Technology-skill complementarity in a globalized world

Esther Ann Bøler*

Abstract

This paper investigates skill-biased technical change at the firm level using rich Norwegian data. In the theoretical framework, firms invest in R&D to enhance their productivity which has a factor-neutral and a skill-biased component. Firms investing in R&D are found to have higher levels and growth rates of skill-biased productivity. The estimated growth rate of skill-biased productivity is sizable enough to account for the majority of the observed increase in the skill premium in Norway over the sample period. The results are supported by exploiting a policy change to estimate the causal effect of innovation on relative skill demand.

*Imperial College Business School, CEPR and CEP. This paper has benefited from discussions with Erling Barth, Andrew Bernard, Swati Dhingra, Jason Garred, Keith Head, Jordi Jaumandreu, Réka Juhász, Julia Lane, Guy Michaels, Andreas Moxnes, Gianmarco Ottaviano, Steve Pischke, Thomas Sampson, Kjetil Storesletten, Catherine Thomas, Karen Helene Ulltveit-Moe and John van Reenen, as well as seminar and conference participants at LSE, University of Oslo, GEP/CEPR Annual Postgraduate Conference 2015, ETSG 2015, and NOITS 2015. This paper has received funding from Research Council of Norway (project number 224956), and is part of the research activities at the Centre for the Study of Equality, Social Organization, and Performance (ESOP) at the Department of Economics at the University of Oslo. Data received from Statistics Norway has been essential for the project and for the paper.

1 Introduction

Firms differ in their relative demand for skilled workers, even within narrowly defined sectors. What could be the driving force behind these differences? This paper focuses on the potential mechanism of skill-biased technological change.

The question of how technological change affects the demand for skills is a long-standing issue in economics. It arose from the observation that the skill premium soared in past decades even though the supply of skills increased substantially at the same time (Bound and Johnson, 1992; Card and DiNardo, 2002). Violante (2008) defines skill-biased technological change as "a shift in the production technology that favors skilled [...] labor over unskilled labor by increasing its relative productivity and, therefore, its relative demand".

Concurrent with the rise in the skill premium, globalization has increased dramatically, and has been shown to increase inequality both in theoretical (Helpman et al., 2010) and empirical (Goldberg and Pavcnik, 2007) work. If skilled workers are relatively less easily substitutable by imported inputs, or better at adjusting to new tasks or production processes introduced in the wake of offshoring, increased offshoring is expected to have similar observational effects on the skill premium as skill-biased technological change. In addition, research intensive firms are more likely to participate in international trade than other firms.[1]

The empirical literature on skill-biased technological change to date has mostly used aggregate (industry- or country-level) data. Using aggregate data is potentially problematic for two reasons. First, there is ample evidence showing a large amount of within-industry heterogeneity of firms.[2] Second, technology and globalization affect the aggregate demand for skill mostly through firm-level decisions. To obtain well-identified estimates, and to better understand the underlying mechanisms, focusing on the relationship between technology and skill demand at the level of the firm is important. Empirical work on productivity estimation and international trade employ firm-level data to a greater extent, but in most applications, firm productivity is usually assumed to be Hicks-neutral, and often exogenous.

In this paper, I model skill-biased technological change at the firm-level as a time-varying,

[1]See Yeaple (2005), Costantini and Melitz (2007) and Atkeson and Burstein (2011) for theoretical work on the innovation responses to an increase in market size. Aw et al. (2011) estimate a structural model of complementarities between innovation and exporting, while Bustos (2011) tests a theoretical model of exporting and technology upgrading. Lileeva and Trefler (2010) provides empirical evidence on the complementarities between exporting and technology. See Glass and Saggi (2001) and Rodriguez-Clare (2010) for theoretical work on the relationship between offshoring and innovation. Goel (2014) investigates the effects of offshoring on innovation, while Bøler et al. (2015) present both theoretical and empirical evidence on the complementarity between the two activities.

[2]See for instance Bernard and Jensen (1995) for evidence of within-sector heterogeneity.

unobservable productivity term, much as unobservable Hicks-neutral productivity has been modeled in the productivity estimation literature (Olley and Pakes, 1996; Levinsohn and Petrin, 2003; Ackerberg et al., 2006). It is, to the best of my knowledge, the first paper to estimate firm-level skill-biased productivity which is endogenously affected by research and development (R&D). To this end, I use rich Norwegian matched employer-employee data covering the period from 1997 to 2010. Importantly, the data contains firm-level information on R&D expenditures, input use, and revenue, as well as the number of years of education and wages for all workers in the firm. The panel structure of the data enables me to track both firms and workers over time, thereby allowing for a dynamic specification in which R&D investments and other firm decisions can potentially affect future skill demand.

