries with the starkest contrast in the case of the US.

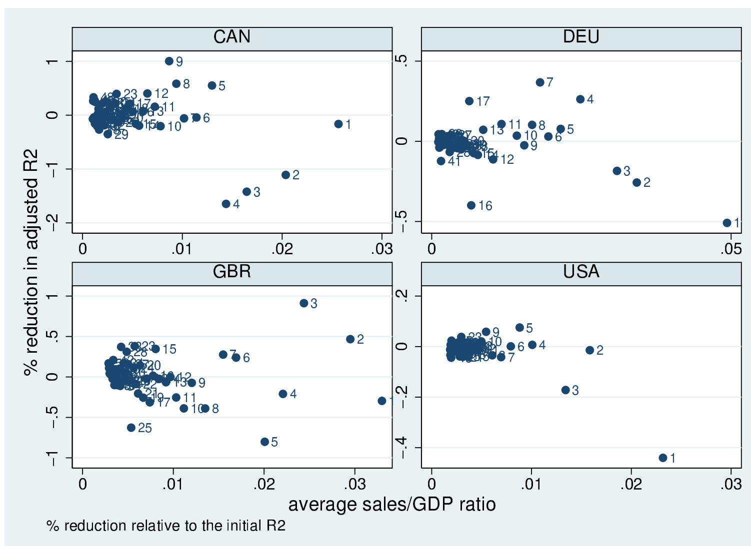


Figure 3.2: Reduction in Adjusted-R^2 after Dropping Top Ranked Firms

In light of this finding, we construct the granular residual with just the top 10 ranks to find that the positive findings in the US and Germany can be explained just as well (see Table 3.3). Table 3.4 additionally shows dropping top 10 ranks reduces the explanatory power significantly and the coefficients are no long significant. In the UK and Canada, however, firms of the same ranks did not play a significant role and dropping them leaves the adjusted R^2 as low as before and the coefficients remain insignificant.[14] Since summing up the granular residuals for the top ranks dilute their explanatory power in the UK and Canada, we suspect the diversification mechanism may be at work still in these two countries.

Table 3.3: Explanatory Power of the Granular Residual with the Top 10 Ranks

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	USA		Germany		UK		Canada	
Γ_t	2.71***	3.30***	-0.01	0.01	0.24	0.17	-0.46	-0.65
	(0.74)	(0.74)	(0.67)	(0.62)	(0.90)	(0.91)	(0.69)	(0.71)
Γ_{t-1}	3.03***	3.34***	1.77**	2.03***	0.99	1.14	0.40	0.39
	(0.72)	(0.70)	(0.66)	(0.64)	(0.88)	(0.92)	(0.69)	(0.70)
Γ_{t-2}		1.77**		0.30		-0.92		0.44
		(0.73)		(0.62)		(0.88)		(0.70)
Intercept	0.02***	0.02***	0.01***	0.01**	0.02***	0.02**	0.02***	0.02***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.01)	(0.00)	(0.00)
N	56	55	21	20	22	21	48	47
R^2	0.36	0.44	0.30	0.40	0.06	0.14	0.02	0.04
adj. R^2	0.34	0.41	0.22	0.29	-0.03	-0.01	-0.03	-0.03

* $p < .1$, ** $p < .05$, and *** $p < .01$. Per capita GDP growth is regressed on the granular residuals calculated over the top 10 ranks.

To further examine where the differences in the contribution of top firms with comparable importance could be coming from, we construct the rank specific weighted idiosyncratic shock and plot its correlation with GDP growth in Figure 3.3.[15] The right column of the figure shows unambiguously that the top ranks in Germany and the US are also positively correlated with GDP growth, but the picture is much more nuanced for Canada and the UK. We conclude that the higher tendency of high ranking firms moving in opposite directions in the UK and Canada neutralizes the effect of idiosyncratic shocks, thus it is not surprising the final measure of granular residual–the sum–do not explain aggregate fluctuations for these two countries.

The last piece of the puzzle remains as to why shocks diversify away in the case of the UK and Canada but not in the US or Germany. The reason may be that top ranked firms in these two

[14]This conclusion is robust to dropping the top 15 ranks.

[15]We compute the correlation between GDP growth and the contemporaneous rank-specific GR, with its one lag and then two lags. In the end we plot the average of the three correlations against the rank's average sales to GDP ratio.

