R&D heterogeneity and implications for growth*

Sigurd Mølster Galaasen[†] and Alfonso Irarrazabal[‡]

May 1, 2016

Abstract

This paper quantifies the determinants of R&D investment heterogeneity and its implications for growth. We estimate a Schumpeterian growth model with heterogeneous firms, à la Lentz and Mortensen (2008), in which firms differ with respect to innovation efficiency. Using observations on size, productivity, and R&D expenditures from a panel of Norwegian manufacturing firms we find that the model has a good fit to the data. In particular, it fits the distribution of R&D investment (mean, dispersion and skewness) as well as the negative correlation between research intensity and size. Moreover, the model generates firm-level investment responses to R&D subsidies that are in line with micro evidence from a natural experiment. The model estimates imply that a large part of aggregate productivity growth (72 percent) is the result of the market directing R&D resources to the more innovative firms. Finally, we study the link between firm heterogeneity and R&D subsidies, and show that the growth effects of subsidies depend crucially on how the policy influences the equilibrium distribution of firms.

JEL Classification: L11, O3,O4 *Keywords*: R&D, Heterogeneous Firms, Subsidies, Growth.

*We would like to thank Ufuk Akcigit, Nils Christian Framstad, Steinar Holden, Per Krusell, Rasmus Lentz, Espen R. Moen and Kjetil Storesletten for useful comments and suggestions. We thank Statistics Norway for providing the data. The analysis, opinions, and findings represent the views of the authors, they are not necessarily those of Norges Bank.

[†]Norges Bank and Statistics Norway, sigurd.galaasen@gmail.com

[‡]Norwegian Business School (BI), alfonso.irarrazabal@bi.no

1 Introduction

R&D surveys report substantial and persistent heterogeneity in R&D intensities.[1] In fact, most firms report zero R&D investment, with some firms reporting some R&D and a few firms reporting large investment in R&D relative to size. What is the source of this heterogeneity? How important is it to allow innovative firms to grow at the expense of less innovative firms? And finally, what is the growth effect of R&D policy (e.g. innovation subsidies) in this environment?

In this paper we address these questions by estimating an equilibrium model of firm-level innovation and growth. We adopt the framework of Klette and Kortum (2004) extended by Lentz and Mortensen (2008). In the model firms grow through product innovation in a Shumpeterian creative-destruction setting, allowing for innovation by both incumbent and entering firms. The two key forces that generates heterogeneity in R&D intensity (R&D expenditures relative to value added) at the firm level in the model are demand shocks and type heterogeneity. Firms invest in R&D which may lead to an innovation. Demand shocks are generated by letting consumer expenditure shares vary across products and undirected innovation, implying that the revenue associated with a new product is random for the firm. Type heterogeneity arises because incumbent firms differ with respect to the quality improvement associated with an innovation, i.e. some firms produce higher quality innovations than others. This heterogeneity is exogenous and realized upon entry. Absent demand shocks and type heterogeneity all firms would have the same R&D intensity[2]. Type heterogeneity generates variations in R&D intensity since high types (those producing high quality improvements) have higher expected return to R&D and thus invest more than low types, both in terms of levels and relative to size. Thus, high types have higher R&D intensity and grow faster than low types.

The model provides a rich, yet tractable, framework that links firm-level dynamics to micro level data. Using observations on size, productivity and R&D expenditures from a panel of Norwegian manufacturing firms, we estimate the model and quantify the relative importance of different sources of R&D heterogeneity. Thus, we contribute to the recent literature that estimates variants of the Klette-Kortum model using micro data on R&D.[3]

---

[1] R&D intensity (RI) is a measure of R&D expenditures relative to size. Firm size is measured in terms of value added.

[2] The R&D production function has the property that a firm's optimal R&D investment is proportional to its size (measured as number of products). Absent demand shocks, value added is proportional to the firm's product size.

[3] Recently, several papers have used R&D information to estimate structural models similar to ours,

We find that the model has a good fit to both R&D and non-R&D moments. In particular, the estimated model fits the observed distribution of R&D expenditures and intensity (mean, dispersion and skewness) as well as the negative correlation between R&D intensity and size. This correlation is driven by small firms in our dataset, which are obtained from a survey of all firms above 50 wokers and a sample of firms between 10 and 50 workers. In contrast, previous studies have used mainly large firms (for survey limitations) and found a zero correlation in their sample (Cohen and Klepper, 1996, Klette and Grilliches, 2000 and others). In our framework, a negative correlation is the result of firms becoming large not only due to innovation activity, but also due to persistent random shocks to the demand for their products. In the estimated model the group of small firms tend to be dominated by those experiencing negative demand shocks. We also find that this shock is key for generating substantial cross-sectional dispersion in R&D intensity, while firm type heterogeneity is important for cross-sectional variation in R&D expenditures.

Firm type differences imply that reallocation of workers from less to more productive firm types generates aggregate efficiency gains. Using data from Danish firms, Lentz and Mortensen (2008) show that around 53 percent of aggregate growth is explained by this reallocation. We use Norwegian manufacturing firms and compute the importance of reallocation for growth in our sample. Crucially, we have R&D information, which can be use to discipline the model along the R&D dimension. To make our estimation comparable to Lentz and Mortensen (2008), we first exclude observations on R&D, and find that the reallocation effect accounts for 44.5 percent of aggregate productivity growth. This magnitude is similar to what they find, which is 49 percent for the manufacturing sector. However, we miss some key empirical R&D patterns: Research intensity is too negatively correlated with size and firms doing R&D are too many, too small and invest too little in R&D relative to data. When re-estimating the model adding R&D moments the new parameters imply a larger role for reallocation, which now explains 72 percent of aggregate productivity growth.

We perform several tests of the model. First, the model produces firm-level response to R&D subsidies that are in line with micro evidence from a natural experiment (Bøler et al. (2014)). In the short run, firms increase their R&D spending with roughly 40 percent in response to a 20 percent R&D subsidy. Using a 2002 Norwegian policy reform, exploiting the fact that only firms with less than 4 million NOK in R&D spending, Bøler et al. (2014) estimates a reform-induced increase in R&D

see for example Akcigit and Kerr (2010) and Acemoglu et al. (2013)

spending by 35 to 72 percent during 2003-2005. Moreover, the model explains several cross sectional and dynamics moments for R&D, size and productivity when we restric the sample to large firms. Finally, the model reproduces features of the life cycle of firms over longer horizon than we consider in the estimation.

Finally, we use the estimated model to explore quantitatively the growth effects of R&D subsidies. Since we find a strong reallocation channel in our estimation, we expect substantial variation in growth effects, depending on how policy is implemented. By studying stylized reforms, we show that failure to target the best innovators may lead to subsidies creating small, or indeed adverse, productivity growth effects. In general, a subsidy's growth effect depends crucially on how it influences the equilibrium distribution of firms and R&D spending. For example, a subsidy that targets small firms (in terms of R&D expenditures) results in a 0.7 percentage point reduction in aggregate productivity growth rate relative to the decentralized equilibrium of 1.47 percent. Compared to subsidizing only incumbent firms, a subsidy to all firms (potential entrants and incumbents) reduces the growth rate from 1.83 to 1.53 percent. The reason for these adverse effects is that subsidies to small firms weaken the reallocation channel, and a larger share of less innovative firms is thus sustained in equilibrium.[4]

Our paper is related to several different literatures. First, it relates to the literature on R&D heterogeneity. Several papers have attempted to account for within industry differences in firm R&D intensity. Cohen and Klepper (1992) proposed a simply mechanism to explain R&D intensity dispersion observed in the data. The authors developed a probabilistic model where firms controls partially the outcome of their R&D effort. They also propose a mechanism that relates R&D spending to size.

More recently, papers have used a structural approach to understand the link between firm dynamics and R&D heterogeneity, see for example Akcigit and Kerr (2010) and Acemoglu et al. (2013). Akcigit and Kerr (2010) develop a model in which firms undertake heterogeneous research activities; exploration (capture new products) and exploitation (improve exciting product lines). Aw et al. (2011) estimate a structural model of producer decision to invest in R&D and export. Their model is partial equilibrium and limits the analysis to within plant productivity gains. Second, it relates to the literature of reallocation (Petrin and Levinsohn, 2012, Foster et al., 2001, Bartelsman et al., 2013 and others). Finally, our paper contributes to the

[4] Acemoglu et al. (2013) find that an optimal R&D policy involves subsidizing both entry and high incumbent type. The key mechanism that drives the difference in policy implications is that, in contrast to the Lentz-Mortensen model, type heterogeneity in Acemoglu et al. (2013) is transitory.

literature on R&D policy (Aghion et al., 2013, Acemoglu et al., 2013, Atkeson and Burstein, 2014, Lentz and Mortensen, 2014).

The paper proceeds as follows. Section 2 describes the data. Section 3 goes through the model. In particular, section 3.2 focuses on the link between Lentz and Mortensen (2008) and Klette and Kortum (2004), and section 3.4 explores the model implication for R&D patterns. In section 4 we go through the empirical implementation and estimation results, and section 5 contains the policy experiments. Section 6 concludes.

2 R&D facts

In this section we describe the data and show some stylized facts about R&D heterogeneity for the Norwegian manufacturing sector. We characterize the relationship between R&D intensity, size and productivity.

2.1 Data

The data consists of a panel of Norwegian manufacturing firms for the period 1997 to 2001. The data is gathered from two sources. First, we use balance sheet data from Statistics Norway's Capital database,[5] which is an annual unbalanced panel of all non-oil manufacturing joint-stock firms. The panel provides information about value added, wage bill and number of workers. Second, we use the Statistics Norway's biennial R&D survey[6], which provides information about firm-level R&D investment. The survey records R&D information for all firms with more than 50 workers. It also contains information for all firms with less than 50 employees with reported intramural R&D activity in the previous survey of more than NOK 1 million or extramural R&D of more than NOK 3 million. Finally, for the remaining firms, with 10-49 employees, a random sample was selected with a sampling rate of roughly 35 percent. We follow Lentz and Mortensen (2008) and exclude entry firms from the sample. Consequently, we follow the 1997 cross section of firms over the period 1997-2001. Before we trim the data we compute the aggregate wage rate in 1997 as the ratio of the aggregate wage bill to aggregate employment, $w_t = \sum_j W_{j,t} / \sum_j N^*_{j,t}$, where $W_{j,t}$ and $N^*_{j,t}$ are total wage bill and employment (number of workers) for firm j in year t. For subsequent years, we compute the wage rate with firms that

[5]For Capital database data documentation, see Raknerud et al. (2004)

[6]See Statistics Norway (2004)

were incumbents in 1997. We also follow Lentz and Mortensen (2008) and construct the quality adjusted employment ($N_{j,t}$) for firm j using $N_{j,t} = W_{j,t}/w_t$, which we use as our measure of firms' employment when constructing empirical moments.

We trim both tails of the employment distribution. At the bottom we drop firms with less than three workers. Many of these firms are single employee companies. At the top we exclude all firms above the top 1 percent of the size distribution. We also exclude all firms with R&D intensity (R&D expenditures over value added) above one in at least one year. Table 9 (appendix A) shows some descriptive statistics for our sample. We have 5290 firms with around 7 percent of those firms reporting positive R&D. The mean R&D intensity of these firms is 8 percent. We also report summary statistics for the 10-50 and above 50 workers categories.

2.2 Stylized facts

Now we present some stylized facts about R&D, size and productivity.

Distributions. Figure 1 panel (a) shows the R&D intensity distribution for all firms with positive R&D in 1997. The R&D distribution is positively skewed with a long right tail. This means that most of the R&D intensity is concentrated at low intensities but there are a few firms with large proportion of R&D expenditure relative to its size. The average R&D intensity for all firms sampled is around 8 percent and around 6 percent for firms with more than 200 workers[7].

Figure 1 panel (b) depicts the employment distribution of performers (positive R&D) and non-performers (reported zero R&D) for 1997, for firms with more than 50 workers.[8] The size distribution for performers has more mass to the right. For this sample performers are on average 1.22 times larger than non-performers in terms of employment.

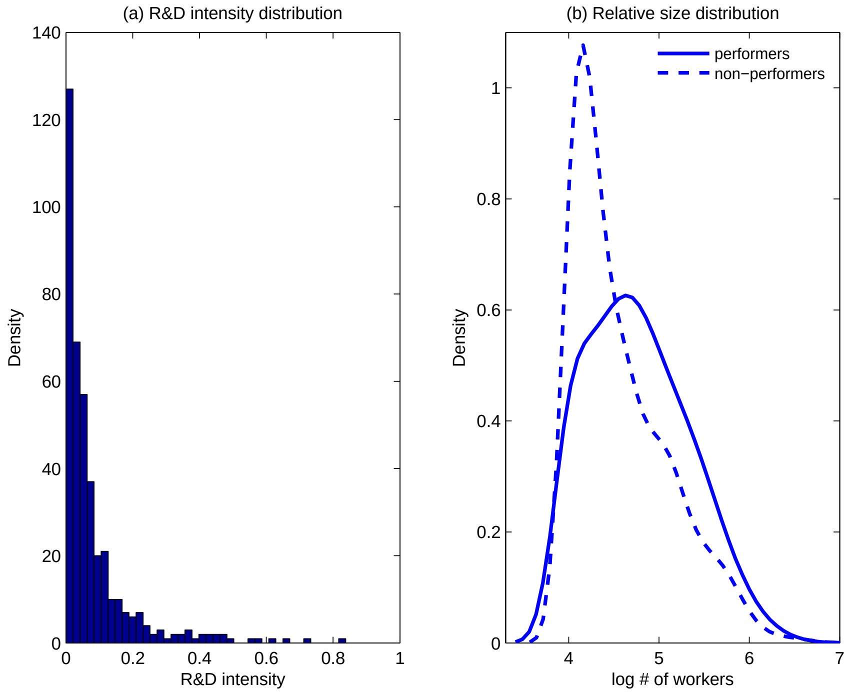
Correlations.