Even with the right type of data, pinning down skill-biased technological change at the firm level is not trivial. First, technological differences between firms are not directly observable. Second, unobservable characteristics of the firm might drive both skill demand and investments in R&D or new technology. Hence, there are potential endogeneity issues posing challenges to identification. Finally, there might be a problem of reverse causality, if having a high share of skilled workers itself affects the likelihood of undertaking new investments in innovation or technology.

This paper tackles these questions in two ways. First, it provides reduced form evidence of the causal impact of innovation on firms' demand for skill, exploiting a policy change. Second, to quantify the extent to which firm productivity is skill-biased, I proceed to structural production function estimation. The estimation strategy exploits the timing of firm-level decisions to pin down the direction of causality. It also isolates the effect of innovation on skill demand from the effect of offshoring and other activities that might be complementary to innovation.

The paper has two sets of empirical results. The first part estimates the reduced form effect of innovation on firms' demand for skill. To break the correlation between unobserved productivity and innovation or technology, I follow Bøler et al. (2015) in exploiting a tax credit affecting the incentives to innovate for a subset of Norwegian firms. I first confirm that the reform had a large and positive effect on R&D investments of treated firms, as has been shown in previous studies.[3] Treated firms seem to have increased their R&D efforts on both the intensive and the extensive margins.

I then show that firms induced to start innovating also increase the skill demand, both in a difference-in-difference specification, and by exploiting the variation from the reform in an instrumental variables framework. Both of these reduced form techniques allow me to separate the effect of innovation on skill demand from the effect of being a larger or more productive firm. On average,

[3] See Hægeland and Møen (2007) and Bøler et al. (2015).

treated firms increased their skill demand by 4.6 percent, while the local average treatment effect of going from zero to positive R&D expenditures points to a 40 percent increase in skill demand.

While these reduced form results show that there is complementarity between innovation and skill demand, there are several channels through which innovation or new technology can increase the relative productivity of skilled workers. First of all, there is potentially a direct effect on workers who are directly involved in research activities within the firm. Second, skilled workers and innovation are likely to be complements in the production function, which is the channel I am focusing on in this paper. Third, Bøler et al. (2015) show that there is complementarity between R&D and imports, and importing might also increase the relative productivity of skilled workers, as mentioned above. While I can show that my reduced form results are not solely driven by firms increasing the demand for R&D workers, separating the effect of innovation from importing is hard in the reduced form framework since treated firms are likely to increase offshoring.

To be able to quantify the extent to which firm productivity is skill-biased, and to isolate the effects of innovation and offshoring on skill demand, I proceed to structural production function estimation. The structural estimates of skill-biased productivity also capture the fact that firms do not have to be directly involved in R&D for their workers to be affected by technological progress. Adaptation of production processes and use of intermediate inputs embedding new technologies also has the potential of changing the marginal productivities of various types of workers within the firm differentially.

The approach to separating skill-biased from Hicks-neutral technological change builds on recent advances in the literature on structural estimation of production functions. This literature aims to tackle the well-known endogeneity problem that arises when estimating production functions because a firm's decisions depend on its productivity, which is not observed by the econometrician. The paper exploits the idea developed by Olley and Pakes (1996) and Levinsohn and Petrin (2003), that the observed input decisions made by the firm can be used to infer its unobservable productivity. It further builds on Doraszelski and Jaumandreu (2014), who extend this methodology to a setting in which productivity is multi-dimensional.

The estimated specification is motivated by a theoretical framework of profit-maximizing firms. Firms have nested constant elasticity of substitution (CES) production functions, in which the aggregate labor input consists of a CES combination of skilled and unskilled workers. There is a productivity-augmenting term associated with skilled workers, reflecting the fact that there might be complementarities between skills and other choice variables of the firm, for instance technology and/or offshoring. Firms can endogenously affect this unobservable skill-augmenting productivity term by investing in R&D, or by offshoring.

Estimation proceeds in two steps. In the first, I exploit the ratio of first order conditions of the two types of labor, which effectively cancels out any Hicks-neutral (observed or unobserved) productivity terms. In the second step, I estimate a gross output production function following the recent critiques by Gandhi et al. (2013) related to the identification of production functions in the presence of variable inputs. I follow the approach suggested by Grieco et al. (2016), making use of the ratio of first-order conditions of the aggregate labor input and materials to deal with the issue of potential correlation between unobserved productivity and input usage.