Table 3.4: Explanatory Power of the Granular Residual without the Top 10 Ranks

	(1) USA	(2)	(3) Germany	(4)	(5) UK	(6)	(7) Canada	(8)
Γ_t	-1.92	-1.90	-4.34	-3.76	0.23	-0.11	0.50	0.25
	(2.50)	(2.53)	(3.10)	(3.09)	(1.13)	(1.58)	(1.33)	(1.34)
Γ_{t-1}	-0.12	-0.33	2.39	3.58	0.81	0.94	0.97	1.37
	(2.49)	(2.56)	(3.12)	(3.54)	(1.13)	(1.52)	(1.32)	(1.34)
Γ_{t-2}		-2.07		1.03		0.43		-0.85
		(2.54)		(3.09)		(1.26)		(1.33)
Intercept	0.02***	0.02***	0.01	0.01	0.01***	0.02**	0.02***	0.02***
	(0.00)	(0.00)	(0.01)	(0.01)	(0.00)	(0.01)	(0.00)	(0.00)
N	56	55	21	20	22	21	48	47
R^2	0.01	0.02	0.21	0.25	0.03	0.04	0.02	0.03
adj. R^2	-0.03	-0.03	0.13	0.11	-0.07	-0.13	-0.03	-0.04

* for $p < .1$, ** for $p < .05$, and *** for $p < .01$. Per capita GDP growth is regressed on the granular residuals calculated over the top 11-100 firms.

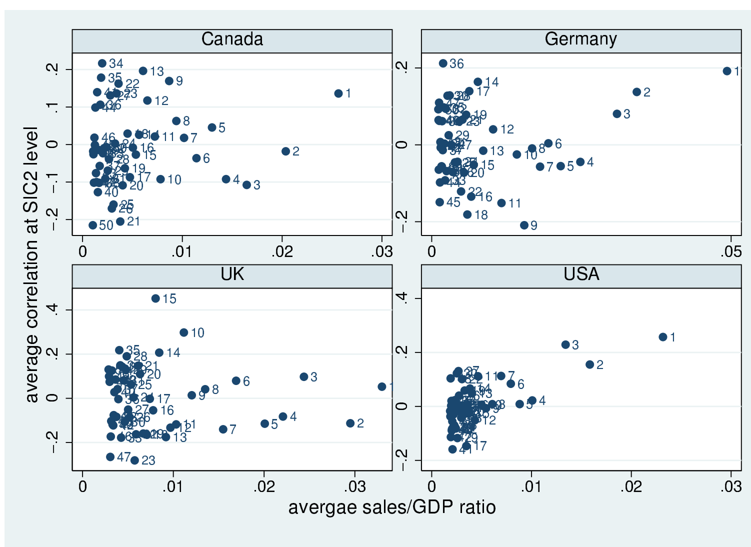


Figure 3.3: Correlation between Firm Specific GR and GDP Growth

Table 3.5: Top 10 Sectors in 2002

Rank	Canada SIC2	# Top Firms	$\frac{S}{Y}$	Germany SIC2	# Top Firms	$\frac{S}{Y}$
1	Elec.	5	6.47	Trans.Eq.	8	16.12
2	Food Strs.	5	5.34	Conglmrts.	3	7.52
3	Comm.	8	4.19	Chem.	14	6.82
4	Trans.Eq.	4	3.64	Comm.	5	3.59
5	Metal Ind.	7	2.57	G.Mechdis.Strs	2	2.59
6	G.Mechdis.Strs.	2	1.28	Metal Ind.	5	2.23
7	Chem.	5	1.16	Trans.Svcs	2	1.83
8	Paper and Allied	5	0.98	Machne.Eq.	14	1.60
9	Lumber	7	0.89	Motr.Warehsing.	1	1.59
10	Railroad Trans.	2	0.84	Whsl Tr.	4	1.30

Rank	UK SIC2	# Top Firms	$\frac{S}{Y}$	US SIC2	# Top Firms	$\frac{S}{Y}$
1	Food Prd.	8	6.41	Trans.Eq.	8	5.09
2	Chem.	7	6.33	G.Mechdis.Strs	8	4.27
3	Comm.	7	5.88	Comm.	11	3.48
4	Food Strs.	5	3.76	Chem.	10	2.64
5	G.Mechdis.Strs	4	3.53	Food Strs.	6	1.71
6	MetalMining	2	3.10	Conglmrts.	2	1.63
7	Trans.Eq.	4	2.18	Food Prd.	7	1.60
8	Biz.Svcs.	6	1.74	Machne.Eq.	6	1.52
9	Whsl Tr.	3	1.51	Biz.Svcs.	4	1.52
10	Prntn.Pub.	4	1.44	Whsl Tr.	4	1.34

Top 10 sectors as ranked by sector sales of top 100 firms (in the overall economy) in 2001.

countries are spread out across more sectors or they are from sectors not as interconnected with the rest of the economy. With our data and methodology we are only able to investigate first possibility here. To do this, we show in Table 3.5 the distributions of top firms by sectors in 2002 and find that they are similar across countries. Thus the positive results in the US and Germany are not purely driven by the fact that top ranking firms are from sectors that are also of disproportionate importance in the economy.

