R&D and Productivity: The Effects of Heterogeneity and Uncertainty*

Job Market Paper

Kevin Andrew[†]

March 10, 2020

Abstract

In this paper I examine the effects of R&D spending on revenue productivity over the life cycle of publicly traded manufacturing firms. I do this by estimating a gross production function and recovering a controlled Markov process which allows productivity to depend on age since listing, lagged productivity and R&D spending. I use a novel identification strategy to isolate the non-linear effects of R&D on productivity over the life cycle. I also estimate the effects of R&D on the volatility of revenue productivity. I find that R&D has non-linear effects on both the mean and the variance of productivity. In particular, more mature firms are more efficient at converting R&D spending into productivity and engage in R&D which is more uncertain. Firms with higher revenue productivity are also more efficient at turning R&D inputs into future productivity. My findings suggest that the R&D process is non-linear and uncertain.

* This research was support by the Social Sciences and Humanities Research Council of Canada and the John Deutsch Institute Doctoral Stipend. Thanks to Allen Head, Marco Cozzi and Huw Lloyd-Ellis for excellent supervision. All errors are my own.

† Queen's University, Department of Economics, Kingston, Ontario, K7L 3N6.

Email: andrewk@econ.queensu.ca

1. Introduction

The relationship between R&D (research and development) spending and productivity has been studied extensively within economics in the last 50 years.[1] This is due to the consensus view that aggregate productivity is of central importance for determining living standards in advanced modern economies. R&D spending measures the deliberate attempts of firms to develop new products, new markets and new production processes. This is why R&D expenditures are a natural place to look to understand the complex processes of productivity accumulation.

My paper addresses the impact of R&D on revenue productivity (TFPR) using a production function approach. I use data on publicly traded manufacturing firms coupled with a novel identification strategy to identify non-linearity and uncertainty in the relationship between R&D and TFPR. My main takeaway from this project is that the outcome of R&D investment is inherently heterogeneous and complex. There are very few simple rules of thumb to describe the impact of measured R&D on productivity. For example, R&D performers are on average more productive, while the marginal R&D dollar appears to lower productivity while reducing unceratinty. This suggests that R&D has effects on productivity at both the intensive and extensive margin while also reducing uncertainty. Furthermore, the marginal R&D dollar is more effective at more productive and more mature firms. I counterintuitively find that exiting firms are more productive on average than continuing firms. This can be explained by the fact that firms are dropped from my sample due to buyouts. I test these results for robustness by looking at different sample periods and different industries within manufacturing. My interpretation of the results do not change after these robustness checks.

I view R&D expenditure as a form of investment in the future profitability of the firm. It is inherently different from investment in tangible capital in that its outcome is harder to measure and is more uncertain. The risks involved in traditional investment include cost overruns, delays and disruption to ongoing business. In contrast, many investments in R&D are abandoned altogether. Furthermore, there are many examples of firms discovering new technologies and failing to commercialize them. For example, the transistor was patented at Bell Labs in 1948. Since Bell was a telephone company they chose not to pursue the transistor, which had creative applications in radios as well as modern computers. Another example is the inno-

[1]There are some comprehensive reviews of this empirical literature. Cohen (2010) begins with the qualitative observations of Schumpeter (1942) on the nature of creative destruction and follows the literature over a long period, placing emphasis on the relationship between firm size and R&D spending. Hall et al. (2010) discusses the literature on identifying the returns to R&D using the knowledge capital model pioneered by Griliches et al. (1979). I do not follow the knowledge capital approach in this paper, but most studies of firm level R&D do.

vations which were developed at Xerox Parc. These included the modern Graphical User Interface, keyboard and mouse for the personal computer. Since Xerox was a photocopy company they chose not to pursue these innovations.[2] These examples make clear that a firm must both innovate and commercialize in order to realize the full return of their R&D dollars.

Existing work has argued that the nature of R&D is heterogeneous with respect to firm size. For example, Akcigit and Kerr (2018) argue that established firms will choose to focus on improving existing product lines while younger firms may focus more on developing new product lines. This means that more established firms will have more certain R&D outcomes. My findings are relevant for this literature. My approach is to make less theoretical assumptions than Akcigit and Kerr (2018). I do not find a significant of firm size on the effect of R&D spending on TFPR and find that larger firms have more uncertain R&D outcomes, which runs counter to the findings in Akcigit and Kerr (2018).[3]

I am not aware of any papers which directly study the link between age since listing and productivity in a way that is comparable to my work. However, there is a large literature using firm level data which suggests that employment growth outcomes are more uncertain for younger firms and that this uncertainty declines as a firm ages.[4] My work is complementary to these results, though I focus on the dynamics of revenue productivity as opposed to employment. The mean and variance of TFPR is declining over the life cycle. It is important to take care in interpreting my results as my focus is on publicly traded manufacturing companies.

My empirical strategy is to estimate a gross production function using data on sales, employment, capital expenditure, cost of goods sold and R&D spending. I take my data from the Compustat database. As part of my empirical strategy I estimate a revenue productivity process which depends on R&D expenditures as well as age since listing and lagged productivity. I also estimate a function which captures the conditional heteroskedasticiy of TFPR dynamics. In this way, R&D can influence both the mean and variance of future productivity. It turns out that there are non-linear marginal effects of R&D on both the conditional mean and conditional variance of productivity.

[2]These examples are taken from Isaacson (2014). The emphasis in this book is on the teamwork required in innovation, both within and between companies as well as the difference between the invention and commercialization, as the two examples show. In order for R&D to show up as revenue productivity the outcome must be commercialized.

[3]Here I follow Akcigit and Kerr (2018) and use productivity as a proxy for firm size. Acemoglu et al. (2018) is another example of a recent endogenous growth paper using firm level data. In this paper the authors try to incorporate the firm life cycle as well as size effects. My findings suggest the life cycle is a very important determinant of productivity dynamics.

[4]This literature is summarized in Decker et al. (2014). For another perspective on the same data see Sterk et al. (2018).

In this paper I make two methodological contributions. First, I extend the methodology of Doraszelski and Jaumandreu (2013) to a more general setting that closer resembles the firm level data available in both publicly and privately available datasets. I do this using the insights of Gandhi et al. (2011) on identification of gross production functions by assuming cost minimization of a variable input. To be more precise, Doraszelski and Jaumandreu (2013) rely on firm level variation in prices and wages for identification. My method combines the materials inversion technique of Levinsohn and Petrin (2003) with a first order condition as in Gandhi et al. (2011) and De Loecker et al. (2019). This allows me to have a structural interpretation of my production function and to recover revenue productivity. To my knowledge this strategy has not been used before in the literature.

I also expand the Controlled Markov Process considered in the paper by Doraszelski and Jaumandreu (2013). I allow age since listing to affect the conditional mean of productivity. This is motivated by a recent literature which identifies strong life cycle properties of firm dynamics. I also include a skedastic function which allows me to systematically examine the channels through which R&D , productivity and age since listing affect the volatility of TFPR growth. Young firms are known to have more volatile employment growth rates than mature firms. I can systematically capture the effect of R&D on volatility. In order to do this I use the tool of the skedastic function.

The study of the impact of R&D on productivity has a long history in economics. The dominant paradigm is the knowledge capital framework introduced in Griliches et al. (1979). A recent survey of this literature is available in Hall et al. (2010) in the context of identifying the returns to R&D spending. I do not disagree with the theoretical concept of knowledge capital, however, there are complications inherent in the construction of this stock variable. These include sensitivity to starting values as well as the depreciation rate. I use flow spending on R&D to avoid this problem. I am not the first to do this, however my approach is relatively new in the literature. Embedded within my framework is the idea that knowledge capital is part of residual productivity.

With the increasing availability of high quality firm level data it has become possible to scrutinize the many theories of firm growth. My work suggests that there is a strong life cycle to TFPR growth, independent of R&D spending and lagged productivity. This is consistent with a firm learning model such as Jovanovic (1982). A more recent quantitative evaluation of this mechanism is found in Arkolakis et al. (2018). This mechanism is complementarry to the acquisition of intangible and tangible capital as found in papers such as Clementi and Palazzo (2016) and D'Erasmo (2007). A nested hypothesis test is beyond the scope of this paper, but the fact that several models of firm growth are consistent with the data further motivates the main finding of this paper: that the relationship between productivity and R&D spending is uncertain and complex.

The paper is organized as follows. Section 2 discusses the empirical methodology used to identify the relationship between productivity and R&D spending. Section 3 introduces the data and discusses variable creation and sample selection. Section 4 presents the main results. Section 5 presents some extensions and robustness exercises. Proofs as well as additional tables are relegated to Section A.1 , Section A.2 , Section A.3 and Section A.4.

2. Empirical Methodology

The purpose of this paper is to use micro-level data to estimate the effects of R&D spending on productivity over the life cycle of newly listed firms.[5] In order to accomplish this I define and estimate a production technology and explicitly model the channels through which R&D spending can influence revenue productivity over the firm life cycle. My empirical approach aims to have a structural interpretation with a minimum of assumptions. In order to do this, I build on a recent literature which provides conditions for the identification of gross production functions.

2.1 A Model of R&D and Productivity

2.1.1 Definitions

The unit of analysis in this paper is an individual firm. A firm may have many separate business locations. Given my focus on manufacturing firms, these business locations can best be thought of as individual plants. Some firms have many different plants, while others have only one. A firm transforms inputs into outputs via its gross production function.

Definition 1. ***(The Gross Production Function)***

$$Y = K^{\theta_k} L^{\theta_l} M^{\theta_m} e^{\omega}$$

It will often be convenient to write the gross production function in natural logs.

$$y = \theta_k k + \theta_l l + \theta_m m + \omega$$

Where productivity is broken up into:

$$\omega = \iota + \tau + z + \varepsilon$$

[5] Throughout the paper I will be discussing revenue total factor productivity (TFPR). This means that R&D may raise the demand for a firms product or lower the costs of producing it. In either case, this effect will show up in the TFPR residual. I will use the terms productivity, revenue productivity and TFPR synonymously.

ι and τ are industry and time effects. z is persistent productivity known to the firm. ε is unforecastable productivity which is not known to the firm within the period.

In this example the inputs are capital (K), labour (L) and materials (M). I have chosen a Cobb-Douglas production technology. This technology can be thought of either as the true underlying technology or as a first order approximation to any smooth production function.[6] Inputs can be classified as either static or dynamic. Static inputs are chosen within period while dynamic inputs must be chosen one period in advance subject to adjustment costs. Capital is a dynamic input while materials is static. Depending on the time period and setting, labour may be static or dynamic. Given my focus on the manufacturing sector, which is heavily unionised, I will assume that labour is a dynamic input. In the preceding definition I have omitted the time, industry and firm subscripts for ease of notation.

My methodology is designed to be compatible with firm level datasets such as the Longitudinal Business Database (LBD) in the United States or T2-LEAP in Canada. Such datasets provide measures of gross output or sales as well as firm age, employment, investment and materials. They can be merged with other datasets which contain information on R&D expenditures. A measure of the capital stock must be constructed from data on investment using a perpetual inventory method. It is common to use industry level price indices to deflate output, investment and materials as it is rare for such datasets to have information about firm level prices. The data form an unbalanced panel where exit is non-random.

Definition 2. ***(Controlled Markov Process)***

$$z' = \mu(z, r, a) + \sqrt{\sigma^2(z, r, a)}\xi'$$

Where $\mu(z, r, a)$ is referred to as the conditional mean function and $\sigma^2(z, r, a)$ is referred to as the skedastic function.

Following Doraszelski and Jaumandreu (2013) I assume that productivity follows a Controlled Markov Process as introduced in Definition 2. This process captures the idea that R&D spending is an investment in the future profitability of the firm. My paper generalizes the approach used in Doraszelski and Jaumandreu (2013) in important ways.

First, I am very interested in the firm life cycle. A recent literature has argued that employment growth follows a distinct life cycle pattern. In particular, job creation is on average higher for young firms. Exit rates are also high for young firms.[7]

[6] This differentiates my work from Gandhi et al. (2011) who use an entirely non-parametric approach. I follow Ackerberg et al. (2015) and Doraszelski and Jaumandreu (2013) in specifying an explicit production technology.