I estimate a skill-biased component of firm productivity that turns out to be economically important. The within-firm growth of skill-biased productivity can account for the majority of the observed rise in the skill premium in Norway over the sample period. Consistent with the reduced form results, innovating firms have higher levels and growth rates of skill-biased technological change. The direct short-run (long-run) impact of own R&D investments on skill demand is roughly 3.5 percent (17.7 percent), while offshoring does not have a significant direct effect. The results are not driven by firms hiring more R&D workers. The results are not dependent on the particular functional form assumptions, as I find complementary results from an alternative specification.[4]

The contributions of the paper are threefold. First, it provides a simple framework to estimate skill-biased productivity using firm-level data. This framework builds on standard estimation techniques and a simple model of firm behavior, and can easily be adapted to focus on other firm decisions.[5] Second, the paper clarifies one channel through which innovation affects inequality – namely that increased innovation leads to increased demand for skill. For policy, this implies that government efforts to increase R&D can have distributional effects. Third, I show that the distribution of estimates of Hicks-neutral productivity depend on whether the skill-biased component of productivity is taken into account or not. This finding will be of particular interest when attempting to evaluate the effect of innovation or technology adoption on firm productivity.

The paper contributes to several strands of literature. First, the paper is related to the literature looking at the effects of technological change on productivity, skill demand or skill premium at the plant or firm level. A typical problem in this literature is the fact that the subset of firms engaging in innovation is not randomly selected, and there are likely unobserved factors affecting both the choice of technology and skill demand.[6] This paper contributes to that literature by using a natural

[4]In the appendix, I estimate Cobb-Douglas production functions in which the input coefficients are allowed to vary with the R&D status of the firm. The results of both exercises point in the same direction, demonstrating that innovation does not affect all inputs symmetrically.

[5]For instance, one can look at whether for instance innovation, importing or exporting affects the relative demand for other groups of workers, like workers performing routine vs. non-routine tasks, in the spirit of Autor et al. (2003).

[6]Doms et al. (1997), Bresnahan et al. (2002) and Abowd et al. (2007) all find that different measures of technology and skill demand are strongly correlated at the plant- or firm-level, but cannot properly deal with unobserved

experiment to estimate the effect of innovation on skill demand, and by providing a structural estimation framework that can be employed in the absence of such exogenous variation.[7]

In the productivity estimation literature that takes this endogeneity problem very seriously, technological change is usually assumed to be Hicks neutral.[8] A notable early exception is Van Biesebroeck (2003), where technological change is assumed to take two different forms, capital-augmenting and Hicks-neutral. Two more recent examples are working papers by Doraszelski and Jaumandreu (2014) and Raval (2015), who look at the labor- and capital-biases of technological change, respectively. In addition, Zhang (2015) estimates a multi-dimensional productivity measure, which allows for capital-augmenting, labor-augmenting, and material-augmenting efficiencies. I draw on the estimation techniques employed in this literature, and contribute by focusing on the potential skill-bias of technological change. Productivity is usually assumed to evolve exogenously in this strand of the literature. I build on Doraszelski and Jaumandreu (2013) by allowing skill-augmenting productivity to be endogenously affected by innovation.

The paper also relates to a growing literature looking at how much of overall inequality can be contributed to within-firm versus between-firm changes. Song et al. (2015) find that the vast majority of overall inequality is driven by increasing dispersion between, not within, firms, using data from the US. Somewhat contrary to their findings, my results suggest that a big part of the increase in the skill premium can be explained by within-firm changes in skill demand. Mueller et al. (2015) find that there is a positive correlation between the within-firm skill premium and firm size, using a panel of UK firms. These findings are consistent with mine, since there is a positive correlation between firm size and investments in innovation and technology.

Finally, the paper relates to the growing literature on the relationship between international trade, technology and the skill premium. There are several papers proposing and calibrating models to pin down the effects of trade and technology on the skill premium.[9] In contrast to these papers, I focus on the effects of within-firm skill-biased technological change and offshoring on

heterogeneity.

[7]Similar in spirit to my paper is Akerman et al. (2015), who estimate production functions in which the input coefficients are allowed to vary with broadband adoption in the firm. Their results indicate that there is complementarity between skills and information technology. Also related are other papers looking specifically at the link between computers and skill demand, either at the firm (Levy and Murnane, 1996) or occupation (Autor et al., 2003) level.

[8]See for instance Griliches (1979), Olley and Pakes (1996), Levinsohn and Petrin (2003), Ackerberg et al. (2006), and Doraszelski and Jaumandreu (2013).

[9]See Acemoglu (2003) and Monte (2011) for theoretical models. Harrigan and Reshef (2015) and Burstein and Vogel (2010) introduce complementarity between skills and productivity in a way that is similar to mine, and calibrate a value for this parameter using the cross-sectional relationship between skill demand and sales. I take a very different approach to estimating skill-augmenting productivity, allowing it be time-varying and endogenously affected by firm's innovation and import activities.

skill demand, not on the between-firm reallocation of inputs in response to trade liberalization.[10] Also, the papers by Kasahara et al. (2013) and Akhmetova and Ferguson (2015) employ similar estimation techniques, looking at the effect of offshoring on skill upgrading and skill-biased technological change.