We also document negative correlations between R&D intensity with size and

[7]Doraszelski and Jaumandreu (2013) report values between 1 to 2.7 percent for Spanish manufacturing firms for a sample of firms with more than 200 workers. Acemoglu et al. (2013), using the Survey of Industrial Research and Development, report values of 9.9 for small firms and 4.2 for large firms. In their sample 32 percent of the firms have more than 500 employees.

[8]It has been reported for other data sets that a considerable fraction of firms report zero innovation. For example, for manufacturing firms with more than 10 workers, Harrison et al. (2008) reports a fraction of non-innovators ranging from 0.47 to 0.6 for four European countries. In our sample, the fraction of firms with zero R&D for firms above 50 workers is 0.65. When we include firms above 3 workers, this fraction rises to 0.92, which is one of the moments we target in the estimation. Notice that we do not target the fraction of zero R&D for firms above 50 workers, but the model gives a fraction of 0.71.

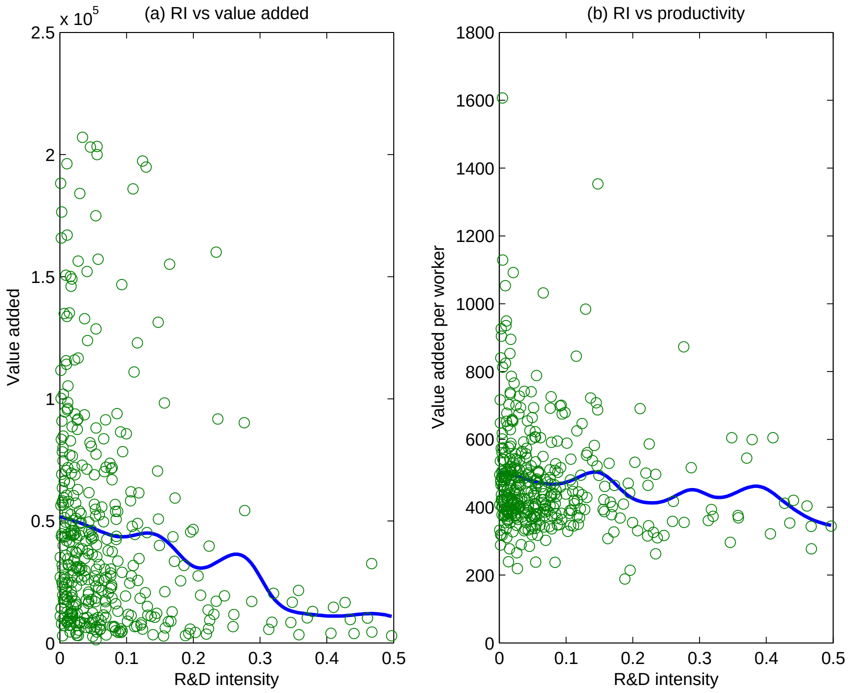
Figure 1: R&D intensity and relative size of performers distributions.



Notes: The data is from 1997. The R&D intensity histogram is computed including all sampled firms. The size distributions considers all firms above 50 workers and depicts kernel densities.

productivity. In Figure 2 panel a, we plot a kernel regression between R&D intensity and value added for 1997. The unconditional correlation is -0.18. It is interesting that most of the correlation is driven by firms with low R&D intensity. In fact, the correlation between R&D intensity and size is -0.02 for firms with 50 or more workers. In our framework, we will be able to explain this negative correlation because firms can become large not only due to innovation activity, but also due to random shock to the demand for their products. Since the latter part is unrelated to R&D it create a negative correlation between size and R&D intensity. In panel b we plot a kernel regression between R&D intensity and productivity measured as value added per worker. The unconditional correlation is -0.18

Figure 2: R&D intensity, value added and worker productivity.



Notes: Panel (a): Kernel regression and scatter plot of R&D intensity on value added (panel a) and value added per worker (panel b), in 1997.

3 The model

This section lays out the model. We will first go through the basics of the Klette-Kortum model, and then incorporate the innovations introduced by Lentz and Mortensen (2008).[9] The model is an endogenous growth model of expanding quality, extending the work of Grossman and Helpman (1991) and Aghion and Howitt (1992) by incorporating research by incumbent firms. This implies that firm size and R&D distributions are endogenous. There is a fixed measure of differentiated goods and innovation improves product quality. Firms compete in a quality ladder setting and invest in R&D to capture market shares.

[9]Readers already familiar with the these models may skip this section. Note that we assume unit price elasticity for all products. In contrast, Lentz and Mortensen (2008) allow for product price to affect product revenue. However, when they estimate the model they impose unit demand elasticities for all product varieties.

3.1 The Klette-Kortum model (2004)

Time is continuous. A representative household maximizes utility $U = \int_0^{\infty} e^{-\rho t} ln(C_t)dt$, and has Cobb-Douglas preferences over a unit continuum of differentiated goods,

$$ln(C_t) = \int_0^1 \ln[y_t(j)A_t(j)]dj, \tag{1}$$

where $y_t(j)$ and $A_t(j)$ measures the quantity and quality, respectively, of good j. Total labor supply (l) is exogenous and homogeneous, and can be used for two activities; production of goods and research. Total expenditures $E_t = P_t C_t$ is normalized to Z for all t. Given an interest rate r_t and the household's Euler equation $\dot{E}/E = r_t - \rho$, this normalization implies that $r_t = \rho$. Furthermore, it means that the consumption price P_t deflates at the rate of consumption growth. Since all goods carry equal log-preference weight, consumers spend Z on each good. The production technology is linear-in-labor and equal across all goods $y(j)$, with factor productivity normalized to 1. The unit (and marginal) cost of production is thus w, the cost per unit of labor.

The innovating firm. Firms are multi-product units. Firms enter with one product and to gain additional products they have to invest in R&D. The outcome of this research effort is stochastic. All firms innovate at the quality frontier and innovations occur randomly with a Poisson arrival rate I, chosen by the firm. Upon a successful innovation, the firm improves the quality of a random good j by a factor $q > 1$. This factor is specific to a firm and applies to all its innovations, but varies across firms. The time t quality of good j is given by

$$A_t(j) = \prod_{i=0}^{J_t(j)} q_i(j), \tag{2}$$

that is, the product of all past innovations $J_t(j)$, where $q_i(j)$ and $q_{J_t(j)}(j)$ are the quality improvement of the i^{th} and last innovation, respectively, and $q_0(j)$ is initial quality. Consider a firm making a successful time t innovation in good j. The innovation creates a blueprint which is a multiplicative improvement q over the blueprint $A_t(j)$ of the current producer. The firm's product quality is then given by $A_t(j)q$. Hence, the innovator combines past quality blueprints embedded in $A_t(j)$ (common knowledge) with its new blueprint to create a superior quality. It receives a patent for this blueprint lasting until a new innovation occurs. Since the innovator is the only one who can produce the frontier quality (all other firms can produce at quality $A_t(j)$) it becomes the sole supplier of that good by Bertrand pricing. The price $p(j)$ is a markup q over marginal cost w, i.e. the price that makes the buyer indifferent between the highest quality version and the second highest quality version, priced at marginal cost. The innovator receives a flow of profit associated with the product, given by

$$\begin{aligned}\Pi(j) &= p(j)y(j) - wy(j) = p(j)y(j)\left[1 - \frac{1}{q}\right] \\ &= Z\pi(q),\end{aligned} \tag{3}$$

where $\pi(q) = 1 - \frac{1}{q}$ is the profit share generated by the quality improvement q. The demand for production workers $l^w(j)$ (and quantity $y(j)$) associated with product j is then

$$l^w(j) = \frac{Z}{qw}. \tag{4}$$

Notice that the model features no social depreciation of knowledge. This is apparent from the definition of $A_t(j)$ in (2), where we see that quality stays constant in the case of no innovation. However, there is private depreciation of knowledge, in the sense that the firm's return to innovation only lasts until the product is overtaken by a competitor.

Innovation choice. The firm's state is the number of products k it currently produces. It invests in R&D to maximize present value of future profits. R&D investment generates new product at a frequency γk. Moreover, any good the firm produces is overtaken by another firm at Poisson rate δ, and firms with k products will see any one of them overtaken at rate δk. The destruction intensity δ is the outcome of aggregate innovation, and thus an equilibrium object.

Investment in R&D requires labor and knowledge capital, measured as the firm's number of products k. The total cost of R&D is $wc(\gamma)k$ where the function $c(\gamma)$ is assumed to be strickly increasing and convex.

The firm's optimal R&D investment solves the Bellman equation

$$rV(k) = \max_{\gamma \geqslant 0} \left\{ \pi(q)Zk - wkc(\gamma) + \gamma k\left[V(k+1) - V(k)\right] - \delta k\left[V(k) - V(k-1)\right] \right\}, \tag{5}$$

where the first two terms represent profit flow from the current portfolio of goods and R&D expenditures, and the last two terms represent the value of gaining and losing a product, respectively. Since the R&D technology features constant returns to scale in labor and number of products, the value and policy functions become proportional to the state variable

$$V(k) = vk$$
$$I(k) = \gamma k. \tag{6}$$

Demand for researchers is thus proportional to number of products

$$l^R(k) = kc(\gamma) \tag{7}$$

The innovation intensity γ and value per product v solve

$$wc'(\gamma) = v \tag{8}$$
$$v = \frac{Z\pi(q) - wc(\gamma)}{r + \delta - \gamma}. \tag{9}$$

Given the firm's innovation choice γ and the aggregate destruction rate δ, the firm's product size k follows a Poisson birth-death process. The waiting time until a firm with k products at time t gains or loses a product, is exponentially distributed with mean $1/(k(\gamma + \delta))$. When the transition occurs, the firm moves to state $k - 1$ with probability $\delta/(\gamma + \delta)$ and to state $k + 1$ with probability $\gamma/(\gamma + \delta)$. When the firm has lost all of its products it permanently exits the market, i.e. $k = 0$ is an absorbing state. As a consequence of the proportionality of the policy function, we can alternatively interpret a size-k firm as being a collection of k firms with 1 product.

3.2 Incorporating Lentz and Mortensen (2008)

Lentz and Mortensen (2008) estimate the Klette-Kortum model on Danish firm-level data. To account for firm heterogeneity, they extend the model along four dimensions.

A. Type heterogeneity In the setup in section 3.1, productivity measured as value added per worker, is independent of firm size k:

$$\begin{aligned} PR &= \frac{kZ}{kl^w + l^R(k)} \\ &= \frac{Z}{Z(wq)^{-1} + c(\gamma)}. \end{aligned} \tag{10}$$

Klette and Kortum (2004) create productivity dispersion across firms through firm specific innovation steps q (and thus profit shares $\pi(q)$). However, they modify the R&D cost function $c(\gamma)$ in such a way that the cost and benefit of large innovation steps are proportional, leaving the optimal creation rate γ constant across firm types. With homogeneous creation rates, firm size is unrelated to productivity.

Using Danish data, Lentz and Mortensen document a positive correlation between productivity and firm output size (value added), and zero correlation between productivity and firm input size (workers). To account for these relationships, they introduce heterogeneity in q as in Klette and Kortum (2004), but allow this to generate heterogeneous innovation intensities. In particular, profitable firms (high q) create larger quality improvements than less profitable firms, with the same R&D cost function $c(\gamma)$. From the firm problem it follows that $\pi(q_\tau) > \pi(q_{\tau'}) \Leftrightarrow \gamma_\tau > \gamma_{\tau'}$. Type τ firms have on average more products (and thus higher value added) than type τ' firms, and from (4) the demand for production workers associated with a product is negatively related to the size of the innovation step. Consequently, firm specific innovation steps can accommodate a positive correlation between labor productivity and value added, and zero correlation between labor productivity and employment.

B. Supply side shocks Value added per worker is perfectly persistent in the basic setup. To address this, Lentz and Mortensen (2008) relax the assumption that the firm-specific innovation step q is constant across innovations. When innovations occur, the type-specific quality jump q_τ is drawn from a Weibull distribution. The quality jumps of a more innovative type dominates (first order stochastic) that of less innovative types. The more innovative firm type is thus more profitable in expectation: $E\left[\pi(q_\tau)\right] > \left[\pi(q_{\tau'})\right] \Leftrightarrow \gamma_\tau > \gamma_{\tau'}$.

C. Demand side "shocks" In the basic Klette-Kortum model, firm growth is independent of firm size (Gibrat's law). In the Danish data, large firms tend to grow slower. To account for this, Lentz and Mortensen (2008) add random product market size by allowing the preference weight to vary across goods, with α_j as the weight on good j :

$$\ln(C_t) = \int_0^1 \alpha_j \ln[y_t(j) A_t(j)] dj,$$

Since $\int_0^1 \alpha_j dj = 1$, the expenditure share on good j is $z_j = \alpha_j Z$. Furthermore, R&D is undirected. Upon a successful innovation, the firm can not choose which good the quality improvement applies to; each good on the unit interval is an equally likely candidate. Thus, from the viewpoint of the innovator, product revenue is uncertain until the particular product variety is realized. This randomness create mean reversion in value added, which potentially can help explain the violation of Gibrat's law.