[7] This literature has grown in recent years due to better measurement of firm age in administrative datasets. Haltiwanger et al. (2013) argues that young firms contribute disproportionately to

This work suggests that a small subset of young firms are extremely important for job creation. Any research which sheds further light on the properties of these job creators would have important policy implications. Therefore, I am interested in testing whether TFPR has similar life cycle properties among newly listed firms. I allow age since listing to effect both the conditional mean and variance of productivity in Definition 2. My hope is to isolate pure age effects from the effects which are due to investment in R&D .

Secondly, I introduce the skedastic function. This is denoted $\sigma^2(z, a, r)$ in Definition 2. The skedastic function captures the fact that the variance of the productivity process may be non-constant with respect to lagged productivity, age since listing and R&D expenditures. My interest in the skedastic function is motivated by the fact that employment growth rates for young firms have higher variance. I want to investigate whether this is due to a pure age effect or due to R&D spending. It is reasonable to assume that R&D investment is uncertain. The skedastic function allows me to isolate such effects.

One of the key themes from Doraszelski and Jaumandreu (2013) that I re-emphasize is the non-linearity of the Controlled Markov Process. In particular, Definition 2 allows lagged productivity, age since listing and R&D expenditures to influence TFPR in flexible ways. I will emphasize the non-linearities that I find when discussing my results.

2.1.2 Assumptions

I now introduce the explicit assumptions of the model, which I number. As I will show in Section 2.2 , these assumptions are sufficient to identify the production function outlined in Definition 1 and the Controlled Markov Process from Definition 2.

Assumption 1. ***(Market Structure)***
Firms take input and output prices as given.

Since the output measure is sales, it is impossible to separately identify supply and demand shocks in this paper. In practice, this assumption guides my choice of deflators for output, materials and wages. In particular, I assume that there is a real wage which is common to firms in all manufacturing industries and that output and materials prices vary across manufacturing sectors, but not within narrow industries.

In Section A.1 I provide a relaxation of this proof. If mark-ups are constant, the elasticity, θ_m has the structural interpretation of the mark-up times the underlying structural elasticity. This shows why it is necessary to be careful to interpret

job creation. Decker et al. (2014) analyses this literature in the context of the falling start-up rate.

productivity as revenue productivity, which captures a combination of demand and supply shocks.

Assumption 2. ***(Cost Minimization)***
I assume that firms minimize costs with respect to materials within period:

$$\min_{M} \; W_m M$$

Subject to:

$$Y \geq \mathbb{E}[K^{\theta_k} L^{\theta_l} M^{\theta_m} e^{z+\varepsilon}]$$

This behavioural assumption is needed to justify using data on the share of spending on materials. The assumption of cost minimization is common in the literature on production function estimation. Gandhi et al. (2011) use this assumption to overcome identification problems similar to mine. Furthermore, De Loecker et al. (2019) use this assumption in a different context to identify firm level mark-ups. The multiplier on the constraint in this problem can be interpreted as the marginal variable cost. If we wish to assume that labour is a static input then its cost must be added to the objective function.

Assumption 3. ***(Exit Policy)***
I assume that exit follows a threshold policy, where:

$$\chi = \begin{cases} 0 & \textit{if: } z \geq z^*(k,n) \\ 1 & \textit{if: } z < z^*(k,n) \end{cases}$$

In lieu of specifying an explicit dynamic programming problem I posit a threshold exit policy. This is the outcome of standard models of investment. Olley and Pakes (1996) prove that exit follows such a policy in the standard framework. Introducing the skedastic function as well as non-linearities implies that a dynamic programming proof of monotonicity in z is not feasible. Assumption 3 makes explicit that we are assuming such a threshold policy.

Assumption 4. ***(Timing and Information Sets)***

$$\mathbb{E}[\varepsilon|\mathcal{I}] = 0$$

$$\mathbb{E}[\varepsilon' + \sqrt{\sigma^2(z,r,a)}\xi'|\mathcal{I}] = 0$$

Where $\mathcal{I}$ represents the current period information set.

The timing assumptions in Assumption 4 clarify which variables one would expect to be endogenous. Since ε is unforecastable to both the econometrician and the firm all current period variables are expected to be predetermined with respect to ε. However, the presence of the skedastic function in Definition 2 implies that only lagged values of R&D expenditures are predetermined. Implicit in this definition is the assumption that current period capital and labour are predetermined, while only lagged materials are predetermined. These timing assumptions are used to create moment conditions for estimation.

The assumptions given are sufficient to justify the estimator introduced in the next section. However, it may be helpful to consider what a dynamic programming problem for the firm might look like. In addition to the assumptions above, I assume convex adjustment costs for the dynamic inputs, capital and labour, as well as R&D expenditures. The firm solves a within period profit maximization problem to choose its materials, given an output and materials price, which it takes as given.

$$\pi(z, K, L) = \max_{M} \{\mathbb{E}[e^{z+\varepsilon} K^{\theta_k} L^{\theta_l} M^{\theta_m}] - W_m M\}$$

the dynamic programming problem would be of the form:

$$V^A(z, K, L) = \max_{R, K', L', \chi} \{\pi(z, K, N) - \Phi_k(K, K') - \Phi_n(L, L') \quad - \Phi_r(R) + \beta(1-\chi)\mathbb{E}[V^{A+1}(z', K', L')|z, R, A]\}$$

Where $\Phi(\bullet)$ represents a convex adjustment cost function. Notice that the firm makes a discrete exit decision, which would lead to a threshold policy such as Assumption 3 if the value funtion is monotonic in z. However, proving global monoticity in the presence of non-linearities and the skedastic function is not feasible.

I re-iterate that I am not explicitly assuming that the firm solves this problem in estimating the production function. Only the numbered assumptions and definitions are required for identification.

$$\underbrace{\frac{\partial \Phi_r}{\partial R}}_{\text{MC}} = \underbrace{\frac{\partial \mathbb{E}[V^{A+1}(z', K', L')|z, R, A]}{\partial R}}_{\text{MB}}$$

However, when interpreting the results, such a model is useful. An optimizing firm will set the marginal benefit of future R&D equal to the marginal cost. The marginal cost depends on the adjustment cost function and the marginal benefit depends on the way that R&D influences future productivity in Definition 2. By definition the way that R&D is profitable is through changes in expected productivity. I will often use this interpretation when discussing the results.

2.2 The Estimator

Now I introduce my estimation procedure which takes as given the definitions and assumptions from Section 2.1.

2.2.1 Stages 1 and 2

A recent literature has argued that proxy function methods do not identify gross production functions. Ackerberg et al. (2015) show that one way to overcome this is to estimate a value added production function and assume that it is Leontief in materials. Gandhi et al. (2011) suggest augmenting the estimation routine with information about the share of spending on materials, which is often available from accounting data. I follow the latter approach because it does not impose the arbitrary restriction that there is no substitutability between materials and other inputs.

Proposition 1 shows how the structural parameter θ_m is related to data on the share of firm spending which goes to materials. This share can be defined as the cost of the materials input divided by sales. Therefore, it is not necessary to view prices of either the input or the output. Proofs of all propositions are relegated to Section A.1 .

Proposition 1. ***(The Share Equation)***

$$\ln S = \ln \theta_m \mathcal{E} - \varepsilon$$

Where:

$$S = \frac{W_m M}{PY} \qquad \textit{and} \qquad \mathcal{E} = \mathbb{E}[e^{\varepsilon}]$$

As a consequence, data on S identifies the structural parameter θ_m.

Notice the inclusion of $\mathcal{E}$. This is the firms forecast of the value of the iid shock.[8] This proposition shows that data on the share of spending on materials identifies the structural parameter θ_m. The interpretation of θ_m is the elasticity of output with respect to materials. Its identification is independent of the estimation of the other parameters of the model.

Given Assumption 3, the optimal cutoff is an unknown function of z, k and l. Following Olley and Pakes (1996) I treat this function non-parametrically and estimate

[8] If I had assumed that ε was measurement error I would not have had to account for this term. However, it is feasible to account for forecast errors in optimization as follows. I regress $\ln S$ on a constant and then recovering the structural parameter, θ_m by defining:

$$\hat{\mathcal{E}} = \frac{1}{NT} \sum_{i,t} e^{\hat{\varepsilon}_{i,t}}$$

a probit model. I refer to $G(k, l, m)$ as the unknown function which I estimate as a 3rd-order polynomial. The probit is specified as:

$$\mathcal{P}(\chi = 1) = \Phi(G(k, l, m))$$

Where Φ is the standard normal distribution. Together, the first two stages of the estimation procedure give $\hat{\theta}_m$ and a set of exit probabilities $\hat{P}_{i,j,t}$.

2.2.2 Stage 3: Moment Estimation

In the final stage I estimate the parameters θ_k and θ_l while recovering the Controlled Markov Process outlined in Definition 2. This process is the primary object of interest in this paper.

Proposition 2. ***(Recovering Productivity)***
Using the cost minimization assumption, productivity can be recovered as:

$$z = constant + (1 - \hat{\theta}_m)m - \theta_k k - \theta_l l$$

Proposition 2 suggests an estimation strategy for the remaining parameters. Given data on R&D spending, age since listing and productivity I can regress productivity on its lag, age since listing and R&D as outlined in Definition 2. The estimates of this regression are conditional on the choice of parameter estimates, θ_k and θ_l. The residuals of this regression are $\varepsilon + \sqrt{\hat{\sigma^2}(z, r, a)}\xi$. This method concentrates out the parameters from the stochastic process to reduce the minimization to two dimensions. This suggests the following moment condition:

$$\mathbb{E}[(\varepsilon' + \sigma(z, r, a)\xi') \otimes \mathbf{w}]$$

Where $\mathbf{w}$ is a vector of instruments which are predetermined. Any variables in the information set $\mathcal{I}$ are appropriate to use. In practice, after concentrating out the Markov Process parameters I will use lagged values of k and l as instruments. The production function is then just identified.

In principle all functions of lagged variables are valid instruments. I could include many additional instruments. This complicates estimation considerably for the following reasons. The first stage estimator, based on an identity weighting matrix yields consistent, but inefficient parameter estimates. In order to find efficient estimates it is necessary to use at least a two stage estimation procedure. Given that I am bootstrapping the standard errors this adds to computation time and also leads to complications with the optimization routine. It is not clear that there is much additional information in further lags of capital and labour to be used in estimation. This is because there is little year to year variation in these inputs. As well, although using the optimal weighting matrix will make the estimates more efficient, it also may bias the results in small samples.[9]

I compute standard errors using a block bootstrap method. To create a bootstrap sample I draw firms with replacement from the sample. If a firm is drawn then I add their entire history of data to the bootstrap sample. I continue adding firms to the sample until the bootstrap sample size, NT^b, is equal to or greater than the actual sample size. For each bootstrap sample I estimate the model. I use these estimates to find bootstrap standard errors. To determine significance I construct symmetric bootstrap confidence intervals. If the 99% confidence interval for a parameter does not overlap with zero then I say this parameter is signficant at the 1% level and so on for 5% and 10%.[10] I use 999 bootstrap samples and visually inspect that the distribution has converged.

3. Data

3.1 Data Sources

The main data source used in this paper is the Standard and Poor's Compustat Fundamentals Annual database. These data cover the substantially long period of 1950 to 2017. Since the data are constructed from financial statements of publicly traded companies, there are well defined measures of sales, cost of goods sold (COGS), R&D expenditures, capital expenditures and employment. The share of publicly traded firms is small but their economic importance is significant since they tend to be very large.[11]

I augment the Compustat data with the NBER-CES Manufacturing Industry Database (NBER-CES), which contains information on price deflators at the 4-digit NAICS industry level. This dataset contains investment and materials deflators as well as average payroll. I limit my paper to the study of manufacturing firms so that I can make use of this rich dataset. I also use 3-digit NAICS industry level price deflators from the BEA to deflate sales.

I make no claim that this dataset is representative. In fact, I have made choices in the sample selection process with the aim of obtaining a clean sample at the expense of a representative one. This has partly been due to data restrictions and

[9]To be transparent, the main reason I opted for a just identified model was the computational burden. Specifically, the fact that the minimization routine often returned corner solutions for many of the bootstrap samples. A just identified model has a clear structural interpretation and is much simpler to estimate. The cost is not being able to test overidentifying restrictions.