The paper is organized as follows. In the next section, I provide details about the data, descriptive statistics and sample selection. Section 3 describes the details of reform, the reduced form empirical strategies and results. The theoretical framework and structural estimation strategy is explained in Section 4. In Section 5, I describe and discuss the main results, showing that technological change is skill-biased at the firm level, and that the within-firm change in this term can explain most of the observed increase in the skill premium over my sample period. The last section concludes.

2 Data and Descriptive Statistics

2.1 Data

The data employed in this paper combines several data sets, all from Statistics Norway. The first is administrative firm-level balance sheet data from the *Capital Database*. This is an annual unbalanced panel of all non-oil manufacturing joint-stock firms.[11] It includes approximately 6000 firms per year, covering around 90 percent of all manufacturing firms. The data set contains standard balance sheet information on for instance revenue, value added, wage bill, number of employees, capital, investments, input use, and the main industry of operation.[12] All variables are deflated using industry-specific price indices. The data covers the years 1997 to 2010.

The second data set is worker-level data from the employee register, which matches employer information with tax records to create a verified individual-level data set of wages and employment categories (full time or part time), among other variables, of all employees in Norway. This data set is merged with demographic data containing information about labor market experience, years and

[10]The paper is also related to the literature looking at the effects of trade and technology on the industry-level skill premium. Feenstra and Hanson (1999) estimate the influence of trade and technology on the relative wage of non-production workers, using data from the US. They find that technology is the most important of the two factors. More recently, Goel (2014) finds that offshoring increases the skill premium, even if it increases the wages of both skilled and unskilled workers, also using industry-level data from the US. Interestingly, she finds that offshoring affects the US workers through inducing innovation, and not so much through substitution of unskilled workers.

[11]The Capital Database is described in Raknerud et al. (2004).

[12]The industry classification used is the NACE Rev. 2. A list of industries can be found in `http://ec.europa.eu/eurostat/ramon/nomenclatures/index.cfm?TargetUrl=LST_NOM_DTL&StrNom=NACE_REV2`.

field of education, gender, municipality of residence and work, and other characteristics by person-year. I classify workers as skilled or unskilled based on the number of years of education. Skilled workers refer to workers with at least 14 years of education, which is equivalent to having some college education, while unskilled workers have less. This data provides information on hourly wages of each worker, which is used to construct firm-level measures of skilled and unskilled wages. To avoid confounding penalties to part-time work with determinants of the skill premium, I include only full time workers in my sample.

The third data set is R&D data, which comes from Statistics Norway's R&D survey. This provides detailed information on firm-level R&D investments and R&D personnel for a subset of the firms in the manufacturing sector. All firms with more than 50 employees are included in the survey each year it is conducted. In addition, a sample of the smaller firms are drawn each year. If firms that are sampled have more than NOK 1 million in R&D expenditures, they are automatically included in next period's sample. Since this sampling procedure introduces some selection issues, I limit my sample to firms above the 50 employee threshold. The R&D data is biennial for the period from 1997 to 2003, and annual from 2003 and onward.

Finally, trade data provides information about firm-level export and import values based on customs declarations. These data make up an unbalanced panel of all yearly export and import values by 8-digit HS product codes and a destination or source country. By aggregating this trade data, I obtain export and import values for each firm-year observation.

All data sets can be merged based on a unique firm identifier. I keep only firms that I can observe both before and after the reform was introduced, and that are included in the R&D data. I am left with a sample of 925000 worker-year observations, based on roughly 4-500 firms each year, spread across 23 industries. The final data covers 1997 to 2010. More details on the data, sample and variables can be found in Section D in the Appendix.

2.2 Trends in Norway

Norway provides a convenient setting for this exercise. Although the Norwegian labor market is characterized by coordinated wage setting and high wage compression, it is a flexible labor market with substantial churning. Income has increased more at the top of the wage distribution than at the bottom also in Norway. OECD (2011) reports that average annual real household income growth between 1986 and 2008 was twice as high for the top decile as for the bottom decile (1.4 versus 2.7 percent).

Inequality can rise for several reasons that are unrelated to skill-biased technological change.

However, also the skill premium has been rising in Norway over the sample period, despite an increasing relative supply of skilled workers. Figure 1 shows the trends in relative supply of skilled workers and the skill premium in the Norwegian manufacturing sector in the period between 1997 and 2008. It shows that the relative supply of skilled workers has increased by a massive 40 percent over the sample period, while at the same time, the skill premium has increased by 11 percent. This points to a substantial increase in the relative productivity of skilled workers.