D. Capital cost Finally, Lentz and Mortensen (2008) add capital to the goods production function, using Leontief technology in labor and capital. Total factor productivity is normalized to 1, and marginal cost is given by $w + \kappa$ where w is cost per unit of labor and κ capital cost per unit of output. Capital cost does not impact firms' profitability, and is thus irrelevant for innovation choice. However, it directly impacts labor's share of value added, and thus pins down the level of value added per worker.

E. Firm problem Adding these features does not alter the firm problem substantially. We need to add the vectors of product demand realizations $z^k = (z_1, ..z_k)$ and innovation steps $q^k = (q_1, ..q_k)$ to the definition of current profit flow in the Bellman equation in equation (5) . However, looking forward, firms expect mean revenue Z and mean profit share $E[\pi(q_\tau)]$ on future innovations. The optimal R&D investment is still proportional to firm product size $I_\tau = \gamma_\tau k$ and the type-specific innovation intensity γ_τ solves

$$\begin{aligned} wc'(\gamma_\tau) &= v_\tau \\ v_\tau &= \frac{ZE[\pi(q_\tau)] - wc(\gamma_\tau)}{r + \delta - \gamma_\tau}, \end{aligned} \tag{11}$$

where v_τ now denotes the type-specific expected value of one product.

3.3 Entry and equilibrium

A. Entry rate There is a constant mass μ of potential entrants choosing an innovation intensity γ_0 to enter with one product. Aggregate innovation rate by entrants is then $\eta = \gamma_0 \mu$. Potential entrants face the same R&D cost function as an incumbent with one product, i.e. $wc(\gamma_0)$. On entry, firms learn their own type, drawn from the discrete distribution of potential firm types, where ϕ_τ denotes the fraction of type τ firms on entry. The expected value of entering with one product is thus given by

$E[v_\tau] = \sum_\tau \phi_\tau v_\tau$. The free entry condition requires that

$$wc'(\gamma_0) = wc'\left(\frac{\eta}{\mu}\right) = \sum_\tau \phi_\tau v_\tau. \tag{12}$$

B. Stationary equilibrium. In a stationary equilibrium, with creation rate γ_τ and destruction rate δ constant, the product birth-death process at the individual firm level give rise to a logarithmic distribution (with parameter $\frac{\gamma_\tau}{\delta}$) in k across firms of a particular type τ. Because firms of different types τ choose different creation rates γ_τ, the distribution differs across types.

Firms choose their innovation intensity γ_τ taking as given the aggregate destruction rate. In equilibrium, aggregate innovation must be consistent with the innovation by incumbents and entrants:

$$\delta = \eta + \sum_{\tau=1}^{n} K_\tau \gamma_\tau. \tag{13}$$

K_τ is the steady state mass of goods produced by type τ firms, given by

$$K_\tau = \frac{\eta \phi_\tau}{\delta - \gamma_\tau}. \tag{14}$$

We have that $\sum_{\tau=1}^{n} K_\tau = 1$, because there is a total mass 1 of goods. Through the process of creative-destruction, the equilibrium distribution of firms, denoted ϕ_τ^*, differs from the entry distribution ϕ_τ. The total steady state mass of type τ firms is given by

$$M_\tau = \frac{\eta \phi_\tau}{\gamma_\tau} \ln\left(\frac{\delta}{\delta - \gamma_\tau}\right), \tag{15}$$

and the equilibrium fraction is then

$$\phi_\tau^* = \frac{M_\tau}{\sum_{\tau=1}^{n} M_\tau}. \tag{16}$$

The stationary equilibrium consists of a constant wage w, aggregate destruction rate δ, entry rate η, type-specific creation rates $\{\gamma_\tau\}_{\tau=1}^{n}$ and distribution of products across types $\{K_\tau\}_{\tau=1}^{n}$, such that η satisfies the free entry condition in (12), creation rates γ_τ solves the firm's optimization problem, aggregate destruction δ and product distribution K_τ satisfies (13) and (14), and the wage rate clears the labor market [10].

[10] This is corresponds to the equilibrium definition in Lentz and Mortensen (2008) p. 1332. We refer to Lentz and Mortensen (2008, appendix C) for the equilibrium solution algorithm.

C. Aggregate growth rate As Klette and Kortum (2005) and Acemoglu et.al. (2013) we assume that at time 0 the economy is in steady state, and normalize initial quality index such that $A_0(j) = q_0(j)\ \forall\ j$. With this normalization, we implicitly normalize the previous quality version of each good $q_{-1}(j) = 1$ and assume that all goods are available in an improved $q_0(j)$ quality version at time 0. Consumption evolves according to

$$\begin{aligned}
\ln(C_t) &= \textstyle\int_0^1 \alpha_j \ln A_t(j)dj + \int_0^1 \alpha_j \ln y_t(j)dj \\
&= \textstyle\int_0^1 \alpha_j \left[\sum_{i=0}^{J_t(j)} \ln q_i(j) \right] dj + \int_0^1 \alpha_j \ln \left[\frac{\alpha_j Z}{w+\kappa} \frac{1}{q_{J_t}(j)} \right] dj \\
&= \textstyle\int_0^1 \alpha_j \left[\sum_{i=0}^{J_t(j)} \ln q_i(j) \right] dj - \int_0^1 \alpha_j \ln \left[q_{J_t}(j) \right] dj + \int_0^1 \alpha_j \ln \left[\frac{\alpha_j Z}{w+\kappa} \right] dj.
\end{aligned}$$

Along the stationary path new innovations arrive at the constant rate δ. We can then apply the law of large numbers to a weighted average (with weights $\alpha(j)$) to get[11]

$$\begin{aligned}
\ln(C_t) &= (\delta t + 1)E[\ln(q)] - E[\ln(q)] + \textstyle\int_0^1 \alpha_j \ln \left[\frac{\alpha_j Z}{w+\kappa} \right] dj \\
&= \delta t E[\ln(q)] + \textstyle\int_0^1 \alpha_j \ln \left[\frac{\alpha_j Z}{w+\kappa} \right] dj,
\end{aligned}$$

where δt is the expected number of innovations per product $J_t(j)$ over time length t, and the average log quality jump given by

$$E[\ln(q)] = \textstyle\sum_{\tau=1}^{n} \frac{\phi_\tau \eta + K_\tau \gamma_\tau}{\delta} E[\ln(q_\tau)],$$

where $\frac{\phi_\tau \eta + K_\tau \gamma_\tau}{\delta}$ is the fraction of new innovation arrivals attributed to firm type τ and $E\ln(q_\tau)$ is the type-conditional average log quality jump. Consequently, aggregate consumption grows at rate

$$g = \delta E[\ln(q)],$$

Growth thus arises due to the arrival of new innovations at rate δ with average quality contribution of $E[ln(q)]$.

[11] Note that since the firm cannot direct innovations to a particular product, $J_t(j)$ and $q_i(j)$ are i.i.d. across the unit continuum of products, and consequently not correlated with the weights α_j. To use the law of large numbers to a weighted average, we apply the Lindeberg Central Limit Theorem.

3.4 Model implications for R&D moments

In this section, we explain the different channels through which the model produces heterogeneity in R&D. This will be useful when we interpret the estimation results in section 4.

The evolution of an individual firm's product size k, is completely determined by the innovation choice γ_τ and the destruction rate δ. In steady state, the type-conditional distribution of number of products is logarithmic, with parameter γ_τ/δ. Given this equilibrium parameter and drawing firm types from the discrete distribution ϕ^*_τ and demand and innovation step sizes from their corresponding distributions, we can produce observations of value added $(Y_{i,t})$, wage bill $(W_{i,t})$, R&D expenditures $(RD_{i,t})$, employment $(N_{i,t})$, labor productivity (PR_i), R&D intensity (RI_i) across firms i at time t as follows

$$\begin{aligned}
Y_{i,t} &= \sum_{j=1}^{k_{i,t}} z_{i,j} \\
W_{i,t} &= w\left(\frac{1}{w+\kappa}\sum_{j=1}^{k_{i,t}}\frac{z_{i,j}}{q_{i,j}} + k_{i,t}c(\gamma_{\tau_i})\right) \\
RD_{i,t} &= wk_{i,t}c(\gamma_{\tau_i}) \\
N_{i,t} &= W_{i,t}/w \\
PR_{i,t} &= Y_{i,t}/N_{i,t}
\end{aligned} \tag{17}$$

Value added $(Y_{i,t})$ created by a firm with $k_{i,t}$ products is the sum of product revenues, and the wage bill $(W_{i,t})$ is the wage per worker times total labor demand (the sum of production and R&D workers). Given a product demand $z_{i,j}$ and the firm's pricing rule, the demand for production workers associated with the product is $\frac{z_{i,j}}{p_{i,j}} = \frac{z_{i,j}}{(w+\kappa)q_{i,j}}$. The optimal R&D investment requires $k_{i,t}c(\gamma_\tau)$ workers. Labor productivity is measured as value added per worker, and research intensity is defined as R&D spending $(RD_{i,t})$ over value added $RI_{i,t} = RD_{i,t}/Y_{i,t}$.

A. Dispersion in R&D The model generates cross-sectional heterogeneity in research intensity mainly through two channels, i) type-heterogeneity in innovation choices γ_{τ_i} and ii) demand shocks. If we shut down demand shocks, $RI_{i,t}$ becomes

$$RI_{i,t} = \frac{wk_{i,t}c(\gamma_{\tau_i})}{\sum_{j=1}^{k_{i,t}} z_{i,j}} = \frac{wk_{i,t}c(\gamma_{\tau_i})}{k_{i,t}Z} = \frac{w}{Z}c(\gamma_{\tau_i}) \tag{18}$$

Due to the proportionality of R&D investment and the state variable k, heterogeneity induced by the product birth-death process will not explain the cross-sectional R&D intensity distribution. Without demand shocks, $k_{i,t}$ drops out of the expression. With demand shocks, the k-distribution does affect the dispersion of R&D intensity across firms. But since R&D spending still scales with size, it will be of second order importance.

For the heterogeneity in the level of R&D spending, $RD_{i,t} = wk_{i,t}c(\gamma_{\tau_i})$, demand shocks are irrelevant. The dispersion is entirely determined by the product birth-death process and heterogeneity in innovation choice γ_τ across firms.

B. Correlation between research intensity and size The correlation, which is negative in the data, is given by

$$corr(RI, VA) = cor\left(\sum_{j=1}^{k_{i,t}} z_{i,j}, \frac{wk_{i,t}c(\gamma_{\tau_i})}{\sum_{j=1}^{k_{i,t}} z_{i,j}}\right) \tag{19}$$

is driven by two opposing forces. First, consider the pure Klette-Kortum (2004) model in which $RI_{i,t} = wc(\gamma)/Z$ is independent of firm size $Y_{i,t} = k_{i,t}Z$. Type-heterogeneity introduced by Lentz and Mortensen (2008) produces a positive relationship. More profitable firm-types choose higher innovation intensity γ than less profitable firms, since they expect a higher profit from a successful innovation. On average, these firm types have more products and invest more in R&D relative to size. On the other hand, demand shocks work in the opposite direction. Firms with a series of lucky product demand draws tend to be large. Since demand shocks are unrelated to the firm's R&D choice, these firms tend to have low R&D intensity.

C. Correlation between research intensity and productivity Among firms with $RI > 0$, R&D intensity and labor productivity is negatively correlated in the data. Firms with high R&D relative to size thus tend to have low productivity.

$$corr(RI, PR) = cor\left(\frac{wk_{i,t}c(\gamma_{\tau_i})}{\sum_{j=1}^{k_{i,t}} z_{i,j}}, \frac{\sum_{j=1}^{k_{i,t}} z_{i,j}}{\frac{1}{w+\kappa}\sum_{j=1}^{k_{i,t}} \frac{z_{i,j}}{q_{i,j}} + k_{i,t}c(\gamma_{\tau_i})}\right) \tag{20}$$

The model generates this pattern through demand shocks. Consider a model without demand shocks. As noted in section 3.2A, type-heterogeneity can accommodate a positive correlation between productivity and value added. If large firms also tend to have high R&D intensity, this translates into a positive correlation between productivity and R&D intensity. Demand shocks will, as with $corr(RI, Y)$, work in the opposite direction. Firms with high R&D intensity tend to have had bad demand draws, and a bad demand draw reduces value added per worker. The reason is the presence of R&D workers in the denominator of (10). This implies that total employment moves less than proportional to value added in response to demand shocks. Finally, supply side shocks (stochastic q) create variation in productivity only, hence pushing the correlation toward zero.

D. Remarks on measurement error In the estimation we allow for log-normal measurement error in value added, wage bill and R&D expenditures. Since employment is computed by dividing the wage bill with w, measurement error in W spills over to N. Measurement error in Y contributes both to R&D intensity variability and a negative correlation between R&D intensity and size and productivity. Measurement error in R&D creates additional R&D intensity dispersion, but pushes correlations towards zero. Wage bill measurement error drives $corr(RI, PR)$ to zero, but does not affect the dispersion in R&D intensity.