[10]In order to construct bootstrap p-values I would need to first run the bootstrap to determine the standard error, then re-run the bootsrap to find the t-statistics.

[11]See Davis et al. (2006) for a discussion of the pros and cons of using a sample of publicly listed firms.

partly due to a belief that understanding the effect of R&D on productivity in the US manufacturing sector is interesting in its own right. My methods can be used on data from different countries and sectors provided a clean sample is obtained. Any similarities and differences across time, geography and industry will be interesting. By focusing on manufacturing, I can use the most accurate price deflators. This reduces, but does not eliminate measurement error. My dataset is large enough to obtain precisely estimated parameters. These estimates paint a clear picture of the post entry dynamics of publicly listed manufacturing firms.

In 1975, accounting standards in the United States changed so that R&D expenditures could no longer be expensed. These expenditures had to be recorded in the period in which they incurred. This means that, prior to 1975, missing observations for R&D expenditures may not reflect that the firm is a non-performer. This motivates my choice of starting the sample in 1976.[12] I end the panel in 2011 as this is the last year for which I have deflators from the NBER-CES database.

Over the sample period there have been some changes in the extensive margin of listed firms. In particular, the 1980s and 1990s saw a large influx of listings, which has then been reversed in the early 2000s. I will estimate my sample on multiple time periods to argue that my conclusions are not sensitive to this. Davis et al. (2006) show that the late 1980s and early 1990s were a period where more young, risky firms listed. Recent work by Gao et al. (2013) shows that this trend has reversed post-2000. The key takeaway is that the margin which determines listing have not been constant over this sample period. In Section 5 I explore the robustness of my approach to sample period selection.

3.2 Variable Creation and Sample Selection

Now I discuss some details about the creation of the dataset. More details are given in Section A.2. My measure of output is gross sales deflated by 3-digit NAICS code deflators from the BEA industry tables. I deflate R&D expenditures using 4-digit NAICS investment deflators. I also use these deflators to create measures of net investment which can be used to determine the capital stock using a perpetual inventory method.

Materials is difficult to measure using the Compustat data. My approach is to use data from the NBER-CES on the average wage at the 4-digit industry level. I substract this from the sum of cost of goods sold and selling general and administrative

[12]See Damodaran (1999) for a discussion of the accounting treatment of R&D expenditures in the United States.

expenses. Materials is therefore:

$$M_{i,j,t} = \frac{COGS_{i,j,t} + XSGA_{i,j,t}}{P^M_{i,t}} - \bar{w}_{j,t} N_{i,j,t}$$

To the extent that wages and price of materials vary across firms within 4-digit industries this approach will introduce some measurement error. This is unavoidable as the purpose of the Compustat data is to measure accounting variables, not economic ones.

Age since listing is computed as follows. I identify the first time a unique firm shows up in the database as the listing year of the firm. Following this, the firm ages organically every year it remains in the the data. The death year of a firm is the last year it appears in the database. Importantly, this may not signify the actual death of the firm. In this data, buyouts lead to significant sample attrition. This fact colours how I interpret the results, particularly with respect to the exit probability.

My target sample is innovative, publicly listed, manufacturing firms. I discuss my sample selection process in detail in Section A.3 . I restrict my attention to NAICS industries which are listed in the NBER-CES database. I want to capture aspects of the life cycle and therefore drop firms which were in the database in 1950 as I cannot find their age since listing.

I drop firms which have mergers and acquisitions activity beyond a certain threshold. Specifically, any firm which has a merger that affects sales by more than 20% is dropped from the sample. The results are not sensitive to this particular choice of threshold. The fact that many firms exit the sample due to acquisitions means that I need to be careful about how I interpret exit. Some firms will exit precisely because they are very productive; this is contrary to organic exits in administrative datasets.

Summary statistics are presented in Table 1. I separate my sample into innovative and non-innovative observations. The top panel shows information about firm year observations where R&D expenditure is greater than zero while the bottom panel shows firms-year observations where R&D spending is zero. The non-performing firms are, if anything, slightly larger, though not by very much. Looking at Sales it is clear that the distributions overlap as the 90th percentile R&D performing firm is larger than the 90th percentile non-performing firm.

The main takeaway from Table 1 is that the firms in my sample are very large. Average sales for R&D performers is $384 million 2010 USD and average number of employees $1,722$. The capital stock is also large and roughly 30% of sales. The average materials share is 0.75. Notice that I have not required this value to be between 0 and 1. I have used average payroll data to construct my materials measure,

Table 1: Summary Stats

	Mean	St. Dev	10th	50th	90th
Sales	384.08	753.59	6.56	89.87	1116.10
Capital	111.77	249.52	1.15	21.73	305.25
Employment	1722.43	2895.86	58.00	538.00	4900.00
Materials	256.49	507.32	4.48	62.97	727.18
Mat. Share	0.75	0.33	0.41	0.69	1.23
R&D	13.83	9.90	3.00	11.00	28.00
Age	14.83	50.36	0.20	2.66	28.85
NT	17,637	17,637	17,637	17,637	17,637
Sales	402.71	705.29	10.83	142.71	1092.29
Capital	143.34	290.10	2.07	42.34	386.38
Employment	1809.53	2729.71	66.00	760.00	4900.00
Materials	270.05	468.40	7.18	91.62	725.12
Mat. Share	0.69	0.22	0.45	0.68	0.92
Age	14.17	10.25	3.00	12.00	29.00
NT	9,906	9,906	9,906	9,906	9,906

All values in millions of constant 2010 USD except for age since listing, which is measured in years and employment which is measured in number of employees.

which induces measurement error. I chose to only exclude values which were extreme outliers (above the 99th percentile). The 90th percentile firm has 4,900 employess, which is very large.

R&D spending varies throughout the distribution. The distribution of R&D expenditure is right skewed, with a mean of $13.83 million USD and a median of $11 million. The 90th percentile firm performs almost 3 times as much R&D as the median firm. This shows that there are a few very large R&D performers in my sample.

The average age since listing is 14 years. Interestingly, bottom half of the age distribution is quite different for performersvs non-performers while the upper tail is similar. The 10th percentile R&D performer is a start-up while the 10th percentile non-performer has been listed for 3 years. This difference is also visible at the median. The median non-performer is 12 years old while the median R&D performer is 2.66. The averages do not though, mostly because the upper tail of the age distribution is similar for both performers and non-performers.

4. Results

4.1 Production Function Estimates

Table 2: Production Function Estimates

	OLS	AR(1)	Age	Ext. Margin	Linear	Full
θ_k	0.0844	0.2341	0.2391	0.2446	0.2703	0.2618
	(0.0009)	(0.0391)	(0.0395)	(0.0386)	(0.0403)	(0.0405)
θ_l	0.2275	0.4729	0.4855	0.4824	0.4785	0.4948
	(0.0011)	(0.0583)	(0.0605)	(0.0583)	(0.0617)	(0.0636)
θ_m	0.7188	0.6158	0.6158	0.6158	0.6158	0.6158
	(0.0011)	(0.0013)	(0.0013)	(0.0013)	(0.0013)	(0.0013)

These estimates are based on the full sample described in Section A.3 using the procedure outlined in Section 2.2. Standard errors are constructed using 999 bootstrap samples. All estimates are significant at the 1% level. The columns refer to different specifications of the function $\mu(z, a, r)$ as described in the text.

The production function estimates are presented in Table 2 . In the first column I show estimates from a simple OLS regression. The interpretation of the parameter values is as the elasticity of gross output with respect to the input in question. For instance, θ_k measures the responsiveness of gross output with respect to capital input. A 1 percent increase in the capital stock would increase output by 0.08% according to the OLS estimates.[13] It is difficult to compare my OLS estimates with the previous literature using Compustat data due to my focus on a gross production function. However, I am confident that these estimates are in the range of recent estimates using gross production functions on different samples.[14] The fact that studies using such different samples seem to find similar ranges for these elasticities is promising. Under OLS all the elasticities are estimated very precisely. However, due to endogeneity there are strong reasons to believe these estimates are biased.

The remaining columns of Table 2 provide production function estimates for various different models for TFP. These models refer to different specifications of the function $\mu(z, r, a)$.[15] I first note the stability of the estimates across these various specifications. The estimate for θ_m is by definition the same in all cases. This is because it is estimated as part of the first stage share equation. The materials elasticity is very precisely estimated. Controlling for endogeneity using my method decreases the materials elasticity relative to OLS. The direction of this change is the same as in Gandhi et al. (2011). It can be shown that failing to control for endogeneity leads to estimates of flexible inputs which are too high and estimates of fixed inputs which are too low.[16] This is exactly the pattern I observe in Table 2. The fact that θ_l is biased downward lends support to the assumption that labour is a fixed input. Recall that my dataset consists of manufacturing firms. These firms are likely to be unionized. In such an environment, labour is best represented as a dynamic input.

It is clear from the estimates in Table 2 that the technology exhibits increasing returns to scale. The sum of the coefficients on capital, labour and materials is 1.37 in the full model. Interestingly, the OLS results do suggest constant returns to scale, summing up to 1.03. The reason for this discrepancy is discussed in Section A.1 where I offer an alternative proof to Proposition 1 , extended to the case where firms have market power. This discussion makes clear why it is important to be careful interpreting revenue productivity. Under constant market power, the estimate for

[13] It is not appropriate to use the "share" interpretation of these estimates as I have not imposed constant returns to scale. Therefore, Euler's theorem is not applicable.

[14] Using OLS, Doraszelski and Jaumandreu (2013) find capital elasticites ranging from 0.051 to 0.109, labour elasticities ranging from 0.2177 to 0.335 and materials elasticities from 0.605 to 0.739 across 10 manufacturing industries in Spain. Gandhi et al. (2011) look at manufacturing data from Colombia and Chile and find a capital elasticity of 0.06 and 0.09 respectively, a labour elasticity of 0.26 and 0.2 respectively and a materials elasticity of 0.72 and 0.77 respectively.

[15] The specification of the skedastic function $\sigma^2(z, r, a)$ does not have an effect on the moment conditions used in estimation. Therefore, variations on this function do not affect the production function estimates.

[16] This is just an application of omitted variable bias where the omitted variable is productivity. Since, by assumption, $m(z, k, l)$ is increasing in z the correlation between z and m is positive. The discussion on p.96 of Wooldridge (2006) shows that in this case the bias will be positive.

the materials elasticity from the share equation identifies $\mathcal{M} \times \theta_m$ where $\mathcal{M} > 1$ is the markup and θ_m is the structural production function parameter. However, I argue that as long as we are willing to interpret the production function elasticities as being proportional to a constant markup, the interpretation of the rest of my results carries through, provided I am careful to note that these results are for revenue productivity. Given the constraints of most datasets (including mine), I need to do this.

De Loecker et al. (2019) find that the average markup has risen from roughly 1.3 to 1.5 over my sample period. Coupled with my results, this implies that constant returns to scale are indeed plausible for the structural production function. In essence, I would be identifying the fact that $\mathcal{M} \times (\theta_k + \theta_l + \theta_m) \approx 1.4$. Therefore, the interpretation of the results in the next section carry through to a situation where there is a constant mark-up.

One of the main takeaways from the De Loecker et al. (2019) paper is that markups are heterogeneous across firms. Once again I emphasize how this should change my interpretation of the production function elasticities, but not necessarily the conditional mean and skedastic function results. Consider that R&D spending may increase market power, leading to higher markups. This will show up in the residual of my gross production function. I am happy to interpret this as an "effect" of R&D on revenue productivity. I am not able to determine whether this effect is due to changes in the production function or demand. This is beyond the scope of most datasets, including mine. The fact that there is now some debate about the magnitude of the increase in market power over the sample period is beyond the scope of this paper.[17]

4.2 Conditional Mean

I now turn to the estimation of the conditional mean of the productivity process outlined in Definition 2. All of the models are special cases of the full model, which I define below:

$$\mu(z,a,r) = \mu_\chi \hat{P} + \begin{cases} \mu_0 + \mu_z z + \mu_a a + \mu_{z,a} za & \text{if } R = 0 \\ \mu_0^r + \mu_z^r z + \mu_a^r a + \mu_r^r r + \mu_{z,a}^r za + \mu_{z,r}^r zr + \mu_{a,r}^r ar & \text{if } R > 0 \end{cases}$$

[17] A recent symposium in the Journal of Economic Perspectives discusses these trends. In particular, Basu (2019) provides reasons to believe that the rise in market power has not been as large as De Loecker et al. (2019), Gutiérrez and Philippon (2017) and Barkai (2017) have emphasized. As well, Traina (2018) is critical of the approach used in De Loecker et al. (2019). This is a new and growing literature in macroeconomics, but the main takeaway that there have been changes in market structure in recent years is robust. There is debate about the magnitude.