2.3 Descriptive Statistics

R&D firms tend to be different from other firms along several margins.[13] Table 1 shows the results of various outcome variables regressed on an R&D dummy, controlling for industry and year fixed effects. They show that firms doing R&D are larger (column (1)), have a higher share of college educated workers (column (2)), have workers with more years of education (column (3)), pay higher wages (column (4)), have higher labor productivity (column (5)), and pay a higher skill premium (column (6)).[14] These results highlight the importance of taking unobserved firm productivity into account, as simply comparing the input coefficients of firms that invest in R&D and firms that do not is likely to confound the effect of being a large or productive firm with the effect to doing R&D.

As is evident from Figure 1, the share of skilled workers increased over my sample period. To better understand what is underlying this change, I decompose it into its within- and between-firm components. Denoting the aggregate share of workers of skill group g across firms in period t by S_{gt}, it can be written as

$$S_{gt} = \sum_f \lambda_{ft} s_{fgt},$$

where λ_{ft} is firm f's share of total employment, and s_{fgt} is firm f's share of workers in skill group g. The aggregate change in S_{gt} can be decomposed as follows:

$$\Delta S_{gt} = \underbrace{\sum_f \Delta\lambda_{ft}\bar{s}_{fg}}_{\text{between}} + \underbrace{\sum_f \bar{\lambda}_f \Delta s_{fgt}}_{\text{within}}, \tag{1}$$

where overbars denote time averages. The first term captures changes in the overall share of workers in skill group g due to changes in employment shares of firms with different skill shares. Entry

[13] I define an R&D firm, or innovating firm, as a firm with positive R&D expenditures. In 2001, a little less than half the firms in my sample had positive R&D expenditures. This is consistent with aggregate numbers from Statistics Norway for manufacturing and mining, showing that 44 percent of firms with more than 50 employees had positive R&D expenditures in 2001.

[14] Labor productivity is defined as value added per worker.

and exit of firms is captured by the employment share of a firm λ_{ft} either going from zero to positive or vice versa. The second term captures the change in the overall share of workers in skill group g due to within-firm changes in the skill shares s_{fgt}.

I report results from this exercise for the sample period 1997-2010 in Table 2. Overall, the share of skilled workers in my sample rose from around 17 percent to 24 percent. For the skilled workers, this change was driven by roughly equal parts within- and between-firm changes. For the unskilled workers, on the other hand, the within-firm changes contributed much more to the overall decline in the employment share, while the between-firm changes counteracted the negative trend somewhat. Hence, it is not the case that the share of unskilled workers is declining because the particularly skill-intensive firms are growing. On the contrary, these descriptive findings are consistent with changes in the relative productivities of skilled and unskilled workers resulting in increased within-firm demand for skills.

3 Reduced Form Findings

In this section, the reform details are described, along with the empirical strategy and results from the reduced form exercises. The first results in this section show that the treated firms indeed increased their innovation efforts along both the extensive and intensive margins, which is crucial for using the reform as a source of exogenous variation in R&D to estimate the effects of innovation on skill demand.[15] The second set of results show the causal effect of R&D on the share of skilled workers. This effect is shown to hold even when disregarding the contribution to the college share of the workers directly involved in the research activities.

3.1 The R&D Reform

The reform, called *Skattefunn*, was put into effect in 2002 as a rights-based tax credit, that allows firms to deduct 20 percent of their R&D expenditures up to a threshold of 4 million NOK.[16] There are several reasons why this reform is useful for my purposes. First of all, projects have to be pre-approved by the Research Council of Norway, so firms can not simply relabel some other expenditures as R&D. An R&D proposal must be submitted and approved before firms receive the

[15] The empirical strategy for the difference-in-difference specifications is borrowed from Bøler et al. (2015), and some of the results with continuous R&D as an outcome variable are similar to the results in that paper.

[16] Initially, only small and medium sized enterprises (SMEs) were eligible for the tax cut, but large enterprises (with more than 100 employees) were included already in 2003. Large enterprises are treated slightly different from SMEs, as they receive a 18 percent tax reduction.

credit. To qualify, a project must be limited and focused, aimed at generating new knowledge, information or experience, to be of use in developing new or improved products, services or manufacturing/processing methods. Second, the details of the reform were only announced a few months before it was introduced, limiting the scope for anticipation effects. Third, the reform was not part of a larger tax reform, nor initiated as a response to any macroeconomic shock. Finally, for firms that are not in a taxable position, the appropriate amount is paid out as a grant. This implies that not only particularly well-performing firms are eligible.