3.5 Conclusion

This paper quantifies the importance of firm level idiosyncratic shocks in explaining aggregate fluctuations. It is motivated by the theoretical framework (Gabaix, 2011) that shows in an economy with fat-tailed size distribution of firms, law of large numbers breaks down and idiosyncratic shocks to firms diversify at a much milder rate and lead to nontrivial effects on aggregate fluctuations. In the data, firm level shocks are estimated as the demeaned productivity growth rates, where the mean is calculated over top firms of the year or of the year and industry. These shocks are then weighted by firms' sales to GDP ratio, and the empirical strategy tests whether the sum of these weighted shocks statistically explains GDP fluctuations.

We find shocks to large firms are of little relevance in the UK or Canada but explain roughly 1/3 of the output fluctuations in the US and Germany. While top ranking firms contribute the most to granular effects, they do not always sum up to play a significant role in every country. We conclude that the reason they did not explain GDP growth in the UK and Canada is because diversification is at work still in these two countries. Our results suggest while firm size distribution is found to be highly skewed in most economies, the ability of the largest firms to transmit shocks is not universal. Future studies on the micro-foundation of aggregate fluctuations should take into account that apart from the granular theory, other transmission mechanisms may be at work, and the importance of which may differ on a country-by-country basis.

3.6 Appendix: Additional Figures

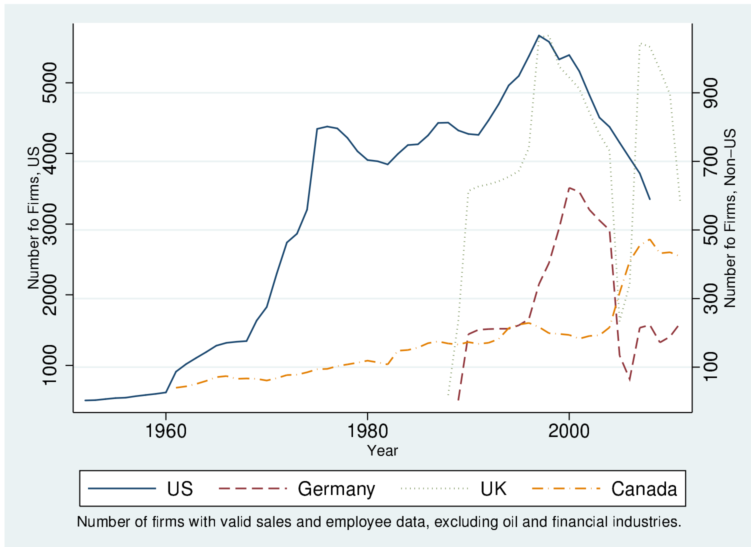


Figure 3.4: Number of Firms with Valid Sales and Employee Data

Bibliography

Acemoglu, D., Carvalho, V., Ozdaglar, A., and Tahbaz Salehi, A. (2012). The network origins of aggregate fluctuations. *Econometrica*, 80(5):1977–2016.

Ackerberg, D., Caves, K., and Frazer, G. (2006). Structural identification of production functions. MPRA Paper 38349, University Library of Munich, Germany.

Ackerberg, D. A., Caves, K., and Frazer, G. (2015). Identification properties of recent production function estimators. *Econometrica*, 83(6):2411–2451.

Aghion, P. and Tirole, J. (1994). The management of innovation. *The Quarterly Journal of Economics*, 109(4):1185–1209.

Agrawal, A., Rosell, C., and Simcoe, T. (2014). How do tax credits affect R&D expenditures by small firms? evidence from canada. *Working Papers*.

Aguirregabiria, V. (1999). The dynamics of markups and inventories in retailing firms. *The Review of Economic Studies*, 66(2):275–308.