Table 3: Conditional Mean Function Estimates

	AR(1)	Age	Ext.	Linear	Full
μ_0	-0.2570	0.0445	-0.0964	-0.2087	-0.1668
	(0.0922)	(0.1074)	(0.1035)	(0.0919)	(0.1008)
μ_z	0.5460	0.4525	0.5019	0.5625	0.5331
	(0.0211)	(0.0267)	(0.0323)	(0.0194)	(0.0306)
μ_a	-	-0.1274	-0.0875	-0.0727	-0.0862
	-	(0.0135)	(0.0132)	(0.0051)	(0.0125)
$\mu_{z,a}$	-	0.0455	0.0165	-	0.0152
	-	(0.0087)	(0.0110)	-	(0.0108)
μ_0^r	-	-	0.1917	0.1664	0.3379
	-	-	(0.0395)	(0.0206)	(0.0461)
μ_z^r	-	-	0.4305	0.5315	0.3800
	-	-	(0.0311)	(0.0212)	(0.0284)
μ_a^r	-	-	-0.1485	-0.0745	-0.1629
	-	-	(0.0159)	(0.0043)	(0.0158)
μ_r^r	-	-	-	-0.0564	-0.1042
	-	-	-	(0.0037)	(0.0102)
$\mu_{z,a}^r$	-	-	0.0591	-	0.0690
	-	-	(0.0095)	-	(0.0085)
$\mu_{z,r}^r$	-	-	-	-	0.0011
	-	-	-	-	(0.0057)
$\mu_{a,r}^r$	-	-	-	-	0.0202
	-	-	-	-	(0.0030)
μ_χ	11.5609	11.3382	11.4422	11.6104	11.6533
	(1.3583)	(1.3189)	(1.2861)	(1.3113)	(1.3289)

These estimates are based on the full sample described in Section A.3 using the procedure outlined in Section 2.2 and variations on the equation for $\mu(z, a, r)$ defined in the text. Standard errors are constructed using 999 bootstrap samples.

This model specifies different stochastic processes for R&D performers vs. non-performers. Of course, a firm can be a performer in one period and a non-performer in another. I also condition on exit for all models. This is done by including the exit probability from the probit estimated in Stage 2. The parameter μ_χ is estimated significantly and positively for all models. The interpretation of the coefficient μ_χ is as a semi-elasticity. Consider two firms, the first of which has an exit probability of 20% and the second which has an exit probability of 30%. The firm with the higher exit probability will on average have productivity that is 1.165% higher.[18] This may seem counter-intuitive as most models of firm dynamics would associate exit with low productivity. However, remember that firms exit my dataset due to both organic firm death and acquisitions. Firms are likely to be acquired precisely because they have a promising business model.[19]

4.2.1 Persistence

A meaningful measure of the persistence of productivity is the derivative of the conditional mean function with respect to lagged productivity. If this measure is positive then a firm which was productive in the past is likely to be productive in the future.[20] The first column of Table 3 shows the estimate of persistence for a simple AR(1) process. In this case, productivity is moderately persistent with an autocorrelation of 0.546. A key theme of my paper is non-linearity and heterogeneity. Simply adding a life cycle term to the model lowers the estimate of μ_z to 0.4525. However, it is clear that this masks heterogeneity over the life cycle. The coefficient $\mu_{z,a}$ is greater than 0 and significant. This means that as firms age, their productivity becomes more persistent. This is an interesting result, which makes intuitive sense. To the extent that productivity accounts for unmeasurable firm characteristics, it is expected that these characteristics are less likely to change as a firm matures.

One unmeasurable characteristic of a firm is its spending on intangible capital. Columns 3 and 4 of Table 3 add R&D to the model for $\mu(z, a, r)$ in different ways. First, look at column 3 (Ext.). I call this the extensive margin model as it uses the same process for both performing and non-performing firms. In this sense a firm makes a discrete choice about whether to perform R&D or not. The coefficient $\mu_{z,a}$

---

[18]This is found by multiplying 0.1 times $\hat{\mu}_\chi = 11.6533$.

[19]Recall that I have not imposed any restrictions on the polynomial which determines the exit probability. Even though this result seems to reject Assumption 3, a weaker version which states that the exit threshold is an unknown function of capital, labour and materials would still be appropriate.

[20]For example, persistence for an R&D performing firm in the full model is:

$$\frac{\partial \mu^r(z, a, r)}{\partial z} = \mu^r_z + \mu^r_{z,a} a + \mu^r_{z,r} r$$

for non-performing firms decreases in magnitude and is no longer significant. This suggests that increases in persistence as a firm matures are concentrated among R&D performers. The coefficient $\mu^r_{z,a}$ confirms this. It is significant and of greater magnitude to the one estimated in Column 2.

Now I consider Column 5 (Full) of Table 3 . For non-performers TFPR is moderately persistent and linear. This suggests that for these firms a simple AR(1) with a life cycle component would be a reasonable model. For R&D performers, things are more complex. First, unconditional persistence is lower, with $\mu^r_z = 0.38$. However, this masks heterogeneity over the life cycle. In particular, TFPR becomes more persistent as firms age. This is consistent with the pure life cycle model. There are no strong non-linearities with respect to R&D expenditures and productivity ($\mu_{z,r} = 0.0011$).

Another way to quantify these non-linearities is to consider that age since listing is measured in natural logarithms. Therefore, the persistence of a newly listed R&D performing firm is 0.38 in the full model. Compare this to a 10 year old firm, which has a persistence of 0.54 and a 15 year old firm which has a persistence of 0.57. This simple example makes it clear that the profile of persistence is concave over the life cycle. The first few years since listing see the biggest rise in persistence.

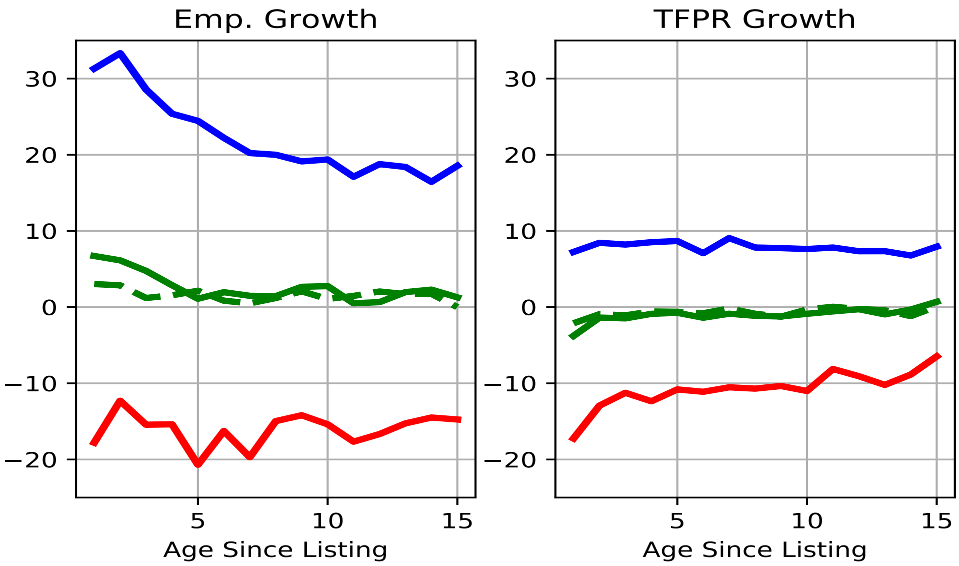
In sum, TFPR is moderately persistent in the sample. The key non-linearity has to do with age since listing. The more mature a firm gets, the more persistent their productivity is. This effect is driven by R&D performing firms. When the model is restricted to be linear in Column 4 (Linear) of Table 3 the coefficient μ^r_a increases substantially. The increase in persistence is concentrated early in the firm life cycle.

4.2.2 The Firm Life Cycle

Next I will consider what the results in Table 3 have to say about the firm life cycle. In particular, I am interested in the marginal effect on $\mu(z, a, r)$ of a change in a. The clearest way to see this is in Column 2 (Age) which abstracts from R&D spending. It is clear that average productivity declines over the life cycle ($\mu_a = -0.1274$). Reading across the table for the different models it is clear that this result is robust. Ceterus peribus there is a decline in revenue productivity over the first few years since listing for publicly traded manufacturing firms. This negative life cycle component is strongest for R&D performing firms as can be evidenced by the coefficient μ^r_a in the extensive margin model.

Turning to the full model, which is easiest to interpret, I conclude that there is a strong negative life cycle effect. Furthermore, it is non-linear. This has already been discussed in terms of persistence in the context of $\mu_{z,a} > 0$ for R&D performers. Another way to view this is that higher productivity firms do not see as strong a

Figure 1: Life Cycle Dynamics



*Weighted percentiles of employment (left panel) and TFPR (right panel) growth, using DHS measure and employment weights. Based on estimation sample.

decline in productivity over the life cycle. Of course, given the specification of the model this is just another way of interpreting the results from the previous section.

Additionally, the coefficient for $\mu^r_{a,r}$ is positive and significant. This means that firms which do more R&D at the intensive margin do not see their productivity decline as much over the life cycle. In this context, R&D can be seen as a way to mitigate the decline in productivity over the life cycle. Firms which are in R&D intensive sectors have a stronger decline and must offset this by doing more R&D .

I have been discussing the life cycle profile of TFPR for newly listed firms. Much of the recent literature on post-entry dynamics emphasizes the characteristics of employment growth rates for young firms. Figure 1 replicates a figure which has been used in this recent work.[21] I replicate the shape of the employment distribution over the life cycle, though not the magnitude. In particular, Compustat firms exhibit less dispersion and a lower average growth rate at all ages. However, the shape is consistent. I interpret this to mean that I am not capturing a random sample of firms, and these firms are more mature than the firms in the Decker et al. (2014) sample. However, the skewed right tail of employment growth is pronounced for young firms.

[21] I am replicating this figure using the method in Decker et al. (2014) which emphasizes that the average, variance and skewness of firm growth rates declines over the firm life cycle. I follow their approach and use weighted percentiles of the employment growth distribution.

Interestingly, this shape is not replicated for revenue productivity. The growth rate distribution for TFPR is relatively stable over the life cycle, excluding the first year after listing in which there is a skewed left tail. This means that the employment growth dynamics observed in the left panel do not appear to be driven in any simple way by TFPR dynamics. Two recent papers which study the life cycle dynamics of firms using US Census Bureau data deserve some consideration at this point. Alon et al. (2018) estimate a firm life cycle for average revenue labour productivity. They find decline growth rates over the life cycle. Of course my paper estimates revenue total factor productivity, which is a broader concept.

Pugsley et al. (2018) estimate the age covariance structure of employment, which they equate with productivity. They do not emphasize firm growth rates, but point out that the correlation between employment at different points in the life cycle does not decline to 0, suggestive of fixed effects in productivity. I have replicated the covariance structure of employment from this paper in Figure 2. The employment dynamics in the left panel show that the covariance between employment of a newly listed firm and a ten year old firm are still 0.7. This suggests that there is heterogeneity in firms at birth which does not disappear. The figure on the right hand side is constructed using my measure of TFPR. A similar conclusion can be drawn. Differences in revenue productivity do not decay, replicating the employment dynamics. My results are strongly suggestive of a fixed effect of productivity which is determined at listing.