4 Empirical implementation

We now estimate the model using a panel of Norwegian firms starting in 1997. We follow Lentz and Mortensen (2008) and use indirect inference methods to estimate the structural parameters on cross-sectional and dynamic moments in 1997 and 2001. We first describe the estimation procedure. Then, we show that the model estimated on Norwegian data (but without using the R&D moments) gives reallocation effects of the same order of magnitude to those of Lentz and Mortensen. However, the estimated model has implications for R&D which are quite different from the observed R&D. Finally, we re-estimate the model (using the R&D moments) and find that reallocation across firms is more important for aggregate growth than what Lentz and Mortensen's study indicates.

4.1 Model estimator

We parametrize the cost function as $c(\gamma) = c_0\gamma^{1+c_1}$. Product revenues are drawn from a Weibull distribution with mean Z, origin at o_z and shape parameter β_z. Quality improvements are drawn from a Weibull with shape parameter β_q (common across firm-types), origin at 1 and type-specific scale parameter ε_{τ}^{q}. We consider three firm types in the estimation, and assume that type 1 does not innovate (i.e. $\varepsilon_1^q = 0$).[12] Finally, we allow for log-additive measurement error ($\zeta_{i,t}^{Y}, \zeta_{i,t}^{W}, \zeta_{i,t}^{RD}$) in value added, wage bill and R&D expenditures, distributed log-normally, $\ln(\zeta_{i,t}^{x}) \sim N(-\frac{\sigma_x^2}{2}, \sigma_x^2)$, $x \in \{Y, W, RD\}$.

In total, the model has 17 fundamental parameters. Two R&D cost function parameters (c_1, c_2), capital cost in goods production (κ), interest rate r, three demand parameters (Z, β_z, o_z), three innovation jump parameters $(\beta_q, \varepsilon_2^q, \varepsilon_3^q)$, the probability of being of type 2 and 3 at entry (ϕ_2, ϕ_3) and three measurement error variances $(\sigma_y^2, \sigma_w^2, \sigma_{rd}^2)$. Given the exogenous labor supply l and mass of potential entrants μ, the wage rate w and entry rate η are both equilibrium objects. However, as Lentz and Mortensen (2008), we estimate w and η, and let l and μ adjust such that the labor market clears and free entry condition holds.

The wage rate $w = 296.5$ is estimated directly from data and the interest rate is set to $r = 0.05$. The remaining 15 parameters are estimated by indirect inference. Given the parameters $\Lambda = \left\{\eta, c_1, c_2, \kappa, Z, \beta_z, o_z, \beta_q, \varepsilon_2^q, \varepsilon_3^q, \phi_2, \phi_3, \sigma_y^2, \sigma_w^2, \sigma_{rd}^2\right\}$, we simulate a firm-year (i, t) panel of value added $(\tilde{Y}_{i,t})$, wage bill $(\tilde{W}_{i,t})$, employment $(\tilde{N}_{i,t})$, productivity $(\tilde{PR}_{i,t})$ and R&D expenditures $(\tilde{RD}_{i,t})$ as follows: Solve for the optimal firm-type R&D choice γ_τ and aggregate δ creation rates. Calculate the aggregate growth rate g and equilibrium distribution of firm types ϕ_τ^*. Then simulate a 5 year firm panel. First, draw the firm type from the distribution ϕ_τ^* and its initial state vector of products k, revenues z and quality jumps q. Using the creation and destruction rates γ_τ and δ, simulate the birth-death process of number of products. This provides us with a panel of firm-year observations using the expressions in (17) and adding measurement error

$$\begin{aligned} \ln(\tilde{Y}_{i,t}) &= \ln(Y_{i,t}) + \zeta_{i,t}^{Y} \\ \ln(\tilde{W}_{i,t}) &= \ln(W_{i,t}) + \zeta_{i,t}^{W} \\ \ln(\tilde{RD}_{i,t}) &= \ln(RD_{i,t}) + \zeta_{i,t}^{RD} \end{aligned} \tag{21}$$

The simulated panel consists of all 5290 incumbent firms in 1997, which we follow until 2001 assuming steady state. Due to firm exit, and the fact that we exclude firm entry (both in the model simulation and in the data), the 2001 cross-section does

[12] Lentz and Mortensen (2008) also assume three firm types, but estimate ε_1^q. Their estimation produces $\varepsilon_1^q = 0$. Moreover, this non-innovating incumbent firm type accounts for 86 percent of all entry firms and 77 percent of equilibrium firms (manufacturing industry estimation).

not reflect steady state. The cross sectional shift from 1997 to 2001 consequently reflects the selected group of surviving firms. In addition, we sample R&D firms as in the survey.

The product size of a firm follows a continuous time birth-death process. To facilitate simulation, we follow Lentz and Mortensen and discretize the time space. A year is divided into 26 sub-periods. In any given sub-period, a firm with k products faces a probability $1 - e^{-\frac{k\delta}{26}}$ of losing a product and a probability $1 - e^{-\frac{k\gamma}{26}}$ of gaining a product.

We compute moments on the simulated panel, repeat 1000 times and store the average simulated moments. In total we have 37 non-R&D moments (same moments as Lentz and Mortensen, 2008) and 21 R&D moments. Tables 10 and 11 (appendix A) list the full set of empirical moments. Along the R&D dimension, we estimate the model on distributions of R&D effort (intensity and level), correlations between R&D intensity and size and productivity, and the fraction of R&D performing firms (firms with positive R&D) and their size relative to non-performers (firms with zero R&D). Both in the data and in the model we treat missing R&D observations, ie; non-sampled firms, as zeros.

Let Ω and $\widehat{\Omega}(\Lambda)$ denote the vectors of empirical and simulated moments, respectively. The parameter estimates are the solution to the minimization problem

$$\min_{\Lambda} \left[\Omega - \widehat{\Omega}(\Lambda)\right]' A \left[\Omega - \widehat{\Omega}(\Lambda)\right] \tag{22}$$

where the weighting matrix A is the inverse of the diagonal covariance matrix of the empirical moments. The squared difference between simulated and empirical moments are consequently weighted by dividing by the variance of the empirical moment. These variances are obtained by bootstrapping the original firm sample 5000 times. Precisely estimated empirical moments are thus given more weight in the minimization. Standard errors for the parameters are estimated by bootstrap.[13]

4.2 Benchmark estimation

Now we replicate the Lentz and Mortensen (2008) estimation on Norwegian data. Specifically, we drop R&D moments from the vector of moments $(\Omega, \widehat{\Omega})$. This provides a useful benchmark to Lentz and Mortensen (2008). They run estimations on

[13] We draw 500 bootstrap data samples from the original data set. For each sample we estimate the model on the bootstrap sample. Both the data moments and simulated moments are recentered around the corresponding moments from the full estimation.

both the entire private sector and on particular industries. Since we use data on Norwegian manufacturing firms, the natural comparison is their corresponding estimation on Danish manufacturing firms (pp. 1366-1367). We report amounts in units of 1000 NOK.

Table 1: Benchmark estimation

c_0/Z	c_1	η	Z	β_z	o_z	β_q	σ^2_Y
103.5	4.931	0.069	8891	0.428	3392	0.426	0.0296
σ^2_W	κ	g	l	μ	δ	$\frac{g_2}{g}$	
0.000	142.9	0.0161	19.0	0.97	0.124	0.445	
	ϕ_τ	ϕ^*_τ	K_τ	γ_τ	ε_τ	π_τ	
τ_1	0.7500	0.615	0.413	0	0	0	
τ_2	0.2499	0.384	0.578	0.0944	0.135	0.154	
τ_3	0.0002	0.0005	0.004	0.1216	1.311	0.409	

Selected moments (1997)	model	data		model	data
$E(PR)$	471.6	477.8	$Corr(PR, N)$	0.000	-0.030
$std(PR)$	173.8	173.8	$Corr(PR, Y)$	0.117	0.124
$E(Y)$	13090	12872	$Corr(Y, \frac{\Delta Y}{Y})$	-0.026	-0.006
$Std(Y)$	23485	23183	$Corr(PR, \Delta PR)$	-0.363	-0.342

Notes: Benchmark estimation targets only non-R&D moments in the data. Minimum of objective function: 167.701.

A. Non-R&D moments Table (1) shows estimated parameters, equilibrium values and a selection of targeted moments from our benchmark estimation. Table 12 in the appendix reports the full set of targeted moments.[14]

Overall, we obtain a very good fit of the model to the data. The estimated model captures the mean, dispersion and median for productivity in 1997, as well as the

[14] The estimation produces a large fraction of non-innovating incumbent firms, consistent with the results in Lentz and Mortensen. In contrast, however, the two R&D-performing incumbent firm types are quite different in terms of innovation intensities (γ_2, γ_3) in our estimation, whereas they are almost identical in Lentz and Mortensen (cf. table VII p 1366). But since the most innovative firm type only produces 0.4% of all goods in our estimation, the implication for reallocation is the same, i.e. the important margin is the reallocation of resources from non-innovating firms to innovating firms.

cross sectional shift to 2001. In contrast, to the Danish manufacturing data, the correlation between value added size and growth is essentially zero (-0.073 in Denmark, -0.006 in Norway) and the model is able to capture this relationship. In addition, the model matches both the persistence and mean reversion in productivity. As in the Danish data, productivity is positively correlated with output size (Y) and uncorrelated with input size (N), a feature the model fits quite well.

B. Reallocation and growth Aggregate growth, estimated to 1.6 percent annually, arises because of the arrival of better quality products, produced with the same amount of labor. The contribution to growth varies across firm types according to their innovation step q_τ and their innovation rate γ_τ. Moreover, high γ types innovate faster and capture market shares at the expense of low γ types. The main goal in Lentz and Mortensen (2008) is to quantify the role of this reallocation from less to more innovative firms in the growth process. They do this by decomposing the contribution to annual growth into three parts:

$$g = \underbrace{\sum_\tau \gamma_\tau E[\ln(q_\tau)]\phi_\tau}_{g_1:\ \textit{within types}} + \underbrace{\sum_\tau \gamma_\tau E[\ln(q_\tau)](K_\tau - \phi_\tau)}_{g_2:\ \textit{between types}} + \underbrace{\eta \sum_\tau E[\ln(q_\tau)]\phi_\tau}_{g_3:\ \textit{entry/exit}} \tag{23}$$

The first term (g_1) measures the growth contribution of continuing firms under the counterfactual that firm types are not allowed to increase their market share K_τ relative to their entry share ϕ_τ. The third term (g_3) measures the net effect of entry and exit. The key measure of reallocation emphasized by Lentz and Mortensen (2008) is captured by the second term (g_2). Consider a cohort of entry firms. Each firm enters with 1 product. Consequently, the distribution ϕ_τ of firm types on entry equals the distribution of products across types on entry. Over time however, more innovative firm types grow faster than less innovative types, thereby gaining an increasing proportion of the cohort's market share. As a result of this selection, the steady state distribution of products K_τ across types differs from the entry distribution ϕ_τ. This selection measures growth induced by the reallocation of market shares across types. Suppose we start out in an equilibrium with $g^* = g_1^* + g_2^* + g_3^*$. Now, if we counterfactually give all incumbent firms the same innovation rate $\gamma = g_1^* / (\sum_\tau E[\ln(q_\tau)]\phi_\tau)$, then there is no selection (i.e. $K_\tau = \phi_\tau$) and growth would decrease by g_2^*. From table (1) we see that imitators (type 1 firms) accounts for 75 percent of entrants' products, but only 41 percent of products in equlibrium. Hence, imitators lose roughly half of their market share to innovators due to selection.

Lentz and Mortensen's (2008) estimation on Danish manufacturing firms[15] implies that this selection accounts for 49% of aggregate growth. The process of entry and exit accounts for 25%, while the within-type contribution is 26%. In other words, if more innovative firms were not allowed to grow at the expense of less innovative firms, aggregate growth would be reduced with 49%. In our benchmark estimation we get similar results. Selection accounts for 44.5%, entry/exit 23.5% and within-type 32%.

Table 2: Benchmark estimation: Non-targeted R&D moments 1997

Moments	Data	Model	Moments	Data	Model
$E(RI)$	0.084	0.038	$corr(RI, Y)$	-0.170	-0.509
$std(RI)$	0.116	0.024	$corr(RI, PR)$	-0.188	-0.150
$Med(RI)$	0.044	0.036	1 $\frac{\#Firms_{RI=0}}{\#Firms}$	0.924	0.808
$E(RD)$	2907	698	$^2corr(RI, RI_{+2})$	0.688	0.810
$std(RD)$	5619	1081	3 $\frac{E(Y)_{RI>0}}{E(Y)_{RI=0}}$	4.44	3.53
$Med(RD)$	1163	455	4 $\frac{E(N)_{RI>0}}{E(N)_{RI=0}}$	4.44	3.21

Notes: Benchmark estimation targets only non-R&D moments in the data. [1]Fraction of firms with a zero R&D observation. [2]Correlation R&D intensity between 1997 and 1999. [3]Average value added for firms with positive R&D observations, relative to firms with zero R&D. [4]Average employment for firms with positive R&D observations, relative to firms with zero R&D

C. R&D moments Table 2 shows the fit of the model along the R&D dimension. The model produces some degree of dispersion in R&D cost and R&D intensity, and gets the correlation signs correct. Size and productivity are negatively correlated with research intensity across firms, and R&D intensity displays considerable persistence over time, $corr(RI, RI_{+2}) = 0.81$.