Amoroso, S. (2014). The hidden costs of R&D collaboration. JRC-IPTS Working Papers on Corporate R&D and Innovation 2014-02, Institute of Prospective Technological Studies, Joint Research Centre.

Athey, S. and Bagwell, K. (2008). Collusion with persistent cost shocks. *Econometrica*, 76(3):493–540.

Audretsch, D. B., Menkveld, A. J., and Thurik, A. R. (1996). The decision between internal and external R&D. *Journal of Institutional and Theoretical Economics (JITE)*, 152(3):pp. 519–530.

Aw, B. Y., Roberts, M. J., and Winston, T. (2007). Export market participation, investments in R&D and worker training, and the evolution of firm productivity. *The World Economy*, 30(1):83–104.

Aw, B. Y., Roberts, M. J., and Xu, D. Y. (2011). R&D investment, exporting, and productivity dynamics. *American Economic Review*, 101(4):1312–44.

Barro, R. J. (1972). A theory of monopolistic price adjustment. *The Review of Economic Studies*, 39(1):17–26.

Beneito, P. (2006). The innovative performance of in-house and contracted R&D in terms of patents and utility models. *Research Policy*, 35(4):502–517.

Bergman, K. (2011). Internal and external R&D and productivity - evidence from swedish firm-level data. Working Papers 2011:27, Lund University, Department of Economics.

Berle, A. and Means, G. (1932). *The Modern Corporation and Private Property*. The Macmillan Company.

Bøler, E. A., Moxnes, A., and Ulltveit-Moe, K. H. (2015). R&d, international sourcing, and the joint impact on firm performance. *American Economic Review*, 105(12):3704–39.

Bonte, W. (2003). R&d and productivity: Internal vs. external r&d - evidence from west german manufacturing industries. *Economics of Innovation and New Technology*, 12(4):343–360.

Carlton, D. W. (1985). *Chapter 15 The theory and the facts of how markets clear: Is industrial organization valuable for understanding macroeconomics?*, volume Volume 1 of *Handbook of Industrial Organization*, pages 909–946. Elsevier.

Carlton, D. W. (1986). The rigidity of prices. *The American Economic Review*, 76(4):637–658.

Carvalho, V. M. (2008). *Aggregate Fluctuations and the Network Structure of Intersectoral Trade*. Universitat Pompeu Fabra, Economics Working Papers.

Cassiman, B. and Veugelers, R. (2006). In Search of Complementarity in Innovation Strategy: Internal R& D and External Knowledge Acquisition. *Management Science*, 52(1):68–82.

Cohen, W. M. (2010). Chapter 4 - fifty years of empirical studies of innovative activity and performance. In Hall, B. H. and Rosenberg, N., editors, *Handbook of The Economics of Innovation, Vol. 1*, volume 1 of *Handbook of the Economics of Innovation*, pages 129 – 213. North-Holland.

Cohen, W. M. and Levinthal, D. A. (1989). Innovation and learning: The two faces of r&d. *Economic Journal*, 99(397):569–96.

De Loecker, J. and Warzynski, F. (2012). Markups and firm-level export status. *American Economic Review*, 102(6):2437–71.

di Giovanni, J. and Levchenko, A. A. (2013). Firm entry, trade, and welfare in Zipf's world. *Journal of International Economics*, 89(2):283–296.

di Giovanni, J., Levchenko, A. A., and Méjean, I. (2014). Firms, destinations, and aggregate fluctuations. *Econometrica*, 82(4):1303–1340.

Domar, E. D. (1961). On the Measurement of Technological Change. *The Economic Journal*, 71(284):709.

Doraszelski, U. and Jaumandreu, J. (2013). Rd and productivity: Estimating endogenous productivity. *The Review of Economic Studies*.

Doraszelski, U. and Jaumandreu, J. (2015). Measuring the bias of technological change. CEPR Discussion Papers 10275, C.E.P.R. Discussion Papers.

Dupor, B. (1999). Aggregation and irrelevance in multi-sector models. *Journal of Monetary Economics*, 43(2):391–409.

Eichenbaum, M., Jaimovich, N., and Rebelo, S. (2011). Reference prices, costs, and nominal rigidities. *The American Economic Review*, 101(1):234–262.

Ellison, G. (2006). Bounded rationality in industrial organization. mimeo.

Foerster, A. T., Sarte, P.-D. G., and Watson, M. W. (2011). Sectoral versus aggregate shocks : A structural factor analysis of industrial production. *Journal of Political Economy*, 119(1):1–38.