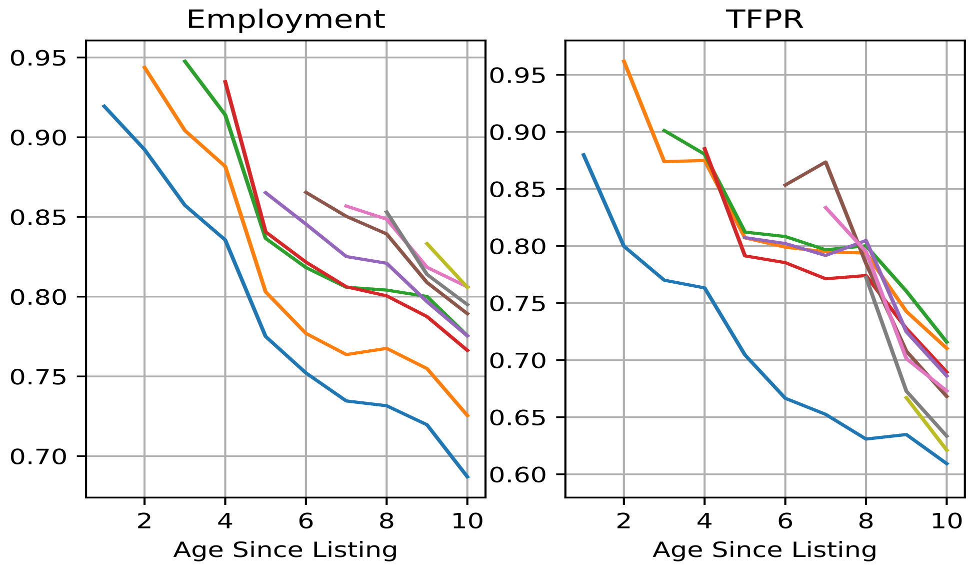
4.2.3 The Return to R&D

In this section I analyse the effect of R&D on productivity through the function $\mu(z, a, r)$. Starting from Column 3 (Ext.) of Table 3 I note that the unconditional mean of productivity for R&D performing firms is higher than that of non-performing firms. That is, μ_0 is not statistically different from zero while $\mu_0^r > 0$ across all specifications. Therefore, the first way that doing R&D can be beneficial to a firm is by increasing the average productivity.

The second robust effect is that productivity is less persistent for R&D performers. This is clearly seen in Column 3 (Ext.) of Table 3. The persistence for non-performers is 0.5019, while that of performers is 0.4305. The main reason for this is that persistence is much more non-linear in age for R&D performers as discussed above. These non-linearities show up as a lower unconditional average productivity.

Next I move to Column 5 (Full). Somewhat surprisingly the intensive margin effect of doing more R&D is negative with $\mu_r^r = -0.1042$ and significant. This is counterintuitive, however, the result needs to be considered in combination with the

Figure 2: Covariance Structure



*Covariance structure of residual employment (left panel) and residual TFPR (right panel) growth after regressing on year and industry dummies. Based on estimation sample.

findings on non-linearity and the skedastic function. I will address this result later. For now I emphasize that this effect is non-linear. In particular, firms seem to "get better" at R&D as they age. Furthermore, the unconditional average productivity is higher for R&D performers.

4.3 Skedastic Function

Next I will introduce the second part of the stochastic process introduced in Definition 2. This I will call the skedastic function, denoted $\sigma^2(z, a, r)$. The structure of this function is similar to the conditional heteroskedasticity function introduced in the previous section.

$$\sigma^2(z,a,r) = \sigma_\chi^2 \hat{P} + \begin{cases} \sigma_0^2 + \sigma_z^2 z + \sigma_a^2 a + \sigma_{z,a}^2 za & \text{if } R = 0 \\ s_0^2 + s_z^2 z + s_a^2 a + s_r^2 r + s_{z,a}^2 za + s_{z,r}^2 zr + s_{a,r}^2 ar & \text{if } R > 0 \end{cases}$$

Note that the function is non-linear and allows productivity, age and R&D spending to effect the volatility of the Markov Process. This process can differ between R&D performing vs. non-performing firms. I also control for exit, although this term turns out to be insignificant.

Table 4: Skedastic Function Estimates

	AR(1)	Age	Ext.	Linear	Full
σ_0^2	0.6550	0.8768	0.8356	0.7362	0.8141
	(0.1059)	(0.1185)	(0.1293)	(0.1054)	(0.1360)
σ_z^2	-0.0773	-0.1482	-0.0988	-0.0541	-0.1020
	(0.0507)	(0.0459)	(0.0532)	(0.0422)	(0.0528)
σ_a^2	-	-0.0873	-0.0779	-0.0566	-0.0788
	-	(0.0143)	(0.0193)	(0.0077)	(0.0191)
$\sigma_{z,a}^2$	-	0.0360	0.0305	-	0.0293
	-	(0.0127)	(0.0184)	-	(0.0186)
s_0^2	-	-	0.0956	0.0811	0.1442
	-	-	(0.0786)	(0.0454)	(0.0866)
s_z^2	-	-	-0.1818	-0.1700	-0.2470
	-	-	(0.0534)	(0.0494)	(0.0538)
s_a^2	-	-	-0.0985	-0.0308	-0.0758
	-	-	(0.0191)	(0.0059)	(0.0209)
s_r^2	-	-	-	-0.0744	-0.1606
	-	-	-	(0.0066)	(0.0136)
$s_{z,a}^2$	-	-	0.0444	-	0.0392
	-	-	(0.0147)	-	(0.0154)
$s_{z,r}^2$	-	-	-	-	0.0201
	-	-	-	-	(0.0065)
$s_{a,r}^2$	-	-	-	-	0.0289
	-	-	-	-	(0.0040)
s_χ^2	-3.4297	-3.8250	-4.0237	-2.8993	-3.4577
	(2.1756)	(2.0718)	(2.0183)	(1.9767)	(2.0444)

*

4.3.1 The Life Cycle, Productivity and Uncertainty

The results for the skedastic function are given in Table 4. Looking at Column 1 (AR(1)), lagged productivity is not significant in determining the volatility of TFPR. This suggests an AR(1) process as a baseline from which to compare future models. The unconditional standard deviation is 0.81. I do not know of any other papers with a similar methodology which focus on volatility, making it difficult to compare this benchmark to previous literature.[22]

Moving to Column 2 (Age) of Table 4 a robust result emerges. The volatility of TFPR is declining over the life cycle. This is true in all models which include age. To put this in perspective, a newly listed firm has an unconditional volatility of 0.94. Ignoring non-linearities, a firm that has been listed for 10 years on average have a volatility of 0.82. A 15 year old firm will have a volatility of 0.8.[23] This shows that volatility declines non-linearly over the life cycle, with the strongest effects in early years. The magnitude of this life cycle component is similar for R&D performers as well as non-performers as can be seen in Column 3 (Ext.).

An important non-linearity emerges with the parameter $\sigma^2_{z,a} = 0.036$. This can be interpreted as follows. We have shown that uncertainty falls as the firm ages. This effect is "smaller" for more productive firms. The life cycle profile of uncertainty is less steep for the most productive firms. The results for both R&D performers and non-performers suggest that more productive firms have lower volatility ($\sigma^2_z < 0$). [24] This suggests that the more productive a firm is, the less volatile their productivity is. The fact that this result is stronger for R&D performing firms is interesting. I am not aware of any prior literature which studies this.

[22] Castro et al. (2015) look at cross sectional variation in TFPR volatility using administrative data from the Census of Manufactures using gross output measures. However, their methodology for recovering the residuals is different from mine. As such, I am not comfortable making comparisons. Their point estimate for volatility is 0.26, considerably lower than mine. This may be for several reasons: they condition on firm size, firm fixed effects and firm age dummies when recovering TFP residuals and I use publicly traded firms.

[23] The calculations are as follows.

$$\sqrt{0.8768} = 0.94 \qquad \sqrt{0.8768 - 0.0873 * \ln(10)} = 0.82 \qquad \sqrt{0.8768 - 0.0873 * \ln(15)} = 0.8$$

[24] Consider the equations:

$$\frac{\partial \sigma^2(z,a,r)}{\partial a} = -0.0873 + .036z \qquad \frac{\partial \sigma^2(z,a,r)}{\partial z} = -0.1482 + .036a$$

4.3.2 R&D and Uncertainty

Now I wish to consider the effects of R&D spending on uncertainty. These effects occur through the extensive margin, the intensive margin and non-linearities with age and lagged productivity. First, look at the full model in Column 5 of Table 4. More R&D spending is associated with lower volatility through the term $s_r^2 = -0.1606$. This suggests that, in addition to the effects of R&D on the conditional mean function, one benefit of R&D is to reduce uncertainty about future productivity. To the extent which firm owners are risk averse, there is a definite benefit in reducing uncertainty.

There are two important non-linearities with respect to R&D expenditures. First, $s_{z,r}^2 = 0.0201$, which shows that more productive firms do R&D projects which are more uncertain. Secondly, $s_{a,r}^2 = 0.0289$ which shows that more mature firms do more uncertain R&D projects. This suggests that more mature firms do R&D that is more volatile. This has interesting implications for a recent literature which stresses the heterogeneity of R&D performance across different types of firms. Akcigit and Kerr (2018) argue that large firms focus more on external innovation, which is inherently less risky. This leads the variance of the growth rate to decrease with size. They do not explicitly discuss the firm life cycle.

It is now possible to discuss the negative marginal effect of R&D spending. For a typical firm, the extensive margin effect of doing R&D is positive ($\mu_0^r > 0$), the marginal effect of the next R&D dollar is negative ($\mu_r^r < 0$) and the marginal R&D dollar reduces uncertainty ($s_r^2 < 0$). Furthermore, it is important to note that we are identifying ex-post "returns" to R&D .[25] Given the uncertainty involved in R&D investments it is important to note that there can be large differences between ex-ante and ex-post returns. The effect of the marginal R&D dollar is both decreasing volatility and decreasing the mean of future productivity. To the extent that managers want to avoid risk, this is a meaningful trade-off. Furthermore, the fact that the effects of R&D on productivity are highly non-linear suggests that for many firms the return to the marginal dollar of R&D is positive. The main point I wish to emphasize is that my results, though surprising, have a reasonable interpretation in a model of R&D investment.

5. Additional Results

This section discusses some additional results. See Section A.4 for the tables referred to in this section. My analysis here shows that the main conclusions of my paper

[25] I put "returns" in quotes to emphasize that I am not identifying a comparable return with the literature following Hall et al. (2010) and Griliches et al. (1979).

are not sensitive to industry or time variation.

5.1 Industry and Sample Period Differences

5.1.1 Inter-Industry Differences

I begin by estimating the full model on a variety of industries. I only run the estimation on industries with more than $1,000$ firm-year observations. The industries which I am able to run the model on are Machinery Manufacturing (NAICS-334), Computer and Electronics Manufacturing (NAICS-334), Miscellaneous Manufacturing (NAICS-339), Wood, Paper and Printing (NAICS-321,322,323) and Electrical Equipment and Appliances (NAICS-335).[26]

Turning to the production function estimates in Table A.4.1 some patterns emerge. First, there is a range of estimates for the materials elasticity. All estimates for θ_m are precise, but they range from 0.5339 for Electrical to 0.8693 for Computers. This suggests that even within the narrow manufacturing industries there are variations in output elasticities. In particular, the Computer Manufacturing industry has very low capital and labour elasticities. This industry is characterized by decreasing returns to scale and a high importance of materials/intermediates in production. On the other hand, machinery manufacturing is characterized by increasing returns to scale. Electrical manufacturing is less dependent on materials and more dependent on labour.

Given such wide variation in production function elasticities, it is interesting to look at which conclusions about the productivity process are robust. I start by looking at the results for the conditional mean function in Table A.4.2. The first thing to note is that the estimates for computers manufacturing are not precisely identified. I will therefore make no conclusions based on this column.

Looking at the persistence measures for the remaining industries, μ_z, it is clear that they are all moderately persistent, ranging from 0.3721 to 0.6706. There is definitely a range of values here, but the main takeaway is the same. As well, the result that $\mu_a < 0$ is robust as well. The intensive margin result that R&D expenditures and productivity are negatively related is also robust and significant across all industries. The qualitative conclusions I would draw from the individual industries match the ones I made from the aggregate analysis. This is interesting in light of the very different production function elasticities.

Next I examine the non-linearities. For non-performing firms, the coefficient, $\mu_{z,a}$ is either positive or insignificant. In the full sample it was insignificant. For R&D performing firms the non-linear terms $\mu_{z,a}$ and $\mu_{z,r}$ are positive when precisely

[26] I tried to estimate the model on Chemical Manufacturing (NAICS-325) but my routine encountered repeated corner solutions.

estimated. This matches the full sample. In all but Computers the term $\mu_{a,r}$is positive. The non-linearities are consistent with the full sample, though in some instances they are not precisely estimated.