The model generates a negative correlation between R&D intensity and size that is much larger than in the data. The positive contribution from firm type heterogeneity is not enough to compensate for the negative impact of demand shocks to the correlation. Moreover, the model needs a substantial amount of demand variation in order to fit the firm size distribution. Simulating the model without demand shocks shows that median and dispersion of value added increases (72 percent) and

[15]p 1367. When they estimate the model on all private sector firms, selection accounts for 53% of growth

decreases (42 percent), respectively. The degree of type heterogeneity, which is primarily tied down by the gap between $corr(Pr, Y)$ and $corr(Pr, N)$, is not sufficient to generate the size dispersion observed in the data.

The mean and dispersion of R&D effort (level and intensity) is too low compared to data. There is also too low aggregate R&D intensity. Among firms with positive R&D, aggregate R&D intensity in 1997 (total R&D expenditures to total value added in 1997) is 65 percent smaller than in the data. Overall, the model produces too many firms with positive R&D, which on average are too small (in value added terms) and invest too little in R&D compared to the data. This indicates that if we add R&D moments to the estimation, reallocation between types becomes more important. The majority of the estimated selection effect comes from moving resources from incumbent firms of type 1 (which do not innovate) to firm type 2 and 3. When fewer firms do R&D, and they do it more intensely, it increases the importance of reallocating resources from non-innovators to innovators.

Summarizing, as Lentz and Mortensen (2008) we find that the model is good at explaining the observed distributions and correlations of size, growth and productivity. However, the model fails to reproduce the (non-targeted) R&D data dimension.

4.3 Estimation with R&D moments

Now we turn to the estimation with R&D, in which we add the 21 R&D moments to the list of moments to match. Table (3) reports the estimated parameters.

Let us first consider how come key parameters are adjusted when adding R&D moments to the estimation. Recall from table (1) the mean R&D expenditure in the benchmark estimation was counterfactually low. Moreover, performers are on average too small (in terms of value added) relative to non-perfomers. In order to increase firm R&D expenditure (measured as spending on R&D), the estimation procedure needs to increase the cost of R&D. From table (3) we observe an increase in c_0/Z and c_1. Therefore, to avoid reducing the incentive to innovate, we would then expect the gains from R&D (expected quality jump) to increase. Indeed, comparing tables (1) and (3) we see that the scale parameters ϵ_τ go up[16]. With higher average quality improvements, the average productivity would increase, as seen in equation 17. Consequently, in order to avoid a counterfactual productivitylevel, the fraction

[16] The coefficient of variation for q has gone up from 0.81 to 1.89 when adding RnD moments and the mean has roughly doubled from 1.38 to 2.15

of firms doing R&D has to fall, relative to the estimation in which we do not target R&D moments.[17]

The growth rate is estimated to be 1.47 percent annually. We find that 87.8 percent of firms do not innovate (ϕ_1^*), and they produce 69.5 percent of all goods (K_1). The expected life-time for these firms is 9 years and they employ on average 22.8 workers. The main bulk of innovation comes from firm type 2. These firms account for 28.7 percent of goods in equilibrium and have an expected profit share of 23.3 percent. They are expected to live for 20 years and have on average 52.3 production workers and 4.4 researchers. Type 3 firms are very innovative, but few, producing only 1.7 percent of all products. They employ 265 production workers and 80 researchers on average and have a life expectancy of 41 years.

The model's fit along non-R&D moments is still good. However, there is some trade-off asociated with matching R&D moments. In particular, the correlation between R&D intensity and valued added is larger than in the benchmark estimation, because a smaller share of performers are now on average larger. Similarly, the estimation with R&D moments generates a large dispersion in value added and productivity.

A. R&D heterogeneity Table 4 compares the fit of the model with R&D moments from the data in 1997[18]. The model fits both the dispersion and skewness of the R&D intensity distribution. Figure 3 depicts the data distribution with those from our estimation with R&D moments and from a counterfactual experiment in which we shut down demand shocks. Even though in the estimation we only target the mean, median and standard deviation, the model captures the shape and location of the entire distribution. In particular, it is able to match the right tail of the distribution fairly well. Figure 3 displays the R&D intensity distribution when we shut down demand shocks. We observe that we miss fit on the right tail of the distribution. The model without demand shocks generates too few firms with high R&D intensity. Intuitively, firms that draw bad demand shocks will have high R&D intensity.

To evaluate further the relative importance of shocks, we shut down each of the shocks and recompute the moments. Table 5 depicts R&D moments under different scenarios. Consider shutting down measurement errors and demand shocks. This eliminates all within-type heterogeneity. Table 5 column (2) shows this case. The

[17] We estimated the model without using the fraction of non-performers as target and have obtained similar results. In particular, the model still generate a substantial fraction of non-performing firms in equilibrium ($\phi_1^* = 0.85$) .

[18] In appendix A, table 14 we present the fit to 2001 R&D moments

Table 3: Parameters: estimation with R&D moments

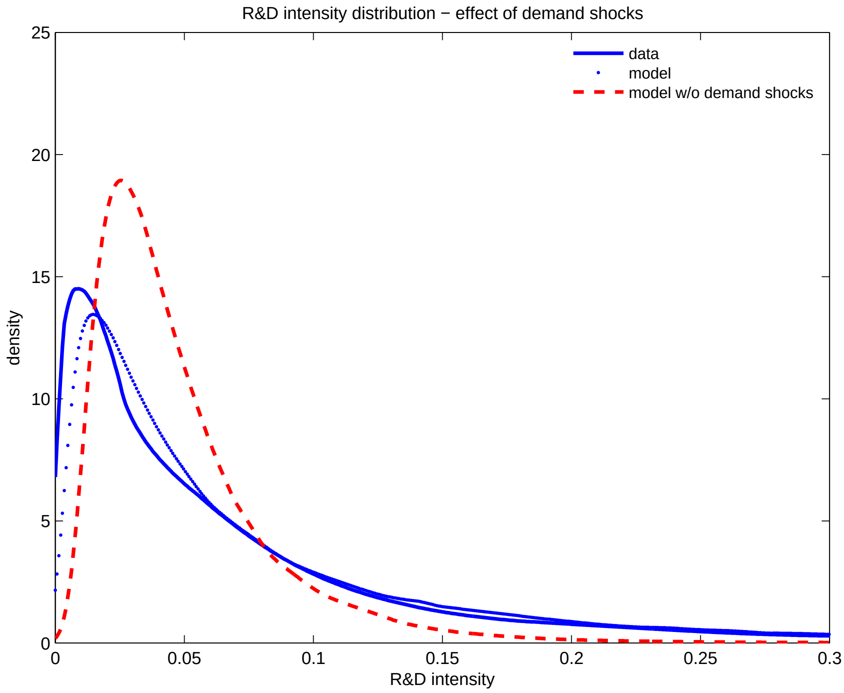
c_0/Z	c_1	η	Z	β_z	o_z	β_q	σ^2_Y
498.3	5.366	0.0814	10296	0.614	1499	0.369	0.0298
(126)	(0.1)	(0.02)	(268.7)	(0.02)	(258.9)	(0.017)	(0.003)
σ^2_W	σ^2_{RD}	κ	g	l	μ	δ	$\frac{g_2}{g}$
0.003	0.360	153.9	0.0147	21.7	1.47	0.1103	0.718
(0.002)	(0.07)	(1.96)	(0.0008)	(0.56)	(0.04)	(0.002)	(0.015)
	ϕ_τ	ϕ^*_τ	K_τ	γ_τ	ε_τ	π_τ	
τ_1	0.942	0.878	0.696	0	0	0	
	(0.003)	(0.005)	(0.01)				
τ_2	0.058	0.121	0.287	0.0938	0.274	0.233	
	(0.003)	(0.005)	(0.01)	(0.003)	(0.03)	(0.01)	
τ_3	0.0002	0.001	0.017	0.1091	1.827	0.442	
	($4{\cdot}10^{-5}$)	(0.0001)	(0.002)	(0.002)	(0.58)	(0.03)	
Selected moments (1997)							
		model	data			model	data
$E(PR)$		468.6	477.8	$Corr(PR, N)$		0.023	-0.03
$std(PR)$		183.9	173.2	$Corr(PR, Y)$		0.14	0.124
$E(Y)$		12988	12872	$Corr(Y, \frac{\Delta Y}{Y})$		-0.02	-0.006
$Std(Y)$		25775	23183	$Corr(PR, \Delta PR)$		-0.365	-0.342

Notes: Estimation w/R&D targets both non-R&D and R&D moments. Standard errors in parenthesis. Minimum of objective function: 331.518

type-specific R&D intensity becomes $wc(\gamma_\tau)/Z$, and we end up with two R&D intensity observations with a dispersion of $std(RI) = 0.008$. Consequently, most of the dispersion is due to demand shocks and measurement error. Since R&D is proportional to the state variable k, demand shocks are important to generate heterogeneity in R&D intensity. Simulation without demand (column 3) reduces the dispersion from 0.121 to 0.036.

Notice that with demand shocks, heterogeneity via type-specific innovation intensities γ_τ could potentially create dispersion through its effect on the number of

Figure 3: R&D intensity, value added and worker productivity.



Notes: The figure shows kernel densities on actual data (1997) and on simulated data, with and without demands shocks

products.[19] But since the estimation produces an equilibrium in which there are 131 firms of type 2 for each type 3 firm, this ends up being relatively unimportant.[20]

In contrast, demand shocks play no role at explaining the shape of the distribution of the level of R&D spending. The results in table 4 show that the estimation misses the average R&D expenditures by 29 percent (2907 in the data, 2072 in the model) but matches both the median and dispersion quite well. Dispersion is explained by three factors: Type-heterogeneity in innovation intensity (γ_τ), within-type dispersion in product size (k), and R&D measurement error (σ^2_{rd}). Overall, the main contribution comes from type-heterogeneity and the product birth-death process, while measurement error plays a minor role. Table 5 column (4) shows simulation results without R&D measurement error. For this case the standard devi-

[19]As we note in section 3.4A, dispersion induced by the product birth-death process is of second order importance.

[20]Note that we sample R&D observations for firms below 50 workers. Among firms with positive R&D after sampling, there are 81 firms of type 2 for each type 3 firm.

Table 4: Estimation including R&D moments: model fit in 1997

Moments	Data	Estimation (benchmark)	Estimation (w/ R&D)	Moments	Data	Estimation (benchmark)	Estimation (w/ R&D)
$E(RI)$	0.084	0.038	0.089	$cor(RI, Y)$	-0.170	-0.509	-0.267
$std(RI)$	0.116	0.024	0.121	$cor(RI, PR)$	-0.188	-0.150	-0.148
$Med(RI)$	0.044	0.036	0.049	[1] $\frac{\#Firms_{RI=0}}{\#Firms}$	0.924	0.808	0.920
$E(RD)$	2907	698	2072	[2] $cor(RI, RI_{+2})$	0.688	0.810	0.440
$std(RD)$	5619	1081	5494	[3] $\frac{E(Y)_{RI>0}}{E(Y)_{RI=0}}$	4.44	3.53	4.40
$Med(RD)$	1163	455	1131	[4] $\frac{E(N)_{RI>0}}{E(N)_{RI=0}}$	4.44	3.21	3.68

Notes: [1]Fraction of firms with a zero R&D observation. [2]Correlation R&D intensity between 1997 and 1999. [3]Average value added for firms with positive R&D observations, relative to firms with zero R&D. [4]Average employment for firms with positive R&D observations, relative to firms with zero R&D.

ation of R&D falls by 14 percent, from 5494 to 4717. Firm type heterogeneity, on the other hand, is important. Computing standard deviation only on type 2 firms gives a dispersion in RD of only 2234. The reason why type heterogeneity is important for dispersion in RD and not for RI, is that R&D spending is proportional to number of products. Firm type 3 invests 2.6 times as much in R&D per product than firm type 2 and has on average 7 times more products. However, since RD scales with number of products, RI is on average only 2.6 times higher. In contrast, the distribution of product size (k) is of first order importance for the distribution of R&D spending, $RD = wc(\gamma_\tau)k$. Even though firm type 3 accounts for only 0.1 percent of firms in equilibrium, the associated type 3 logarithmic product-size distribution has a large dispersion (more than 142 times the variance of type 2) which contributes disproportionally to the overall dispersion in R&D spending.

Table 5: Effect of shocks on simulated moments in 1997.

Moments	All shocks	no shocks	no demand shock	no measurement error	no q shocks
	(1)	(2)	(3)	(4)	(5)
$E(RI)$	0.089	0.043	0.049	0.075	0.089
$std(RI)$	0.121	0.008	0.036	0.067	0.121
$Med(RI)$	0.049	0.042	0.039	0.052	0.049
$E(RD)$	2072	2159	2129	2035	2065
$std(RD)$	5494	4900	5535	4718	5449
$Med(RD)$	1131	1313	1170	1306	1127
$corr(RI, Y)$	-0.267	0.284	-0.023	-0.357	-0.270
$corr(RI, PR)$	-0.148	0.30	-0.040	-0.181	-0.421
$\frac{\#Firms_{RI=0}}{\#Firms}$	0.920	0.929	0.923	0.922	0.920
$corr(RI, RI_{+2})$	0.440	1	0.074	0.842	0.434
$\frac{E(Y)_{RI>0}}{E(Y)_{RI=0}}$	4.40	4.39	4.19	4.51	4.42
$\frac{E(N)_{RI>0}}{E(N)_{RI=0}}$	3.68	3.66	3.51	3.77	3.64

Notes: Table shows the simulated R&D moments when we counterfactually shut down shocks. (1) displays the results from the estimation. (3) shut down demand shocks. (4) shut down measurement error. (5) shut down stochastic quality improvement. (2) combines (3)-(5), i.e. no demand shocks, no q-shocks and no measurement error

B. Non-performers From table 4 we observe that the model gets the fraction of firms with zero R&D with a value 0.92. Note that this fraction is higher than the true fraction of non-performers, due to the sampling of R&D observations for firms with less than 50 workers. Notice also that we do not target the fraction of non-performers for firms above 50 workers, but the model gives a fraction of 0.71, in line with what we observe in the data (0.65). Firms reporting positive R&D are on average 4.40 times larger than firms with zero R&D. However, in terms of employment, the model underpredicts the relative size by 17 percent. Consequently, the model predicts too high value added per worker for R&D performers relative to non-performers.