A worry when estimating the effects of a reform such as Skattefunn is that the administrative effort required to receive the tax credit introduces selection bias, in the sense that only the firms with the best projects and most resources will spend time and money on the application. There are two main reasons why this should not be the case for this particular reform. First of all, it is a rights-based tax credit. That implies that all firms fulfilling the criteria for projects will receive the credit, as long as they send in the formal application. Second, Cappelen et al. (2010) estimate the average costs related to applying for and receiving Skattefunn to be modest, at roughly 4 percent of the total tax credit paid out. More details on the reform can be found in Section C in the Appendix.

3.2 Reduced Form Framework

Even though all firms can apply for the tax deduction, the reform has a built in discontinuity in the form of a cut-off for when a firm's marginal cost of R&D is affected. The firms are entitled to tax deductions for R&D expenses up to 4 million NOK, roughly equivalent to 0.5 million USD. Firms with lower levels of initial R&D spending will experience a 20 percent reduction in the cost on their marginal R&D investment, and hence might be induced to increase their R&D spending.

This feature of the reform can be exploited in a difference-in-differences (DID) type of specification, where I compare the differences in various outcomes between the pre- and post-reform periods for the group of firms that experienced decline in the marginal cost of R&D to the firms that did not. Following Bøler et al. (2015), a firm is classified as "treated" if its average pre-reform spending is less than 4 million NOK, and as a "control" firm if it exceeds 4 million NOK. Figure 2 illustrates the variation picked up by the DID. The dotted line represents the average R&D expenditure for the treated firms, while the solid line shows average for the control firms. The trends prior to the reform seem fairly similar, but after the introduction of the reform in 2002, there is a spike in the R&D activities for the treated group.

A standard DID specification is given by:

$$y_{jt} = \alpha_j + \beta \left(T_{1j} \times \delta_t\right) + \delta_t + \varepsilon_{jt}, \tag{2}$$

where T_{1j} is a treatment indicator taking the value one if the firm is classified as treated, and zero otherwise. α_j is a firm fixed effect, δ_t is year fixed effects, and β is the vector of coefficients on the interactions between the year dummies and the treatment indicator. If the pre-reform trends are similar, the βs should be small and insignificant before the introduction of the reform. If the reform affected the R&D investments of the two types of firms differentially, the βs should be positive and statistically significant in the years after the reform.

An extension of this framework allows for the inclusion of firm-specific trends. This permits treatment and control firms to have different trends, and hence should reduce concerns about reversion to the mean or other forms of spurious correlation between the group of treated firms and the outcome variable.

$$y_{jt} = \alpha_j + \delta_t + g_j t + \beta \left(T_{1j} \times Post_t \right) + \varepsilon_{jt}, \tag{3}$$

where $Post_t$ is a dummy that equals one in the post-reform years, and zero otherwise, and g_j is the firm-specific coefficient on the trend, t. This framework allows the treatment term $(T_{1j} \times Post_t)$ to be arbitrarily correlated with either the firm fixed effect or the firm-specific trend. By differencing this I obtain a triple differences model that is estimated by fixed effects:

$$\Delta y_{jt} = \Delta\delta_t + g_j + \beta\Delta \left(T_{1j} \times Post_t \right) + \Delta\varepsilon_{jt}. \tag{4}$$

In my baseline specification, the outcome variable is log R&D expenditures, since the distribution of R&D is highly skewed. Using logs also eases the interpretation of the coefficients. To check the robustness of the baseline result, I can replace the outcome variable with the level of R&D, or alternatively the R&D intensity, defined as R&D investments over sales. Since these variables are non-negative (and zero for some observations), I use a fixed effects Poisson quasi-maximum likelihood (FE-PQML) estimator following Santos Silva and Tenreyro (2006). The conditional expectation function is given by:

$$E \left[RD_{jt} \right] = exp \left[\alpha_j + \delta_t + g \left(t \times T_{1i} \right) + \eta \left(T_{1j} \times Post_t \right) \right]. \tag{5}$$

where y_{jt} is the level of R&D expenditures or R&D intensity for firm i at time t.[17] Since this framework cannot handle negative values, it is not possible to difference this equation. Group-specific trends $g \left(t \times T_{1i} \right)$ are therefore included to deal with potential differential trends for the two groups.

[17]The distribution of y_{jt} does not have to be discrete for this estimator to provide consistent estimates. See Wooldridge (1999).