Moving to the Skedastic Function estimates in Table A.4.3 first note that the values for σ_0^2 vary across sectors. This means that the baseline level of uncertainty has a distribution. The value s_0^2 is not precisely estimated. For non-performing firms, only machinery has precise estimates and these mirror the estimates of the full sample. That is volatility is declining in age and lagged productivity. The life cycle component is also estimated precisely for Miscellaneous Manufacturing.

The result that volatility is declining in R&D expenditures is robust across industries. The result holds for Miscellanious and Materials Manufacturing and is insigificant but negative in the rest of the industries. The non-linearities are not very precisely estimated, except in Machinery Manufacturing where they match the broad pattern given in the full sample.

It is noteworthy that, despite variation in the production function elasticities, the conclusions I draw from the conditional mean and skedastic functions are essentially the same. TFPR is moderately persistent and non-linear. Productivity declines over the life cycle. R&D expenditure is more effective over the life cycle. Estimating the model on subsets of the data leads to issues around statistical significance and in the case of Chemicals Manufacturing issues around identification. However, I believe that the narrative around R&D as being non-linear, uncertain and complex follows through.

5.1.2 Time Period Differences

Now I estimate my model on two different time periods. The first period I look at is the decade from 1986-1995 and the second is from 1996-2005. I tried to estimate on the period 1976-1985 but there were issues with corner solutions. Furthermore, there was not enough data for the period from 2006-2012 to obtain precise estimates.

Looking at the production function estimates, a pattern emerges. First, the materials elasticity is precisely estimated and grows from 0.5936 to 0.7308 between the two time periods. The data suggest fundamental changes in the production process. This increase in materials elasticity could be due to increases in market power as outlined in De Loecker et al. (2019). I am using a subset of their data, which documents a pronounced rise in mark-ups over this period. Since one interpretation of this elasticity is as the markup times the structural elasticity, this needs to be taken in to account.

Another structural change could be due to outsourcing parts of the manufacturing process. Both the capital and labour elasticities have fallen over the sample.

However, the fall in the labour elasticity is pronounced and striking. My sample reflects a period where employment in American manufacturing has been in secular decline. The rise of intermediate goods and the fall of labour as an important factor of production show up in my sample. Notice that the labour elasticity for the full sample is 0.4948. This suggests that the fall in this value has been occurring over the full sample period. This would match the trends observed in aggregate data on domestic manufacturing employment.

Given such striking trends in the shape of the production function, it is useful to examine the estimated TFPR process across sample periods. For non-R&D performing firms the conditional mean of productivity is not drastically different across time. Revenue productivity is more persistent in the 1986-1995 period. Furthermore, there appears to be a stronger non-linearity between age since listing and lagged productivity in the 1996-2010 period. This may be related to the fact that selection is not as strong. Notice that μ_{χ} is decreasing over time. This means that the correlation between exit probability and productivity is less pronounced in the latter part of the sample.

Looking at R&D performers, I first note that none of the signs of the coefficients change. The qualitative story that I would tell about the conditional mean function is unchanged. However, there are some clear changes in magnitudes. In particular, $\mu_{z,a}$, the non-linearity between age since listing and productivity, is increasing for R&D performers as well. As well, the coefficient μ_a declines in magnitude. This suggests that there has been a shift in the nature of the firm life cycle over time. This change in the firm life cycle can also be seen in a decrease in the complementarity between R&D and age since listing.

Turning to the skedastic function, I first emphasize that the signs of the coefficients are consistent across the two time periods. The absolute value of the variance is declining. This is consistent with a literature which has shown that the volatility of publicly listed firms has been declining over time. These studies usually emphasize changes in the variance of employment growth.

The life cycle effect on the variance of productivity is negative in both cases, but the magnitude is decreasing over time. It is clear that there is a quantitative shift in the nature of the life cycle over time. What is unclear from my work is whether this is specific to my sample or whether this is a more general conclusion. The non-representativeness of publicly listed firms is an issue here. My work would need to be extended to more representative datasets to identify whether these trends are robust.

5.2 Robustness

In this section I consider further checks of the robustness of my results.

5.2.1 Robustness to Exit

It has become common in the literature on production function estimation not to control for exit. The argument is that controlling for exit does not have a large effect on the estimates of the production function elasticities. Given my focus on a Controlled Markov Process it was not clear whether controlling for exit would have a significant impact on my estimates of this process. Furthermore, given that my sample features attrition primarily due to mergers and acquisitions, it is not clear that the common rules of thumb from the literature apply. Lastly, my estimation procedure is a hybrid of some of the methods which have been used in the past.

For these reasons it is prudent to examine the effects of controlling for exit on my results. I find that failing to control for exit does in fact have a substantial impact on the production function estimates in Table A.4.7. Despite substantial differences in the estimates for the capital and labour elasticity, the results for both the conditional mean function and the skedastic function are robust. Given that the Controlled Markov Process is the focus of my paper I take comfort in these findings.

5.2.2 Robustness to Materials Measure

One potential area of concern with my data is the use of a materials measure which has measurement error. This is a problem in all datasets. In my case, I need to use data on average wages at the industry level as well as firm level accounting data to back out a measure for materials that matches the economic definition. To the extent which a firms average wages differ from the industry average, this introduces measurement error.

Two approaches have been used in the literature to deal with this. The first approach is to use a sparsely populated variable in Compustat call Expenditure on Labour (XLR). Unfortunately, after sample selection, this variable is only present in a few hundred observations. This is not enough to run my estimation procedure and receive significant results.

Another recent approach has been to assume that Cost of Goods Sold (COGS) is synonymous with materials. Cost of Goods Sold includes what economists traditionally view as materials as well as the expense of labour. However, not all types of labour are included in COGS. This is why, in Column 4 of Table A.4.7 the sum of the "materials" and capital elasticities are so low. It is unclear whether this approach leads to a cleaner measure of productivity than the one I have used in the

Table 5: Lag Structure

μ_{-1}	$\mu_{r_{-1}}$	$\mu_{r_{-1},z}$	$\mu_{r_{-1},a}$
0.0352	-0.0681	0.0061	0.0105
(0.0136)	(0.0169)	(0.0075)	(0.0046)
σ^2_{-1}	$\sigma^2_{r_{-1}}$	$\sigma^2_{r_{-1},z}$	$\sigma^2_{r_{-1},a}$
0.0264	-0.0756	0.0325	0.0060
(0.0205)	(0.0382)	(0.0154)	(0.0087)

*

main part of the text.

With that in mind, when looking at the conditional mean function estimates in Column 4 of Table A.4.8, the main difference seems to be that R&D and productivity are positively related at the intensive margin. It is unclear why this is the case. Keep in mind that the notion of residual productivity is very different between the baseline and the COGS results.

Turning to the skedastic function results, the main impact is on statistical significance. In the instances where the signs change all of the new results are statistically insignificant. It remains for future work to determine which measure of materials is the most correct one. My stance in this paper is that COGS is an accounting measure which does not reflect materials. There is measurement error inherent in both approaches. Estimating a gross production function has been shown to be quite robust, even to large variations in the production function estimates.

5.2.3 Further Lags of R&D

As a last robustness measure I consider further lags of R&D in the Controlled Markov Process. These results are summarized in Table 5. The estimates for the production function as well as the rest of the conditional mean and skedastic function are relegated to the Appendix. Those tables show that neither the production function estimates or the stochastic process estimates are changed in a major way by including further lags. The main effect is that the impact of one period lagged R&D is lowered by including a further lag, which is to be expected.

Turning to the top panel of Table 5 I examine the conditional mean effects of lagged R&D . The coefficient μ_{-1} is positive and significant. This means that 2 years lagged R&D has a positive unconditional increase on mean productivity. Similar to the results from earlier in the paper, lagged R&D has a negative intensive margin effect on productivity, which is once again puzzling. Lastly, the complementarity between age since listing and lagged R&D is positive and significant. This means that firms "get better" at R&D as they age, which is consistent with the main results of the text. It doesn't appear that there are any new insights from including lagged R&D in the model and the other insights are largely robust to this extension.

Looking at the bottom panel, which explores the effects of lagged R&D on the skedastic funciton I find that lagged R&D reduces the volatility of TFPR. This is complementary to my baseline results. Lastly, there is a positive complementarity between lagged productivity and lagged R&D . More productivity firms do more uncertain R&D . There is an insignificant complementarity between age since listing and lagged R&D .

6. Conclusion

In this paper I have used a novel identification strategy to examine the non-linearities and uncertainties in the relationship between revenue productivity and R&D expenditure over the firm life cycle. I have discovered that non-linearity is important in the effectiveness of R&D . More productive and more mature firms are better at transforming R&D expenditures into productivity. I have also find that there is a positive extensive margin effect on mean productivity of doing R&D , yet the intensive margin consistently yields negative coefficients. This puzzle could possibly be resolved by noting that more R&D spending consistently lowers uncertainty.

The main takeaway from my work is that the process of converting R&D spending into revenue productivity is uncertain and non-linear. My approach is flexible and requires few assumptions to be valid. This methodology can be extended to different industries, time periods and geographies. It can also be extended to identify spillovers of aggregate R&D spending. I do this in another paper.

References

Acemoglu, D., Akcigit, U., Alp, H., Bloom, N., and Kerr, W. (2018). Innovation, reallocation, and growth. *American Economic Review*, 108(11):3450–91.

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

Akcigit, U. and Kerr, W. R. (2018). Growth through heterogeneous innovations. *Journal of Political Economy*, 126(4):1374–1443.

Alon, T., Berger, D., Dent, R., and Pugsley, B. (2018). Older and slower. *Journal of Monetary Economics*, 93:3–24.

Arkolakis, C., Papageorgiou, T., and Timoshenko, O. A. (2018). Firm learning and growth. *Review of Economic Dynamics*, 27:146–168.

Barkai, S. (2017). Declining labor and capital shares. *University of Chicago*.

Basu, S. (2019). Are price-cost markups rising in the united states? a discussion of the evidence. *Journal of Economic Perspectives*, 33(3):3–22.

Castro, R., Clementi, G. L., and Lee, Y. (2015). Cross sectoral variation in the volatility of plant level idiosyncratic shocks. *The Journal of Industrial Economics*, 63(1):1–29.

Clementi, G. L. and Palazzo, B. (2016). Entry, exit, firm dynamics, and aggregate fluctuations. *American Economic Journal: Macroeconomics*, 8(3):1–41.

Cohen, W. M. (2010). Fifty years of empirical studies of innovative activity and performance. In *Handbook of the Economics of Innovation*, volume 1, pages 129–213. Elsevier.

Cooley, T. F. and Quadrini, V. (2001). Financial markets and firm dynamics. *American economic review*, 91(5):1286–1310.

Damodaran, A. (1999). Research and development expenses: Implications for profitability measurement and valuation.

Davis, S. J., Haltiwanger, J., Jarmin, R., Miranda, J., Foote, C., and Nagypal, E. (2006). Volatility and dispersion in business growth rates: Publicly traded versus privately held firms [with comments and discussion]. *NBER macroeconomics annual*, 21:107–179.

De Loecker, J., Eeckhout, J., and Unger, G. (2019). The rise of market power and the macroeconomic implications.

Decker, R., Haltiwanger, J., Jarmin, R., and Miranda, J. (2014). The role of entrepreneurship in us job creation and economic dynamism. *Journal of Economic Perspectives*, 28(3):3–24.

D'Erasmo, P. (2007). Investment and firm dynamics. *Available at SSRN 994663*.

Doraszelski, U. and Jaumandreu, J. (2013). R&d and productivity: Estimating endogenous productivity. *Review of Economic Studies*, 80(4):1338–1383.

Gandhi, A., Navarro, S., and Rivers, D. (2011). On the identification of production functions: How heterogeneous is productivity? Technical report, CIBC Working Paper.

Gao, X., Ritter, J. R., and Zhu, Z. (2013). Where have all the ipos gone? *Journal of Financial and Quantitative Analysis*, 48(6):1663–1692.

Griliches, Z. et al. (1979). Issues in assessing the contribution of research and development to productivity growth. *Bell Journal of economics*, 10(1):92–116.