Gabaix, X. (2011). The Granular Origins of Aggregate Fluctuations. *Econometrica*, 79(3):733–772.

Goldberg, P. K. and Hellerstein, R. (2011). How rigid are producer prices? Technical report, Federal Reserve Bank of New York. Staff Reports.

Hall, R. L. and Hitch, C. J. (1939). Price theory and business behaviour. *Oxford Economic Papers*, (2):12–45.

Himmelberg, C. P. and Petersen, B. C. (1994). R&D and Internal Finance: A Panel Study of Small Firms in High-Tech Industries. *The Review of Economics and Statistics*, 76(1):38–51.

Horvath, M. (1998). Cyclicality and Sectoral Linkages: Aggregate Fluctuations from Independent Sectoral Shocks. *NBER Chapters*, 1(4):781–808.

Horvath, M. (2000). Sectoral shocks and aggregate fluctuations. *Journal of Monetary Economics*, 45(1):69–106.

Hottenrott, H. and Peters, B. (2012). Innovative capability and financing constraints for innovation: More money, more innovation? *The Review of Economics and Statistics*, 94(4):1126–1142.

Hou, J. and Mohnen, P. (2013). Complementarity between internal knowledge creation and external knowledge sourcing in developing countries. UNU-MERIT Working Paper Series 010, United Nations University, Maastricht Economic and social Research and training centre on Innovation and Technology.

Hulten, C. R. (1978). Growth accounting with intermediate inputs. *The Review of Economic Studies*, 45(3):511–518.

Kashyap, A. K. (1995). Sticky prices: New evidence from retail catalogs. *The Quarterly Journal of Economics*, 110(1):245–274.

Katayama, H., Lu, S., and Tybout, J. R. (2009). Firm-level productivity studies: Illusions and a solution. *International Journal of Industrial Organization*, 27(3):403–413.

Klenow, P. J. and Malin, B. A. (2010). *Chapter 6 - Microeconomic Evidence on Price-Setting*, volume Volume 3 of *Handbook of Monetary Economics*, pages 231–284. Elsevier.

Klette, T. J. and Griliches, Z. (1996). The inconsistency of common scale estimators when output prices are unobserved and endogenous. *Journal of Applied Econometrics*, 11(4):343–361.

Lach, S. and Tsiddon, D. (1992). The behavior of prices and inflation: An empirical analysis of disaggregated price data. *Journal of Political Economy*, 100(2):349–89.

Levinsohn, J. and Petrin, A. (2003a). Estimating production functions using inputs to control for unobservables. *Review of Economic Studies*, 70(2):317–341.

Levinsohn, J. and Petrin, A. (2003b). Estimating production functions using inputs to control for unobservables. *The Review of Economic Studies*, 70(2):317–341.

Loecker, J. D. (2011). Product differentiation, multiproduct firms, and estimating the impact of trade liberalization on productivity. *Econometrica*, 79(5):1407–1451.

Lokshin, B., Belderbos, R., and Carree, M. (2008). The productivity effects of internal and external r&d: Evidence from a dynamic panel data model. *Oxford Bulletin of Economics and Statistics*, 70(3):399–413.

Long, Jr, J. B. and Plosser, C. I. (1983). Real business cycles. *Journal of Political Economy*, pages 39–69.

Manski, C. F. (2007). Identification of endogenous social effects: The reflection problem. *The Review of Economic Studies*, 60(3):531–542.

Midrigan, V. (2011). Menu costs, multiproduct firms, and aggregate fluctuations. *Econometrica*, 79(4):1139–1180.

Olley, G. S. and Pakes, A. (1996). The dynamics of productivity in the telecommunications equipment industry. *Econometrica*, 64(6):1263–97.

Pakes, A. (2003). A Reconsideration of Hedonic Price Indexes with an Application to PC's. *American Economic Review*, 93(5):1578–1596.

Peters, B., Roberts, M. J., Vuong, V. A., and Fryges, H. (2013). Estimating dynamic r&d demand: An analysis of costs and long-run benefits. Working Paper 19374, National Bureau of Economic Research.

Rotemberg, J. J. (1982). Monopolistic price adjustment and aggregate output. *The Review of Economic Studies*, 49(4):517–531.

Rust, J. (1987). Optimal replacement of gmc bus engines: An empirical model of harold zurcher. *Econometrica*, 55(5):pp. 999–1033.

Schmiedeberg, C. (2008). Complementarities of innovation activities: An empirical analysis of the German manufacturing sector. *Research Policy*, 37(9):1492–1503.