Later, I will exploit the variation from the reform in an instrumental variable framework in which the instrumented variable is a dummy variable for R&D investments. Recall that the details of the reform does not allow me to estimate the extensive margin of R&D directly. Since all the firms in the control group by definition have positive R&D expenditures, there is no variation in the extensive margin for these firms. However, looking at other measures of R&D also provides an indirect way of evaluating whether the reform also induced some treated firms to start investing in R&D. I can compare the coefficients of a regression of where the outcome variable is the level of R&D expenditures for firms doing R&D in both periods, to one which includes also the firms with zero R&D. If the coefficient in the latter case is greater in magnitude compared to the first, this indicates that some treated firms that were previously not investing in R&D started doing so, or that some control firms stopped investing in R&D after the reform. The latter case is a rare event, as there is a lot of persistence in R&D within firms over time. The results presented below are, reassuringly, robust to dropping the few observations where control firms stop investing in R&D post-reform.[18]

3.3 Reduced form effects on R&D expenditures

The results of running regressions (2) and (6) are given in columns (1) and (2) of Table 3, respectively. Focusing on column (1) first, the coefficients are, as expected, small and insignificant before the introduction of the reform, and large, positive and highly statistically significant in the years after. They show that treated firms indeed did increase their R&D expenditures. Column (2) show the results from the triple difference specification. The coefficient of 0.433 is statistically significant and economically substantial: the reform led to an increase in R&D investments of roughly 50 percent.[19]

Table 4 provides several robustness checks to verify the magnitude and relevance of the estimated effects.[20] The results from the triple difference specification in column (2) of Table 3 are reproduced in column (1) for comparison. In column (2), the dependent variable is log R&D intensity. Again, the estimated effect is economically substantial and statistically significant. So far, only the effects on the intensive margin have been estimated, since the outcome variable is in logs. In columns (3) to (5), the FE-PQML specification is employed. Results from an FE-PQML specification on the same (intensive margin) sample as in column (1) are provided in column (3), and show that the results are comparable in size. The estimated effect decreases somewhat, to 43

[18]Detailed results available upon request.

[19]$100 * (\exp(0.433) - 1) = 54.2$ percent.

[20]For more robustness checks on the effectiveness of the reform, see Bøler et al. (2015).

percent.

When the sample is increased to include the zeros (column (4)), the effect more than doubles compared to the results in column (3). This reflects the fact that some treated firms with no pre-reform R&D expenditure started investing in R&D, pulling up the average effect of the reform. This is consistent with results from Hægeland and Møen (2007), who find that firms are more likely to start investing in R&D after the reform was introduced. The estimated effect of the reform on the extensive and intensive margins combined is large, implying more than a doubling of R&D expenditures for the average treated firm. Column (5) presents the results using R&D intensity as the outcome variable. Again, the estimated effect is positive, significant and substantial – treated firms increased their R&D intensity by 170 percent.[21] The fact that the estimated percentage change in R&D intensity is much larger than for R&D expenditures again is likely to reflect the fact that many firms with no (or very small amounts of) pre-reform R&D expenditures started investing (or increased their investments) in R&D, which for these firms entailed a massive increase in R&D intensity.

3.4 Reduced form effects on skill demand

In this paper, I am interested in whether innovation affects the skill demand of firms. If technological change is skill-biased, we would expect firms that are induced to start performing R&D to increase their relative demand for skilled workers, as these workers become relatively more productive. In Figure 3, I plot the evolution of the college shares of the treated and control firms. Also in this case, the pre-reform trends seem quite similar, but the college share for the treated firms show a much more rapid growth after 2002 than what is the case for the control firms. This is consistent with increased R&D at the firm level driving up the relative demand for skilled workers. I estimate the reduced form effects of the reform on skill demand in two ways. First, I proceed with the framework presented above. Second, I exploit the variation from the reform as an instrumental variable.

3.4.1 Skill demand: DID

I begin by running a specification that is very similar to equation (6) above. In that specification, differencing implies that the only variation picked up by the coefficient on the treatment term $(T_{1j} \times Post_t)$ comes from the first year after the introduction of the reform. This is not ideal when the outcome variable is likely to adjust with some lag, as is the case when the outcome variable is

[21]The median R&D intensity across years in the sample is 0.002. Conditional on having positive R&D, it is 0.01.

related to the composition of the workforce. Instead, I estimate a model of the following kind:

$$\Delta y_{jt} = \Delta\delta_t + g_j + \beta\left(T_{1j} \times Post_t\right) + \Delta\varepsilon_{jt}, \tag{6}$$

which again is estimated by fixed effects. The outcome variable is a measure of skill demand at the firm level. Hence, the coefficient on the treatment term, β picks up differential changes in skill demand for the treated firms in the years after the reform. Results are presented in Table 5. In column (1), the outcome variable is the college share. Since this college share variable is highly skewed, in column (2) I use the log of the college share.