Gutiérrez, G. and Philippon, T. (2017). Declining competition and investment in the us. Technical report, National Bureau of Economic Research.

Hall, B. H., Mairesse, J., and Mohnen, P. (2010). Measuring the returns to r&d. In *Handbook of the Economics of Innovation*, volume 2, pages 1033–1082. Elsevier.

Haltiwanger, J., Jarmin, R. S., and Miranda, J. (2013). Who creates jobs? small versus large versus young. *Review of Economics and Statistics*, 95(2):347–361.

Isaacson, W. (2014). *The innovators: How a group of inventors, hackers, geniuses and geeks created the digital revolution*. Simon and Schuster.

Jovanovic, B. (1982). Selection and the evolution of industry. *Econometrica: Journal of the Econometric Society*, pages 649–670.

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

Olley, G. S. and Pakes, A. (1996). The dynamics of productivity in the telecommunications equipment. *Econometrica*, 64(6):1263–1297.

Pugsley, B. W., Sedlacek, P., and Sterk, V. (2018). The nature of firm growth.

Schumpeter, J. A. (1942). *Capitalism, socialism and democracy.* Routledge.

Sterk, V., Sedlacek, P., and Pugsley, B. (2018). The nature of firm growth.

Traina, J. (2018). Is aggregate market power increasing? production trends using financial statements. *Production Trends Using Financial Statements (February 8, 2018).*

Wooldridge, J. M. (2006). *Introductory econometrics: A modern approach.* Nelson Education, 4 edition.

A. Appendices

A.1 Proofs

A.1.1 Proof of Proposition 1

For a variable input, M, the cost function is:

$$VC(z, K, L, Y) = \min_{M} \; W_m M$$

Subject to:

$$Y \geq \mathbb{E}[e^{z+\varepsilon} K^{\theta_k} L^{\theta_l} M^{\theta_m}]$$

Which leads to a first order condition of:

$$\frac{W_m}{\lambda} = \theta_m \mathcal{E} K^{\theta_k} L^{\theta_l} M^{\theta_m - 1}$$

Where λ is interpreted as the marginal cost. Under Assumption 1 this is constant and equal to P. Now multiply each side by M and divide by Y to get:

$$\frac{W_m M}{PY} = \frac{\theta_m \mathcal{E}}{e^{\varepsilon}}$$

Taking natural logs yields:

$$\ln S = \ln \mathcal{E}\theta_m - \varepsilon$$

A.1.2 Proof of Proposition 2

Take the first order condition from the cost minimization problem above and denote the real price of materials as w_m. Therefore:

$$M(z, K, L) = (\frac{\theta_m \mathcal{E} e^z K^{\theta_k} L^{\theta_l}}{w_m})^{\frac{1}{1-\theta_m}}$$

Take natural logs and simplify to find:

$$(1 - \theta_m)m = \ln \theta_m + \ln \mathcal{E} - \ln w_m + \theta_k k + \theta_l l + z$$

Simplifying gives the parametric inversion used in the paper:

$$z = -(\ln \theta_m + \ln \mathcal{E} - \ln w_m) + (1 - \theta_m)m - \theta_k k - \theta_l l$$

A.1.3 Alternative Proof of Proposition 1 with Market Power

In the more general case where there is an inverse demand curve, $P(Y)$, the marginal revenue will be the usual formula:

$$MR(z,K,L,Y) = \frac{\eta - 1}{\eta} P(Y)$$

Where η is the negative of the price elasticity of demand. So long as η is a constant, the optimal pricing rule will be:

$$P(Y) = \frac{\eta}{\eta - 1} MVC(z,K,L,Y)$$

Recalling the first order condition from cost minimization includes λ, the shadow value on the constraint, which is just marginal cost, the optimal pricing rule will set:

$$\lambda = \frac{\eta - 1}{\eta} P$$

Inserting this into the share equation from Proposition 1 generalizes the result to cases where the firm has goods market power. The share equation now becomes:

$$\ln S = \ln \mathcal{E}\mathcal{M}\theta_m - \varepsilon$$

Where $\mathcal{M}$ denotes the markup. This makes it clear that in essence we are identifying the markup times the materials elasticity from the share equation.

A.2 Variable Construction

Inevitably I have had to make difficult choices in preparing the data. These choices are outlined in this section.

A.2.1 Age Since Listing

I define the age since listing as follows. The first time that an observation for either sales or employment is present in the Compustat data I flag that as the birth year for a firm with a given GVKEY. From then on I let the firm age organically from year to year. Any firm which first appears in 1950 (the first year the data is available) I flag as right censored. For most of the analysis these firms will be dropped as I am interested in studying the post listing dynamics of firms.

An exit from the Compustat sample may be due to firm death, going private or a merger/buyout with a larger firm. I will discuss how I deal with mergers and acquisitions during sample selection. Unfortunately, there is no way to tell if a firm is going private.

A.2.2 Capital Stock

I use a perpetual inventory method to construct the capital stock. Capital at firm i in industry j evolves according to:

$$K_{i,j,t+1} = (1-\delta)K_{j,t} + I_{j,t}$$

I make use of the NBER-CES Manufacturing Industry Database which has industry level investment deflators for the years 1958 to 2012. I define the growth in the capital stock as:

$$\Delta K_{i,j,t+1} = \frac{I_{i,j,t} - \delta K_{i,j,t}}{P^I_{j,2010}}$$

Where the deflator is industry specific and converts net investment to constant year 2010 USD. Using Compustat I choose to measure net investment as the change in Property Plant and Equipment, net of depreciation (PPENT in Compustat).

Specifically:

$$\Delta K_{i,j,t+1} = \frac{\text{PPENT}_{i,j,t+1} - \text{PPENT}_{i,j,t}}{P^I_{j,2010}}$$

Before doing so I clean the data as follows:

1. I set capital equal to deflated Gross Property, Plant and Equipment for age 0 firms:[27]

$$K_{i,j,t} = \frac{\text{PPEGT}_{i,j,t}}{P^I_{j,2010}}$$

[27] Or in cases where there is no net investment or gross capital until later on I initialize the first observation of PPEGT.

2. If a firm has a missing observation for PPENT in the midst of an investment spell I interpolate its value using neighbouring observations.

A.2.3 Wage Bill

The NBER-CES database has a measure of the total employment and total wages in a given manufacturing industry. I use this to construct an industry level average wage as follows:

$$\bar{w}_{j,t} = \frac{\text{Payroll}_{j,t}}{\text{Employment}_{j,t}}$$

This of course is a nominal value of the average wage paid to production employees in industry j. I use the GDP Deflator as well as the Compustat variable EMP to construct firm level payrolls:

$$\text{Payroll}_{i,j,t} = \frac{\bar{w}_{j,t}\text{EMP}_{i,j,t}}{P^{Y}_{2010}}$$

A.2.4 Materials

I construct materials using Compustat information on Cost of Goods Sold (COGS) and Selling, General and Administrative Expense (XSGA). I deflate using the NBER-CEA materials deflator for industry j and remove average payroll to correct for labour costs.

$$M_{i,j,t} = \frac{\text{COGS}_{i,j,t} + \text{XSGA}_{i,j,t}}{P^{M}_{j,2010}} - \frac{\text{R}_{i,j,t}}{P^{I}_{j,2010}}\text{Payroll}_{i,j,t}$$

Notice that I also subtract R&D expenditures. This is because these are contained in Selling, General and Administrative expenses. Since I explicitly model R&D and this paper it is not appropriate to leave it in materials.

My estimation strategy requires a measure of the materials share in output. This is defined as:

$$S^{M}_{i,j,t} = \frac{M_{i,j,t}}{\text{SALE}_{i,j,t}}$$

A.2.5 Other Variables

At this point I use the Manufacturing Gross Output Deflator from the BEA to deflate the Compustat variable SALE for firms in all industries.[28] Labour input is the number of employees as measured by EMP in Compustat.

Notice that I deflate R&D expenditures with the investment deflator for the appropriate industry.

I have now created an unbalanced panel of firms with measurements for:

$$\mathcal{Y}_{J,I,T} = \{Y, K, N, M, S^M, R, A, X\}$$

Where the variables are gross output, capital, employment, materials, the share of materials cost in gross output, R&D expenditures, age since listing and an exit dummy which equals 1 if a firm is in its last year in the sample. These variables, in addition to industry and time dummies are all that is required to implement my estimation strategy. I have tried my best to align the measurement with the economic definition of the variable.

[28] I will need to figure out the concordance between BEA Industries and my SIC or NAICS industries in order to deflate with industry specific deflators.

A.3 Sample Selection

Below I discuss my sample selection choices in detail. Table A.3.1 shows how many firm-year observations are dropped at each stage. The final sample has $27,543$ firm-year observations on $1,813$ firms from 1976 to 2011.

1. Start from a sample of manufacturing firms with 2 Digit NAICS Codes in range $31-33$.
2. Drop firms without ages which are present in database in 1950.
3. Drop observations which have merger activity greater than 20% of sales.
4. Keep only observations from 1976 to present.
5. Drop age 0 firms as they do not have a measure of capital stock.
6. Keep only American firms.
7. Drop firms which have 2 or more mergers in their lifetime.
8. Drop rows with missing observations in important variables. Often this will be abnormal values for materials (less than 0).
9. Trim outliers from the data. I trim observations which are above the 99th percentile for sales, capital, materials, employment, material share or R&D intensity. There are some abnormally large values which may influence results.
10. Run the first stage regression on time and industry dummies to recover $\hat{y}_{i,j,t}$ and its lags. Drop variables with missing observations for $t-1$.
11. Keep only observations with $R > 0$. I will not always do this, but for some of my results I use the "innovative" sample which requires that R&D spending be positive.

Table A.3.1: Sample Selection

Stage	Count	Dropped
(1)	161,070	(-)
(2)	139,363	21,707
(3)	61,134	78,229
(4)	53,864	7,280
(5)	50,855	2,999
(6)	48,217	2,638
(7)	43,654	4,563
(8)	33,132	10,522
(9)	30,820	2,312
(10)	27,543	3,277
(11)	17,637	9,906

A.4 Additional Tables

I estimate the full model on subsets of the data. Table A.4.1 , Table A.4.2 and Table A.4.3 replicate Table 2 , Table 3 and Table 4 in Section 4 of the main text. I define industries corresponding to the following NAICS codes: Machinery (NAICS-333, Machinery Manufacturing), Computers (NAICS-334, Computer and Electronic Product Manufacturing), Misc. (NAICS-339, Miscellaneous Manufacturing), Wood (NAICS-321, Wood Product Manufacturing, NAICS-322, Paper Manufacturing, NAICS-323, Printing and Related Support Activities), Electrical (NAICS-335, Electrical Equipment, Appliance and Component Manufacturing). I do not run the model on industries with less than 1,000 observations or Chemical Manufacturing. Furthermore, I do not run the model on Chemical Manufacturing (NAICS-325) as I encountered identification issues.

Next I estimate the model on two different time periods. The earlier time period of 1986-1995 and the later period of 1996-2010. Recall that the full sample goes from 1977-2010. I chose the decades I did based on them having the most observations. These results are presented in Table A.4.4, Table A.4.5, and Table A.4.6.

Lastly, I present some tables which evaluate the robustness of my estimator in Table A.4.7, Table A.4.8, and Table A.4.9. In the second column (No Exit) I show estimates which omit the controls for exit. In the third column (R&D Lags) I include an additional lag of R&D spending. These results are complementary to the ones presented in Table 5. The last column estimates the model under a different assumption about variable inputs. The variable Cost of Goods Sold is assumed to be a variable input which is complementary to capital.