Sheshinski, E. and Weiss, Y. (1977). Inflation and costs of price adjustment. *The Review of Economic Studies*, 44(2):287–303.

Sheshinski, E. and Weiss, Y. (1992). Staggered and synchronized price policies under inflation: The multiproduct monopoly case. *The Review of Economic Studies*, 59(2):331–359.

Slade, M. E. (1991). Market structure, marketing method, and price instability. *The Quarterly Journal of Economics*, 106(4):1309–1340.

Slade, M. E. (1998). Optimal pricing with costly adjustment: Evidence from retail-grocery prices. *The Review of Economic Studies*, 65(1):87–107.

Stiglitz, J. E. and Weiss, A. (1981). Credit rationing in markets with imperfect information. *American Economic Review*, 71(3):393–410.

Tirole, J. (1989). *The Theory of Industrial Organization*. MIT Press.

Álvarez, L. J., Dhyne, E., Hoeberichts, M., Kwapil, C., Bihan, H. L., Lünnemann, P., Martins, F., Sabbatini, R., Stahl, H., Vermeulen, P., and Vilmunen, J. (2006). Sticky prices in the euro area: A summary of new micro-evidence. *Journal of the European Economic Association*, 4(2):575–584.

Curriculum Vitae

Education

Ph.D., Economics, Boston University, Boston MA, May 2016 (expected)
Dissertation Title: *Innovation and Productivity Growth with Heterogeneous Firms*
Dissertation Committee: Stefania Garetto, Jordi Jaumandreu and Marc Rysman

M.A., Political Economy, Boston University, Boston MA, May 2012
B.S.Comm. (*Honours*), Economics and Finance with a minor in Mathematics,
DePaul University, Chicago IL, May 2009
Honors Thesis Title: *The Transferability of China's Development Strategy: Case Studies of Special Economic Zones in Africa*

Fields of Interest

International Industrial Organization, Applied Econometrics, Economic Development

Working Papers

"The Dynamics of R&D Organization and Productivity Growth" (Job Market Paper).
"Innovation and Prices"
(with Jordi Jaumandreu), July 2015.
"Do Firm-Level Shocks Generate Aggregate Fluctuations?"
(with Maria Francisca Perez), July 2014.

Work in Progress

"Trade and Skill Upgrading in the Chinese Manufacturing Sector"
"Import Competition and Heterogeneous Innovation"
"The Anatomy of China's Exports to Africa"

Conference and Seminar Presentations

6th Annual ZEW/MaCCI Conference, Mannheim, Germany, June 2015
13th Annual International Industrial Organization Conference, Boston, April 2015
MaCCI Annual Conference, Mannheim, Germany, March 2015
NBER Productivity Lunch Seminar, Boston, September 2014
12th Annual International Industrial Organization Conference, Chicago, April 2014

Teaching Experience

Lecturer, Modeling Business Decisions, School of Management, Boston University, Spring 2015, Fall 2015, Spring 2016

Teaching Fellow, Statistics I, Mathematics Department, Boston University, Fall 2014

Co-Head Teaching Fellow, Introductory Macroeconomic Analysis, Economics Department, Boston University, Spring 2014

Teaching Fellow, Introductory Microeconomic Analysis, Economics Department, Boston University, Fall 2013

Instructor, Introductory Macroeconomic Analysis, Economics Department, Boston University, Summer 2012

Teaching Fellow, Introductory Macroeconomic Analysis, Economics Department, Boston University, Fall 2011

Work Experience

Research Assistant for Claudia Olivetti, Boston University, Spring 2012 - Spring 2013

Economic Development Research Intern, Lincoln Park Chamber of Commerce, Chicago IL, 2009

Corporate Relations Intern, The Chicago Council on Global Affairs, Chicago IL, 2009

McCormick Tribune Intern, South-East Asia Center, Chicago IL, 2008-2009

Legal Consulting Intern, Huron Consulting Group, Chicago IL, 2008

Fellowships and Awards

IED Travel Grant, Boston University, 2014

Funding for a Research Assistant, MA-RA Mentor Program, Boston University, Department of Economics, Summer 2013, Spring 2014

Teaching Fellowship, Boston University, 2011, 2013

56th Presidential Inauguration Conference Scholar, 2009

DePaul University Centennial Scholarship, 2005 - 2009

Computer Skills: Stata, Matlab, Gauss, LaTeXand MS Office

Citizenship: China/F1