In the data, average value added per quality adjusted worker is the same for the two groups, whereas in the model, performers have 25 percent higher worker productivity. Consider the case where we shut down measurement error, and demand

and supply shocks. Then, worker productivity is given by

$$PR = \frac{Z}{Z(w+\kappa)^{-1}q_{\tau_i}^{-1} + c(\gamma_{\tau_i})}.$$

Note that for non-performers, $q = 1$ and $\gamma = 0$. Innovating firms ($q > 1$) need fewer production workers per product than firms that don't innovate. This contributes to higher productivity PR for innovating firms. On the other hand, the presence of research workers in the denominator, $c(\gamma_{\tau_i})$, works in the opposite direction. The estimated parameters imply that the former effect dominates.[21]

C. Productivity, size and R&D intensity Table 4 shows that the estimated model produces a correlation between research intensity and productivity of -0.148. As explained in section 3.4C, this negative relationship can be generated by a combination of demand shocks and value added measurement error.[22] In addition, type heterogeneity can in general drive the correlation in either direction. To isolate the effect of type heterogeneity, recall that the correlation between R&D intensity and productivity can be written as follows

$$corr\left(\frac{wc(\gamma_\tau)}{Z}, \frac{Z}{Z(w+\kappa)^{-1}q_{\tau_i}^{-1} + c(\gamma_{\tau_i})}\right)$$

where all sources of heterogeneity (demand, supply and measurement error), except firm types, are shut down. The correlation is now completely determined by type differences in quality jumps q_τ and innovation intensity γ_τ. In this case, we see from table (5) column 2 that $corr(RI, PR)$ is positive at 0.3, which implies that type 3 ends up with higher R&D intensity and higher productivity than type 2. The labor saving feature of larger quality jumps ($q_3 > q_2$) dominates the higher demand for research workers ($c(\gamma_3) > c(\gamma_2)$). Consequently, demand shocks and measurement error in value added produce the negative relationship in the estimated model. Simulation

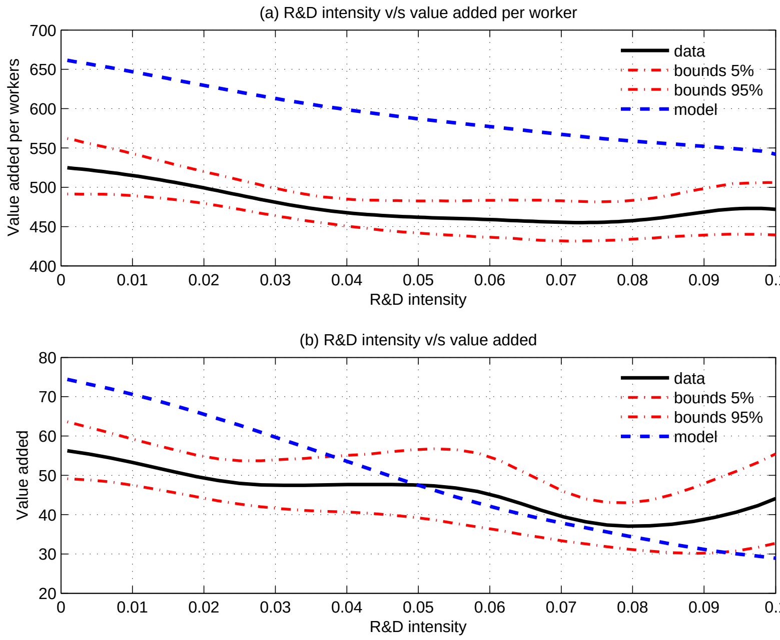
[21]There are several factors that might explain why the data shows identical worker productivity across the two groups. First, if we assume that the fraction of the firm's employment involved in R&D are the same in the data as in the model (8 percent), then firms that report positive R&D have about 9 percent higher value added per production worker than firms with zero R&D. In addition, because of the sampling of R&D observations in the data, some firms are observed with zero R&D even if they do innovate. Finally, the equality of worker productivity between the two groups in the data is a result of the quality adjusted measure of employment. If we instead use non-quality adjusted employment, firms that report positive R&D have 11.2 percent higher value added per worker than non-performing firms.

[22]Notice that R&D measurement errors and q shocks will drive down this correlation to zero.

without demand variations gives a correlation of -0.04 (table 5, column 3).

Regarding the relationship between R&D intensity and size there is also a tension between type-heterogeneity (causing positive correlation) and demand and value added measurement error (causing a negative correlation). With firm types as the only source of heterogeneity, the correlation between size and R&D intensity is 0.28. However, demand shocks and value added measurement error drives the correlation down to -0.26, most of which is due to demand shocks. Shutting down this variation gives a correlation of -0.02.

Figure 4: R&D intensity, value added and worker productivity.



Notes: Kernel regressions of R&D intensity on value added per worker (panel a) and value added (panel b). Both panels compares using actual data (1997), with 95 percent confidence bounds, with simulated data from the estimated model.

Figure (4) depicts both size and worker productivity against different levels of R&D intensity. For the relationship between R&D intensity and productivity, panel (a), the model matches the slope, because we overestimate the average productivity among R&D performers as explained in section 4.3B. Regarding, the relationship between R&D intensity and size, panel (b), the model succeeds in explaining this relationship over the range between the median and mean of the R&D intensity distribution. But firms in the lower end of the R&D intensity distribution are larger in the model than in the data. In the model, firms with large demand draws tend to be large in size and have low R&D intensity relative to firms with small demand draws.

For the benchmark estimation we found that the correlation between R&D intensity and size was too negative compared with the data (see table (1)). When adding R&D moments to the estimation the model correlation becomes less negative without resorting to smaller demand shocks.[23] The reason is that firms have more products on average relative to the benchmark (42 percent) which means the impact of demand shocks falls. In addition, type heterogeneity contributes more positively to this correlation in the R&D estimation than in the benchmark.[24] Consequently, when adding R&D moments we get more heterogeneous innovators. Demand shocks are key to allow for more type heterogeneity and at the same time not violate the negative correlation between R&D intensity and size.

D. Persistence: $corr(RI, RI_{+2})$ We get too low R&D intensity persistence in the model (0.44) compared to data (0.69), due to measurement error in R&D. Without shocks (shutting down demand, supply and measurement error) would imply perfect persistence (i.e. constant R&D intensity over time). Simulation without demand shocks drives the persistence down to 0.075, while simulating without measurement error gives a persistence of 0.842. To get a correlation of 0.44 we thus need product demand shocks. These shocks stay with the firm until it loses the product and thereby create persistence in value added which carries over to research intensity.

E. Selection effect: Recall that in the benchmark estimation, the fraction of R&D performing firms is 0.19. This is more than twice the fraction we observe in the data. The R&D estimation correctly gets the fraction of firms.[25] Consequently, when we also match R&D moments we end up with fewer performers, as shown in table 6. These firms are also bigger and more research intensive than in the benchmark estimation. Relative to non-performers they have on average 4.40 (3.53) times higher value added and 3.68 (3.21) more workers in the R&D estimation (benchmark estimation). The aggregate research intensity is now 4.6 percent compared to 2.2 percent

[23]The standard deviation of demand shocks is 15500 and 15000 in the benchmark and R&D estimation respectively.

[24]To make this comparison we shut down all shocks, leaving firm types as the only source of heterogentiy in R&D intensity. The correlation is then 0.28 and 0.14 in the R&D estimation and benchmark, respectively.

[25]The true fraction of R&D performing firms differs from the observed, due to sampling of R&D observations. The equilibrium fraction of R&D performing firms is 0.38 in the benchmark and 0.12 in the R&D estimation.

in the benchmark.[26] Overall, we end up with innovating firms being fewer, bigger, and more research intensive. As a result, the growth contribution of reallocation across firm types increases from 44.5 percent to 71.8 percent.[27]

Table 6: Model fit II. Estimation with R&D moments.

	Data	Estimation (w/R&D)	Estimation (benchmark)
Firms with $RI > 0$	8%	8%	19%
$E(Y_{RD>0})/E(Y_{RD=0})$	4.44	4.40	3.53
$E(RD_{RD>0})/E(Y_{RD>0})$	6.4%	4.6%	2.2%

Notes: Estimation w/R&D targets both non-R&D and R&D moments. Minimum of objective function: 331.518

4.4 External validity

We have shown that including R&D moments have improve the ability of the model to match cross-section moments. Now we perform a series of robustness checks, to asses the performance of the model for moments we did not match.

4.4.1 R&D reponse to tax-credit

We start comparing the model response to a R&D subsidy with the outcome of a 2002 R&D reform in Norway, analyzed in Bøler et al. (2014). They exploit the introduction of a tax-credit scheme that enabled firms to deduct 20 percent of R&D expenditures (up to a threshold of NOK 4 million) from their tax bill. Effectively, this reduced the marginal cost of R&D with 20 percent for firms with less than 4 million in R&D expenditures. They conclude that the reform induced firms that had positive R&D pre-reform, but less than NOK 4 million, to increase R&D expenditures with between 35 and 72 percent (depending on identification strategy) during the period 2003-

[26] Aggregate research intensity is measured as the ratio of average R&D spending to average value added among R&D performers. It is not an explicit target in the estimation. However, it is implicitly targeted as we target relative value added between performers and non-performers, average value added and average R&D expenditures.

[27] This result is is robust to leaving out the fraction of non-performing firms as a targeted moment in the estimation.

2005.[28]

We use our estimated model to evaluate the impact of a similar, but simplified reform. In the model economy, we implement a 20 percent subsidy to incumbent firms' R&D investment, with no upper threshold. The optimal innovation choice is then still proportional to number of products $I = \gamma_\tau k$ and we can analyze the effect on innovation intensities γ_τ using the solution to the firm problem in (11):

$$\begin{aligned} wc'(\gamma_\tau) &= v_\tau \\ v_\tau &= \frac{ZE\pi(q_\tau) - wc(\gamma_\tau)}{r + \delta - \gamma_\tau} \end{aligned}$$

We look at the effect of this policy reform from two perspectives. First we look at the firm's response to a subsidy (s) equal to 20 percent of R&D expenditures, ignoring equilibrium effects, i.e. we hold the wage rate w and aggregate destruction δ constant. Firms now face a net (of subsidy) R&D cost of $(1 - s)wc(\gamma)$. Firm type 2 and 3 respond by increasing their innovation intensity by 5.3 and 5.9 percent, respectively. Consequently, gross R&D expenditures per product increase by 39 percent (type 2) and 44 percent (type 3).

We interpret this as a short run effect. With a higher innovation intensity, firms will gain more products over time and thus invest more in R&D. If we still assume a constant wage and aggregate destruction rate, firm type 2 will on average have 27 percent more products and thus a total increase in gross R&D expenditures of 76 percent.[29] However, in equilibrium, when individual firms innovate more intensely, the aggregate destruction rate δ goes up. In the new stationary state, firm types increase their innovation intensity by 4.1 and 4.2 percent, and the aggregate destruction rate increases by 3.5 percent. On average, firms employ 30 percent more researchers.[30] Equilibrium effects consequently dampen the initial response, since innovation intensities and firm product size are decreasing in δ.

[28] Bøler et al. (2014) report an increase in R&D expenditures between 0.30 and 0.54 log points. We use the transformation $X_t / X_{t-1} = exp(logpoint) - 1$ to arrive at the percentage change.

[29] Keeping the destruction rate δ constant imply that firm type 3 choose $\gamma_3 > \delta$ and the expected number of products is $+\infty$. In equilibrium, however, the aggregate destruction rate adjusts such that $\gamma_3 < \delta$. Notice that the number of researchers also grows by 76 percent as the wage rate is constant.

[30] The reason we report the stationary state effect of R&D subsidy on R&D workers is that the subsidy increases the growth rate of the economy. Hence, in the new stationary equilibrium, real wage and thus the research wage bill grows at a higher rate than in the initial equilibrium. Hence, it is not possible to compare the level of R&D expenditures across the two equilibria.

4.4.2 Large firms

Recall that in our estimation we use R&D moments including all firms. Therefore, we use the ability of the model to match moment for large firms in our sample as an external validity test. In table (7) we display how the model compares with the data for firms with more than 50 workers. Indeed, the model matches quite well the level, dispersion and skewness of the R&D intensity and R&D expenditures. Also it generates a lower correlation between R&D and size and at the same time it maintains a high negative correlation between R&D intensity and productivity. This indicates that the negative correlation of R&D intensity and size is driven by small firms. The intuition behind this result, is that larger firms tend to have more products than smaller firms, and hence the impact of product-specific demand shocks is lower. Moreover, the model matches the relative size (in terms of employment and value added) of performers. Finally, the model underestimate the persistence of R&D intensity for large firms.