Table A.4.1: Full Model Production Function Estimates

	Machinery	Computers	Misc.	Wood	Electrical
θ_k	0.5927	0.0792	0.4064	0.1673	0.3206
	(0.1193)	(0.0359)	(0.1152)	(0.0571)	(0.1374)
θ_l	0.5550	0.0358	0.3764	0.4758	0.5918
	(0.0028)	(0.3136)	(0.1346)	(0.0955)	(0.1853)
θ_m	0.8525	0.8693	0.5440	0.6134	0.5339
	(0.2386)	(0.0075)	(0.0036)	(0.0036)	(0.0060)
NT	3,987	4,840	2,761	1,696	1,619

*

Table A.4.2: Conditional Mean Function Estimates

	Machinery	Computers	Misc.	Wood	Electrical
μ_0	0.8525	0.0425	-0.1679	0.2580	0.6729
	(0.2386)	(0.2266)	(0.2684)	(0.1101)	(0.3438)
μ_z	0.3721	-0.0851	0.5707	0.6706	0.3793
	(0.0674)	(0.2178)	(0.0692)	(0.0710)	(0.1241)
μ_a	-0.2159	0.0071	-0.0618	-0.0080	-0.2381
	(0.0587)	(0.1028)	(0.0325)	(0.0295)	(0.0931)
$\mu_{z,a}$	0.0746	0.0840	0.0045	-0.0447	0.1174
	(0.0295)	(0.0533)	(0.0229)	(0.0311)	(0.0460)
μ_0^r	-0.1924	-0.0119	0.2001	0.3763	-0.1141
	(0.1553)	(0.3068)	(0.1024)	(0.1906)	(0.2798)
μ_z^r	0.4755	-0.0513	0.3953	0.3741	0.4367
	(0.0460)	(0.1399)	(0.0561)	(0.1299)	(0.1050)
μ_a^r	-0.1317	0.0069	-0.1292	-0.1045	-0.1252
	(0.0331)	(0.1344)	(0.0342)	(0.0629)	(0.0561)
μ_r^r	-0.1599	0.0176	-0.0995	-0.1856	-0.2564
	(0.0300)	(0.1275)	(0.0224)	(0.0709)	(0.0515)
$\mu_{z,a}^r$	0.0418	0.0327	0.0638	0.0013	0.0670
	(0.0164)	(0.0463)	(0.0215)	(0.0529)	(0.0345)
$\mu_{z,r}^r$	0.0343	0.0059	-0.0194	0.0240	0.0476
	(0.0118)	(0.0071)	(0.0141)	(0.0252)	(0.0124)
$\mu_{a,r}^r$	0.0190	0.0017	0.0270	0.0436	0.0416
	(0.0062)	(0.0224)	(0.0079)	(0.0200)	(0.0145)
μ_χ	6.9151	2.5960	10.8375	4.8837	6.4023
	(1.5777)	(0.9948)	(2.5278)	(1.3280)	(1.6868)
NT	3,987	4,840	2,761	1,696	1,619

Table A.4.3: Skedastic Function Estimates

	Machinery	Computers	Misc.	Wood	Electrical
σ_0^2	0.9829	0.2740	1.4417	0.4402	0.8572
	(0.1668)	(0.2852)	(0.3144)	(0.0971)	(0.5280)
σ_z^2	-0.1716	-0.0562	-0.3567	0.0421	-0.0454
	(0.0736)	(0.2366)	(0.1936)	(0.0865)	(0.2956)
σ_a^2	-0.1531	0.0410	-0.1765	0.0402	-0.0850
	(0.0582)	(0.0879)	(0.0888)	(0.0399)	(0.1758)
$\sigma_{z,a}^2$	0.0582	0.0541	0.1047	-0.0657	-0.0063
	(0.0350)	(0.0692)	(0.0584)	(0.0368)	(0.0954)
s_0^2	0.2531	0.1316	-0.3911	-0.1895	0.2145
	(0.2106)	(0.1302)	(0.2579)	(0.1910)	(0.6957)
s_z^2	-0.3428	0.2026	-0.1130	0.2043	-0.3015
	(0.0818)	(0.2524)	(0.0769)	(0.1549)	(0.2948)
s_a^2	-0.1067	-0.0280	-0.0447	0.1908	-0.0482
	(0.0508)	(0.0537)	(0.0297)	(0.0969)	(0.1491)
s_r^2	-0.2017	-0.0177	-0.1030	-0.0685	-0.1861
	(0.0320)	(0.0681)	(0.0257)	(0.0489)	(0.1198)
$s_{z,a}^2$	0.0406	-0.0564	0.0104	-0.1796	-0.0037
	(0.0283)	(0.0630)	(0.0259)	(0.0819)	(0.1020)
$s_{z,r}^2$	0.0437	-0.0024	0.0238	0.0599	0.0370
	(0.0111)	(0.0128)	(0.0126)	(0.0291)	(0.0147)
$s_{a,r}^2$	0.0276	-0.0001	0.0086	-0.0183	0.0137
	(0.0068)	(0.0142)	(0.0080)	(0.0151)	(0.0374)
σ_χ^2	-2.4277	-2.5163	-5.7696	-2.4827	-1.3459
	(1.6040)	(1.1203)	(2.8266)	(0.9303)	(1.7839)
NT	3,987	4,840	2,761	1,696	1,619

Table A.4.4: Production Function Estimates by Time Period

	Baseline	1986-1995	1996-2010
θ_k	0.2621	0.4220	0.3449
	(0.0405)	(0.1082)	(0.0183)
θ_l	0.4945	0.3618	0.0520
	(0.0636)	(0.2406)	(0.0384)
θ_m	0.6158	0.5936	0.7219
	(0.0013)	(0.0021)	(0.0016)
NT	27,543	9,104	11,853

Table A.4.5: Conditional Mean Function Estimates by Time Period

	Baseline	1986-1995	1996-2010
μ_0	-0.1663	-0.0243	0.0471
	(0.1009)	(0.1736)	(0.1560)
μ_z	0.5321	0.6670	0.4300
	(0.0306)	(0.0387)	(0.0653)
μ_a	-0.0862	-0.0809	-0.0663
	(0.0125)	(0.0153)	(0.0081)
$\mu_{z,a}$	0.0152	0.0286	0.0930
	(0.0108)	(0.0200)	(0.0256)
μ_0^r	0.3375	0.3417	0.1678
	(0.0461)	(0.1097)	(0.0294)
μ_z^r	0.3789	0.4909	0.3256
	(0.0285)	(0.0434)	(0.0461)
μ_a^r	-0.1628	-0.1616	-0.0915
	(0.0158)	(0.0331)	(0.0107)
μ_r^r	-0.1041	-0.1603	-0.0651
	(0.0102)	(0.0288)	(0.0084)
$\mu_{z,a}^r$	0.0691	0.0688	0.1188
	(0.0085)	(0.0219)	(0.0167)
$\mu_{z,r}^r$	0.0011	0.0256	0.0128
	(0.0057)	(0.0072)	(0.0110)
$\mu_{a,r}^r$	0.0202	0.0428	0.0155
	(0.0030)	(0.0059)	(0.0036)
μ_χ	11.6515	6.5865	1.6757
	(1.3289)	(5.0335)	(1.4818)
NT	27,543	9,104	11,853

Table A.4.6: Skedastic Function Estimates by Time Period

	Baseline	1986-1995	1996-2010
σ_0^2	0.8429	0.3821	0.1731
	(0.1398)	(0.1562)	(0.1103)
σ_z^2	-0.0995	-0.2391	-0.3257
	(0.0537)	(0.0805)	(0.0903)
σ_a^2	-0.0790	-0.0890	-0.0032
	(0.0195)	(0.0196)	(0.0079)
$\sigma_{z,a}^2$	0.0289	0.0212	0.0615
	(0.0191)	(0.0287)	(0.0344)
s_0^2	0.3276	0.2915	0.0295
	(0.0978)	(0.1583)	(0.0319)
s_z^2	-0.3203	-0.4241	-0.2107
	(0.0602)	(0.0875)	(0.0553)
s_a^2	-0.1292	-0.1440	-0.0038
	(0.0252)	(0.0372)	(0.0078)
s_r^2	-0.1786	-0.1595	-0.0611
	(0.0165)	(0.0312)	(0.0094)
$s_{z,a}^2$	0.0617	0.0408	0.0279
	(0.0178)	(0.0318)	(0.0220)
$s_{z,r}^2$	0.0236	0.0025	0.0112
	(0.0071)	(0.0115)	(0.0098)
$s_{a,r}^2$	0.0329	0.0318	0.0151
	(0.0050)	(0.0089)	(0.0033)
s_χ^2	-3.9090	4.3127	0.6710
	(2.1113)	(3.0637)	(0.9405)
NT	27,543	9,104	11,853

Table A.4.7: Robust Production Function Estimates

	Baseline	No Exit	R&D Lags	COGS
θ_k	0.2621	0.4482	0.2652	0.1350
	(0.0405)	(0.0090)	(0.0421)	(0.0117)
θ_l	0.4945	0.1013	0.4932	-
	(0.0636)	(0.0103)	(0.0664)	-
θ_m	0.6158	0.6158	0.6158	0.6151
	(0.0013)	(0.0013)	(0.0013)	(0.0013)

Table A.4.8: Robust Conditional Mean Function Estimates

	Baseline	No Exit	R&D Lags	COGS
μ_0	-0.1663	0.2379	-0.1772	0.6039
	(0.1009)	(0.0186)	(0.1027)	(0.0605)
μ_z	0.5321	0.5954	0.5327	0.5615
	(0.0306)	(0.0339)	(0.0304)	(0.0354)
μ_a	-0.0862	-0.0741	-0.0854	-0.0909
	(0.0125)	(0.0054)	(0.0124)	(0.0164)
$\mu_{z,a}$	0.0152	0.0667	0.0155	0.0790
	(0.0108)	(0.0139)	(0.0106)	(0.0136)
μ_0^r	0.3375	0.1347	0.3069	-0.0376
	(0.0461)	(0.0206)	(0.0470)	(0.0441)
μ_z^r	0.3789	0.4379	0.3793	0.5017
	(0.0285)	(0.0225)	(0.0285)	(0.0252)
μ_a^r	-0.1628	-0.0944	-0.1630	-0.0624
	(0.0158)	(0.0052)	(0.0156)	(0.0077)
μ_r^r	-0.1041	-0.0592	-0.0438	0.0699
	(0.0102)	(0.0051)	(0.0165)	(0.0044)
$\mu_{z,a}^r$	0.0691	0.1167	0.0704	0.0744
	(0.0085)	(0.0096)	(0.0086)	(0.0077)
$\mu_{z,r}^r$	0.0011	0.0252	-0.0057	0.0004
	(0.0057)	(0.0031)	(0.0078)	(0.0029)
$\mu_{a,r}^r$	0.0202	0.0139	0.0119	-0.0103
	(0.0030)	(0.0019)	(0.0049)	(0.0019)
μ_χ	11.6515	-	11.6426	-0.9046
	(1.3289)	-	(1.2952)	(0.5630)

Table A.4.9: Robust Skedastic Function Estimates

	Baseline	No Exit	R&D Lags	COGS
σ_0^2	0.8141	0.5039	0.8355	1.0444
	(0.1360)	(0.0216)	(0.1361)	(0.0846)
σ_z^2	-0.1020	-0.3323	-0.0987	-0.4252
	(0.0528)	(0.0587)	(0.0536)	(0.0555)
σ_a^2	-0.0788	-0.0662	-0.0769	-0.1092
	(0.0191)	(0.0069)	(0.0185)	(0.0295)
$\sigma_{z,a}^2$	0.0293	0.0426	0.0282	0.0903
	(0.0186)	(0.0242)	(0.0186)	(0.0238)
s_0^2	0.1442	0.1084	0.3071	0.1392
	(0.0866)	(0.0367)	(0.0955)	(0.0873)
s_z^2	-0.2470	-0.4095	-0.3215	-0.4967
	(0.0538)	(0.0682)	(0.0596)	(0.0569)
s_a^2	-0.0758	-0.0721	-0.1280	0.0214
	(0.0209)	(0.0105)	(0.0252)	(0.0250)
s_r^2	-0.1606	-0.1296	-0.1122	0.0147
	(0.0136)	(0.0088)	(0.0428)	(0.0138)
$s_{z,a}^2$	0.0392	0.0416	0.0618	-0.0357
	(0.0154)	(0.0260)	(0.0179)	(0.0232)
$s_{z,r}^2$	0.0201	0.0268	-0.0062	-0.0268
	(0.0065)	(0.0065)	(0.0166)	(0.0113)
$s_{a,r}^2$	0.0289	0.0288	0.0286	0.0478
	(0.0040)	(0.0033)	(0.0103)	(0.0053)
s_χ^2	-3.4577	-	-3.9125	-3.7067
	(2.0444)	-	(2.0808)	(0.6625)