Table 7: Fit of the model for large firms (non-targeted)

Moments	Data	Model	Moments	Data	Model
$E(RI)$	0.06	0.04	$corr(RI,Y)$	-0.02	-0.05
$std(RI)$	0.09	0.04	$corr(RI,PR)$	-0.13	-0.15
$Med(RI)$	0.03	0.03	$^{1}\frac{\#Firms_{RI=0}}{\#Firms}$	0.66	0.71
$E(RD)$	2906	2071	$^{2}corr(RI,RI_{+2})$	0.68	0.34
$std(RD)$	5618	5494	$^{3}\frac{E(Y)_{RI>0}}{E(Y)_{RI=0}}$	1.36	1.79
$Med(RD)$	1163	1131	$^{4}\frac{E(N)_{RI>0}}{E(N)_{RI=0}}$	1.34	1.50

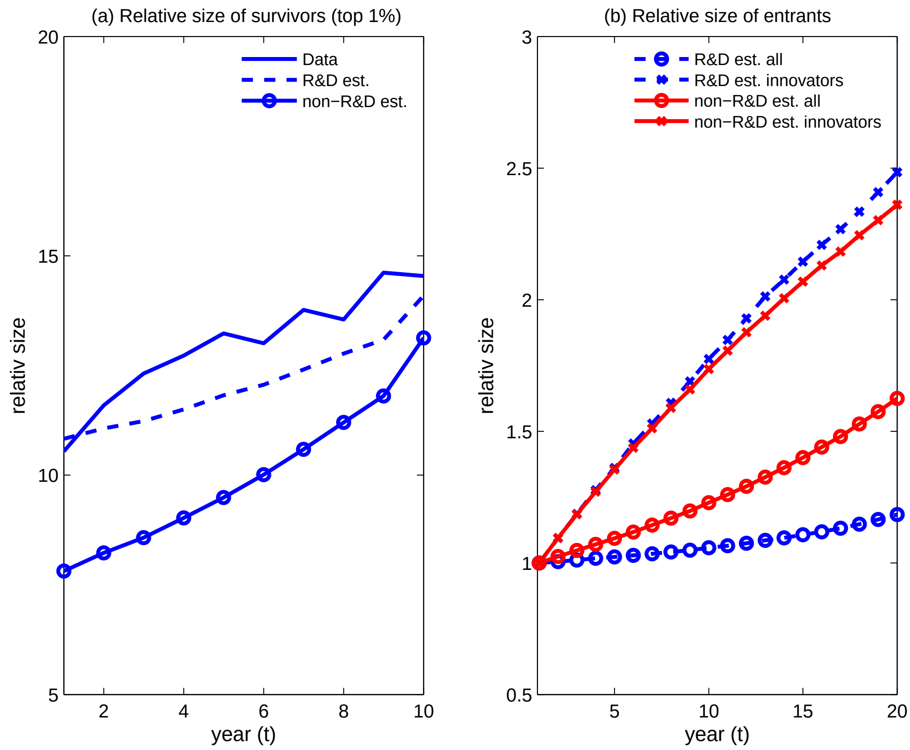
Notes: We use parameters from estimation with R&D moments. Moments for firms above 50 workers were not used in the estimation. [1]Fraction of firms with a zero R&D observation. [2]Correlation R&D intensity between 1997 and 1999. [3]Average value added for firms with positive R&D observations, relative to firms with zero R&D. [4]Average employment for firms with positive R&D observations, relative to firms with zero R&D

4.4.3 Life cycle of firms

Luttmer (2011) finds that type heterogeneity is needed to match the size distribution and the relative young age of large firms. We check the size history of those firms that reach the top 1 percent in the size distribution. Again, this a feature that we have not included in our estimation. [31]

In figure (5) panel (a), we plot the relative size of firms that end up among the top 1 percent largest firms. Both in the data and in the model, we construct a 10 year sample and select the firms that are in the top 1 percent at the end of the sample. We do not include entrants and exitors in both the data and simulation. Then, we compare how the mean value added of these firms has grown relative to the average firm size in the economy over sample period. In the simulation we compute this ratio for both the benchmark and the R&D estimation.

Figure 5: Relative size of survivors (top 1%) and entrants.



Notes: Panel (a): graphs show the average value added of continuing firms, conditional on being in the top 1 percent of the size distribution in period t=10, relative to average value added of incumbents over time. For the data, t=1 corresponds to 1996, and t=10 to 2005. Panel (b): Average value added of a simulated cohort of entrants (period t=1) relative to average value added of incumbents over time (t+1 normalized to one).

We find that in the R&D estimation, the both the initial size and the growth history of large firms are in line with the data. In contrast, for the benchmark estima-

[31] Recall that in our estimation we only match relative size after 4 years)

tion firms grow faster during the transition to the top 1 percent. We also see from table (12) and (13) that the implied growth of average value added from 1997-2001 is higher in the benchmark compared to the R&D estimation. To shed some light on the mechanism behind why firms grow faster for the bechmark estimation we study the relative growth of a cohort of entrants. Figure (5) panel (b) shows that entry firms grow faster than compared to the R&D estimation. The reason is that in the benchmark, an entry cohort consists of a larger fraction of fast growing innovating firms. Recall that the entry share of type 1 firms (non-performers) is 0.75 and 0.94 in the benchmark and R&D estimation, respectively.

5 Policy: R&D subsidy

As in most endogenous growth models, the decentralized equilibrium growth rate in the Klette-Kortum (2004) model need not equal the welfare maximizing growth rate. Hence, there is a potential role for policy interventions. There are several sources of inefficiencies which can lead to too little growth in equilibrium.[32] First, all firms innovate at the technology frontier (frontier quality is common knowledge). Consequently, the social return to a successful innovation lasts forever. In contrast, the private return to a firm only lasts until the product, and its associated profit, is lost. This private depreciation of R&D investments contributes to insufficient innovation. In addition, the reallocation channel within the Lentz-Mortensen framework suggests that firm heterogeneity plays an important role. Through the process of equilibrium selection, the steady state distribution of products K_τ across types differs from the entry distribution ϕ_τ. In general, the bigger share of goods K_τ produced by firms generating high quality innovations, the larger is growth. However, the amount of reallocation generated in equilibrium tends to be too low due to a business stealing effect.

Real world subsidies, however, often fails to target the most effective innovators. By studying stylized reforms, we show that failure to target the best innovators may lead to subsidies creating small, or indeed adverse, growth effects. The effectiveness of a subsidy depends crucially on how it influences the reallocation channel. In general, policies that increase R&D (higher γ) and shift the composition of products (K) to firms producing high quality innovations will stimulate growth.

We run four policy experiments where we subsidize R&D expenditures as in section 4.4 and evaluate the impact on the growth rate along balanced growth paths.

[32]See for example Acemoglu et al. (2013), Li (2001), and Atkeson and Burstein (2014)

Table 8: The effect of R&D subsidies on aggregate growth.

Reform:	g	K_1	K_2	K_3
Estimation with R&D moments	1.47	69.6	28.7	1.7
Subsidy on:				
(I) Incumbents and entrants	1.53	69.5	28.6	1.8
(II) Incumbents	1.82	67.0	28.5	4.4
(III) Entrants and firm type 2	1.40	69.9	29.8	0.30
(IV) Firm type 3	2.45	66.2	21.5	12.3

Notes: numbers as percentages

In each experiment, we fix the total subsidy to 0.7 percent of the wage income of production workers (which is a proxy for GDP), financed by a non-distortionary tax on consumers. The results are shown in table 8.

In experiment (I) all firms receive R&D subsidy, while in (II) only incumbent firms are subsidized. The growth rate increases from its initial level of 1.47 percent, to 1.53 when we subsidize all firms and to 1.82 percent when we only subsidize incumbents. Clearly, the growth effect is much larger when the policy reform targets only incumbent firms. The key to understand this result is to look at the change in the composition of products across firm types. In the second experiment, the share of products produced by the high type is 4.4 percent, compared to 1.8 percent in the first experiment. Table 8 shows that firm type three (the high type) grows at the expense of type one firms (the imitators). Innovation by low types has an adverse effect on the incentive to innovate for high types (through the impact on aggregate destruction). In particular, high entry innovation reduces the incentive for incumbent firms to innovate, and thus helps sustain a large fraction of firm type 1 in equilibrium, which is bad for growth. This is a source of misallocations of R&D spending across firms, which can have large negative growth effects. [33]

Overall, going from reform (I) to (II) raises the growth rate with 19.0 percent.

[33] Lentz and Mortensen (2014) extends Lentz and Mortensen (2008) by solving for the planners balanced growth path allocation. Compared to the equilibrium outcome, the optimal allocation implies a twice as high growth rate.

In comparison, in the benchmark estimation, going from reform (I) to (II) raises the growth rate with only 6.6 percent. The intuition behind this result is that the implementation of the R&D subsidy matters more in once we account for R&D in the estimation, because reallocation becomes more important.

Consider the effect of implementing stylized size-dependent subsidies (size being measured in terms of R&D spending).[34] Reform (III) gives a subsidy to incumbent firm type 2 and potential entrants, while reform (IV) targets only firm type 3. Reform (III) induces a reallocation of products from type 3 firms to type 2 firms. This effect is so strong that aggregate growth falls. The reason aggregate growth falls is a combination of two effects. First, when subsidizing entrants and firm type 2, these firms raise their innovation intensity and gain products at the expense of firm type 3 (reallocating products away from the high quality innovator). In addition, the induced increase in aggregate innovation rate adverseley impacts the incentive for firm type 3 to innovate (through the creative destruction channel). This causes its innovation intensity to fall, which induces further reallocation of products away from the high quality innovator. Reform (IV) on the other hand, creates a massive reallocation of products to firm type 3, and hence the growth rate increases substantially.[35]

Acemoglu et al. (2013) finds that an optimal R&D policy involves subsidizing both entry and high incumbent type. In that model there is also firm type heterogeneity with respect to innovation ability (high and low types), realized upon entry. As in our model, subsidizing the high type encourages the expansion of this type at the expense of the low types. However, we get different implications of an entry subsidy. The reason for this is that, in contrast to our model, in Acemoglu et al. (2013) all incumbent firms innovate. Moreover, the high type faces a fixed probability of becoming a low type over time. Consequently, inflow of new firms (both high and low) helps to sustain a high type in equilibrium and an entry subsidy raises the equilibrium share of firms producing high quality innovations. Our approach also differ from Acemoglu et al. (2013) in the sense that we estimate the model using observation on all firms, whereas they focus on firms with positive innovation activity. Hence, our approach allows for a share of products being produced by non-innovators, and we find that an important margin of adjustment is from non-

[34] It is not a true size-dependent reform, in the sense that it targets firm types rather than actual size, but it highlights some important negative reallocation effects arising from subsidizing small firms in the model. The reason is that small firms in the model are predominantly firm type 2.

[35] It is important to emphasize that a type-dependent reform exaggerates the negative reallocation effects relative to a true size-dependent reform, since also highly innovative firms will be below the size cut off (in particular young firms) and some less innovative firms will be above the cut off (in particular after the reform).

innovators to innovators.

The growth effects in our experiments should be interpreted with caution, since we only consider balanced growth path effects. First, because firm type 3 constitute a very small fraction of entry firms (only 0.02 percent), the transition to a new balanced growth path might be slow. More broadly, Atkeson and Burstein (2014) show, through numerical examples, that R&D subsidies in Klette-Kortum type endogenous growth models tend to have small short to medium-term effects. It should be noted, however, that they consider a framework in which there is no misallocation of research activity across firms (type heterogeneity). In our model, a subsidy generates more research effort in the aggregate, but in addition it also influences the allocation of R&D (and products) across firms.

6 Conclusion

In this paper we estimate a general equilibrium model of firm-level innovation using observations on size, productivity and R&D expenditures from a panel of Norwegian manufacturing firms. In particular, we estimate an extended version of the Klette-Kortum model, developed in Lentz and Mortensen (2008), to quantify the relative importance of different sources of R&D heterogeneity and the link to reallocation and growth.

We find a larger role of reallocation to aggregate growth when ask the model to explain (in addition to other moment) R&D intensity dispersion and a negative correlation between R&D intensity and size observed in the data. We first replicate the study of Lentz and Mortensen (2008) on a sample of Norwegian manufacturing firms. To make our estimation comparable to Lentz and Mortensen, we first exclude observations on R&D, and find that the reallocation effect accounts for 44.5 percent of aggregate growth. This magnitude is similar to what they find, which is 49 percent for the manufacturing sector. However, we miss some key empirical R&D patterns: Firms doing R&D are too many, too small and invest too little in R&D relative to data. Then, we re-estimate the model adding R&D moments. We find that the model has a good fit to both R&D and non R&D moments. The estimated model fits the empirical distribution of R&D effort (mean, dispersion and skewness) as well as the negative correlation between research intensity and size. More importantly, now the new parameters imply a larger role for reallocation, explaining 72 percent of aggregate growth. Quantitatively, product demand shocks, measurement error, and firm differences in the ability to do R&D are all key to account for the observed heterogeneity in R&D effort.

Then, we find that demand shocks and innovative differences are important to explain the shape of the R&D distribution and the correlation between R&D intensity and size. Interestingly, when we shut down demand shocks we observe that we miss fit on the right tail of the distribution. The model without demand shocks generates too few firms with high R&D intensity. Intuitively, firms that draw bad demand shocks will have high R&D intensity.

We also conducted several external validity tests. We find that our model is consistent with the firm-level response to R&D subsidies that are in line with micro evidence from a natural experiment (as in Bøler et al. (2014)). In the short run, firms increase their R&D spending with roughly 40 percent in response to a 20 percent R&D subsidy. Furthermore, the model is able to explain several cross sectional and dynamics moments for R&D, size and productivity for large firms. Also, the model is able to explain features of the life cycle of firms over longer horizon.

Finally, we use the estimated model to explore quantitatively the growth effects of R&D subsidies. We find that subsidies are successful in increasing growth, but that the effect depends crucially on how it influences the reallocation channel.

The model abstracts from openness and trade, which could potentially affect our results. For example, diffusion of ideas across borders could reduce the importance of domestic R&D, while foreign competition and access to foreign markets could change the firms' incentives to innovate and the link between R&D subsidies, reallocation and growth. Eaton and Kortum (2001) and Atkeson and Burstein (2010) consider innovation and growth in open economies, but abstract from product innovation by incumbent firms. An interesting extension would be to consider the Klette-Kortum, Lentz-Mortensen model in an open economy setting to understand the link between international competition, firm dynamics and innovation.

References

[1] Acemoglu, D., Akcigit, U., Bloom, N., and W. Kerr (2013): "Innovation, Reallocation and Growth", NBER Working Paper No. 18993.

[2] Aghion, P, Akcigit, U. and P. Howitt (2013): "What do we learn from Schumpeterian Growth Theory?", NBER Working Paper No. 18824.

[3] Aghion, P. and P. Howitt (1992): "A Model of Growth Through Creative Destruction", *Econometrica*, 60, 323-352.

[4] Akcigit, U. (2009): "Firm Size, Innovation Dynamics and Growth", University of Pennsylvania, Working Paper.

[5] Akcigit, U. and W. Kerr (2010): "Growth Through Heterogeneous Innovations", NBER Working Paper No. 16443

[6] Atkeson, A. and A. Burstein (2014): "Aggregate Implications of Innovation Policy", Working Paper, UCLA.

[7] Atkeson, A. and A. Burstein (2010): "Innovation, Firm Dynamics, and International Trade", Journal of Political Economy, 118(3), 433-484.

[8] Aw, B.Y., Roberts, M.J. and D.Y. Xu. (2011): "R&D Investment, Exporting, and Productivity Dynamics", *American Economic Review*, 101(4), 1312-1344.

[9] Bartelsman, E., Haltiwanger, J. and Scarpetta, S. (2013): "Cross-Country Differences in Productivity: The Role of Allocation and Selection", *American Economic Review*, 103(1), 305-334.

[10] Bøler, E.A., Moxnes, A. and K.H. Ulltveit-Moe (2014): "Technological Change, International Sourcing and the Joint Impact on Productivity", Working Paper: https://dl.dropboxusercontent.com/u/3719891/inno_paper32.pdf

[11] Cohen, Wesley M., and Steven Klepper (1992): "The Anatomy of Industry R&D Intensity Distributions", *American Economic Review* 82 (4), 773–799.

[12] Cohen, Wesley M., and Steven Klepper (1996): "A Reprise of Size and R&D", *Econimic Journal* 106 (July), 925–951.

[13] Doraszelski, U. and J. Jaumandreu (2013): "R&D and productivity: Estimating endogenous productivity", *Review of Economic Studies*, 80(4), 1338-1383.

[14] Eaton, J. and S. Kortum (2001): "Technology, trade, and growth: A unified framework", *European Economic Review*, 45,742-755.

[15] Foster L., Haltiwanger J. and Krizan C. (2001): "Aggregate productivity growth. Lessons from microeconomic evidence", NBER chapters, in: New Developments in Productivity Analysis, 303–372.

[16] Grossman, G.M., and E. Helpman (1991): Innovation and Growth in the Global Economy. MIT Press, Cambridge, MA, 1991.

[17] Harrison, R., Jaumandreu, J., Mairesse J. and B. Peters (2008): "Does innovation stimulate employment? A firm-level analysis using comparable micro data on four European countries", NBER Working Paper No. 14216.

[18] Klette, T.J. and S. Kortum (2004): "Innovating Firms and Aggregate Innovation", *Journal of Political Economy* 112(5), pp. 986-1018

[19] Klette, T.J. and Z. Griliches (2000): "Empirical Patterns of Firm Growth and R&D Investment: A Quality Ladder Model Interpretation", *The Economic Journal*, 110, pp. 363-387

[20] Lentz, R. and D.T. Mortensen (2008): "An empirical model of growth through product innovation", *Econometrica*, 76(6), pp. 1317-1373

[21] Lentz, R. and D.T. Mortensen (2014): "Optimal Growth Through Product Innovation", Unpublished Working Paper

[22] Li, C.W (2001): "On the Policy Implications of Endogenous Technological Progress", *The Economic Journal*, 111, C164-C179

[23] Petrin, A. and J. Levinsohn. (2012): "Measuring Aggregate Productivity Growth Using Plant-Level Data", *The RAND Journal of Economics*, 43(4), 705-725.

[24] Raknerud, A., Rønningen, D. and T. Skjerpen (2004): "Documentation of the Capital Database", Statistics Norway Documents 2004/16.

[25] Statistics Norway (2004): "Research and development in the Business Enterprise Sector 2001-2002". *Official Statistics of Norway* D301

A Appendix

Table 9: Descriptive statistics (year 1997).

Statistics	All firms	10-50 workers	> 50 workers
Number of firms	5290	2087	684
Average Value added	12872	10475	57436
Average Employment	27.5	21.9	124.1
Average Productivity	482.2	479.2	467.8
Average R&D expenditure	247	75.6	1470
% R&D>0	7.6	6.9	35.1
Average RI (performers)	0.08	0.09	0.068

Notes: Value added and R&D expenditure are reported in units of 1000,- NOK. The above 50 and 10-50 worker categories are based on quality-adjusted workers). Employment is number of quality-adjusted workers, Productivity is expressed units of 1000,- NOK per quality-adjusted worker. R&D intensity (RI) is the ratio of R&D expenditure divided by value added

Table 10: non-R&D data moments (standard errors in parenthesis)

Moments	1997	2001	Moments	1997	2000
$E(Y)$	12872	14998	$Cor(PR, PR_{+1})$	0.735	0.697
	(317.8)	(479.1)		(0.017)	(0.034)
$std(Y)$	23183	28161	$Cor(PR, \Delta PR)$	-0.342	-0.371
	(796.0)	(1797.3)		(0.043)	(0.048)
$Med(Y)$	4985	6254	$Cor(PR, \frac{\Delta Y}{Y})$	-0.126	
	(105.0)	(140.4)		(0.016)	
$E(W)$	8144	10828	$Cor(PR, \frac{\Delta N}{N})$	0.037	
	(196.0)	(321.8)		(0.017)	
$std(W)$	14394	18991	$E(\frac{\Delta Y}{Y})$	-0.007	
	(472.6)	(1060.3)		(0.009)	
$Med(W)$	3157	4655	$std(\frac{\Delta Y}{Y})$	0.612	
	(67.8)	(114.2)		(0.040)	
$E(PR)$	477.8	503.6	$Cor(Y, \frac{\Delta Y}{Y})$	-0.006	
	(2.4)	(2.8)		(0.009)	
$std(PR)$	173.8	168.0	*within*	0.330	
	(7.4)	(7.8)		(0.074)	
$Med(PR)$	444.1	473.5	*between*	0.311	
	(1.68)	(2.02)		(0.071)	
$Cor(Y, W)$	0.950	0.949	*cross*	0.133	
	(0.004)	(0.007)		(0.059)	
$Cor(PR, N)$	-0.030	-0.016	*exit*	0.226	
	(0.010)	(0.015)			
$Cor(PR, Y)$	0.124	0.142			
	(0.014)	(0.019)			
survivors	5290	3564			
		(33.8)			

Notes: Average growth rate $E(\Delta Y/Y)$ includes exiting firms. They contribute with a -1 observation. The within, between, cross and exit moments are the components of a standard empirical labor productivity growth decomposition, over the period 1997-2001. See Lentz and Mortensen (2008, p. 1335)

Table 11: R&D data moments (standard errors in parenthesis)

Moments	1997	2001	Moments	1997	2001
$E(RI)$	0.084	0.090	$Cor(RI, Y)$	-0.170	-0.103
	(0.006)	(0.006)		(0.038)	(0.045)
$std(RI)$	0.117	0.117	$Cor(RI, PR)$	-0.188	-0.063
	(0.010)	(0.012)		(0.041)	(0.059)
$Med(RI)$	0.044	0.050	1 $\frac{\#Firms_{RI=0}}{\#Firms}$	0.924	0.897
	(0.004)	(0.003)		(0.004)	(0.005)
$E(RD)$	2906	3463	2 $Cor(RI, RI_{+2})$	0.688	
	(280.7)	(416.3)		(0.076)	
$std(RD)$	5619	8019	3 $\frac{E(Y)_{RI>0}}{E(Y)_{RI=0}}$	4.44	
	(880.7)	(1751)		(0.243)	
$Med(RD)$	1163	1250	4 $\frac{E(N)_{RI>0}}{E(N)_{RI=0}}$	4.44	
	(102.4)	(110.7)		(0.233)	

Notes: [1]Fraction of firms with a zero R&D observation. [2]Correlation R&D intensity between 1997 and 1999. [3]Average value added of firms with positive R&D observation, relative to firms with zero R&D. [4]Average employment of firms with positive R&D observation, relative to firms with zero R&D.

Table 12: Model fit non-R&D moments: Benchmark estimation. Data top row, model bottom.

Moments	1997	2001	Moments	1997	2000
$E(Y)$	12872	14998	$Cor(PR, PR_{+1})$	0.735	0.697
	13090	15932		0.706	0.699
$std(Y)$	23183	28161	$Cor(PR, \Delta PR)$	-0.342	-0.371
	23485	27823		-0.363	-0.364
$Med(Y)$	4985	6254	$Cor(PR, \frac{\Delta Y}{Y})$	-0.126	
	5187	6209		-0.093	
$E(W)$	8144	10828	$Cor(PR, \frac{\Delta N}{N})$	0.037	
	8230	9904		0.060	
$std(W)$	14394	18991	$E(\frac{\Delta Y}{Y})$	-0.007	
	14071	16492		-0.017	
$Med(W)$	3157	4655	$std(\frac{\Delta Y}{Y})$	0.612	
	3351	4016		0.735	
$E(PR)$	477.8	503.6	$Cor(Y, \frac{\Delta Y}{Y})$	-0.006	
	471.6	507.1		-0.026	
$std(PR)$	173.8	168.0	*within*	0.330	
	173.8	186.5		0.717	
$Med(PR)$	444.1	473.5	*between*	0.311	
	442.9	475.4		0.113	
$Cor(Y, W)$	0.950	0.949	*cross*	0.133	
	0.964	0.963		0.059	
$Cor(PR, N)$	-0.030	-0.016	*exit*	0.226	
	0.000	0.006		0.110	
$Cor(PR, Y)$	0.124	0.142			
	0.117	0.126			
survivors	5290	3564			
	5290	3581			

Benchmark estimation targets only non-R&D moments. Minimum of objective function: 167.701

Table 13: Model fit non-R&D moments: Estimation with R&D. Data top row, model bottom.

Moments	1997	2001	Moments	1997	2000
$E(Y)$	12872	14998	$Cor(PR, PR_{+1})$	0.735	0.697
	12988	15145		0.704	0.704
$std(Y)$	23183	28161	$Cor(PR, \Delta PR)$	-0.342	-0.371
	25775	31218		-0.365	-0.354
$Med(Y)$	4985	6254	$Cor(PR, \frac{\Delta Y}{Y})$	-0.126	
	5285	5920		-0.116	
$E(W)$	8144	10828	$Cor(PR, \frac{\Delta N}{N})$	0.037	
	8096	9292		0.061	
$std(W)$	14394	18991	$E(\frac{\Delta Y}{Y})$	-0.007	
	14373	17069		-0.0393	
$Med(W)$	3157	4655	$std(\frac{\Delta Y}{Y})$	0.612	
	3472	3869		0.621	
$E(PR)$	477.8	503.6	$Cor(Y, \frac{\Delta Y}{Y})$	-0.006	
	468.6	500.8		-0.02	
$std(PR)$	173.8	168.0	*within*	0.330	
	183.9	201.2		0.650	
$Med(PR)$	444.1	473.5	*between*	0.311	
	447.1	475.4		0.172	
$Cor(Y, W)$	0.950	0.949	*cross*	0.133	
	0.962	0.962		0.025	
$Cor(PR, N)$	-0.030	-0.016	*exit*	0.226	
	0.023	0.034		0.153	
$Cor(PR, Y)$	0.124	0.142			
	0.140	0.154			
survivors	5290	3564			
	5290	3530			

Estimation w/R&D targets both non-R&D and R&D moments. Minimum of objective function: 331.518

Table 14: Model fit: R&D moments 2001

Moments	Estimation (w/ R&D)	Data	Estimation (Benchmark)	Moments	Estimation (w/ R&D)	Data	Estimation (benchmark)
$E(RI)$	0.087	0.090	0.037	$cor(RI, Y)$	-0.270	-0.103	-0.506
$std(RI)$	0.115	0.117	0.023	$cor(RI, PR)$	-0.153	-0.063	-0.162
$Med(RI)$	0.049	0.050	0.034	$\frac{\#Firms_{RI=0}}{\#Firms}$	0.893	0.897	0.765
$E(RD)$	2359	3463	837				
$std(RD)$	5888	8018	1242				
$Med(RD)$	1300	1250	517				