Again, results confirm the positive effect of R&D on skill demand. Lastly, in column (3), I use the log skill ratio as the dependent variable. This is the variable used in the structural estimation in Section 4, hence the coefficient in column (3) provides a reduced form estimate that I can compare to the estimated effect of R&D on skill demand from the structural framework. This preferred estimate suggests that R&D increases skill demand by roughly 5 percent. I expect this reduced form effect to be somewhat larger in magnitude compared to the structural estimate, as the reduced form effect incorporates other complementarities and spillovers associated with increased R&D that will affect skill demand, like for instance increased offshoring. The structural estimation framework is designed precisely to rule out these other competing channels, and isolate the direct effect of the firm's own R&D investments.

To check the robustness of these results, I can employ a placebo test. Instead of using the actual threshold, I classify firms as treated or control firms based on a counterfactual threshold of NOK 8 million. The results from this placebo can be found in Table 11 in the Appendix. Reassuringly, they show that there is no effect on skill demand in this counterfactual setting.

3.4.2 Skill demand: instrumental variables

The group of treated firms include many firms that do not take up the reform, and estimating the change in skill demand in the DID framework presented above yields somewhat imprecise estimates. To examine the extent to which R&D affects skill demand, I can also exploit the exogenous variation in an instrumental variables framework. The baseline instrumental variables specification takes the form:

$$\ln \text{Skill ratio}_{jt} = \alpha_j + \rho RD_{jt-1} + \delta_t + \varepsilon_{jt}, \tag{7}$$

where $\ln \text{Skill ratio}_{jt}$ is the log skill ratio firm j at time t, and α_j is again a firm fixed effect. Since employment is likely to adjust with a lag, and to better mirror the results from the structural estimation, RD_{jt-1} is a dummy taking the value one if the firm has positive R&D expenditure in the previous year, and zero otherwise. This R&D dummy is instrumented using the variation introduced by the reform. The instrumental variable technique allows me to separate the effect of innovation on skill demand from the effect of being a larger or more productive firm. The first stage is given by:

$$RD_{jt} = \gamma_j + \kappa \left(T_{1j} \times Post_t \right) + \tau_t + \varepsilon_{jt}, \tag{8}$$

where γ_j is a firm fixed effect, τ_t is a time dummy, and $(T_{1j} \times Post_t)$ is the same treatment term used in the previous section (i.e. the treated dummy interacted with the post-reform dummy).

Results are given in Table 6. I begin by presenting results from an OLS specification in column (1), dropping the firm fixed effect γ_j and not exploiting the instrument. The resulting coefficient of 0.404 is highly statistically significant, implying that there is a positive association between investing in R&D and the share of college educated workers in the firm. However, based on these results there is no way of telling if the larger college share is caused by the technological change, or if it is just the case that more productive firms are investing in R&D and also happen to have higher college shares than other firms.

In column (2), I include γ_j to account for time-invariant firm fixed effects. The associated increase in college share remains positive, but the size decreases. In this specification, only within-firm changes in innovation activities are estimated, but the potential simultaneity problem where changes in unobserved firm productivity drive both the college share and innovation remains unsolved. In addition, this coefficient is particularly susceptible to measurement error. If there is some degree of measurement error in the R&D variable, this will be exacerbated in the fixed effects specifications. The instrumental variables technique alleviates concerns related to this.

In column (3) I combine the instrumental variables technique with firm fixed effects (FE-IV). In the first stage, I regress the R&D dummy on the treated firm dummy as given by equation (8). The first stage coefficient (κ) is 0.171 and statistically significant at the 1 percent level. The first stage is reasonably strong, with an F statistic of 35.34. The second stage coefficient (ρ) is 0.334. Comparing this to the OLS result of 0.404 indicates that the endogeneity problem described above implies an upward bias in the OLS coefficient, as expected. The local average treatment effect of the instrumental variables approach suggests that starting to invest in R&D leads to a 40 percent increase in skill demand.

While the instrumental variables technique is helpful for pinning down the direction of causality, the observed increase in the college share could be driven by treated firms hiring research personnel as they start performing R&D. While this might partly explain the observed positive correlation between skill demand and proxies for technology, the skill-biased technological change hypothesis speaks to a more general shift in the production function, where technology is complementary to tasks performed by skilled workers, and less so to tasks performed by unskilled workers. If this is indeed what is underlying these results, they should be robust to removing the workers directly involved in R&D.

To evaluate the contribution of this "mechanical" channel relative to the skill-biased technological change channel, I assume that all workers denoted as R&D workers in the firm level data are in fact college educated workers. I remove the number of R&D workers from the number of college educated workers, and re-estimate the college share regression.[22] The results from this exercise is presented in column (4) of Table 12 in the Appendix, and show that the college share effect is even stronger when all R&D workers are taken out. This result might seem counter intuitive, but it likely reflects that when firms start engaging in R&D, this increases the productivity of the skilled workers, and that this effect is much stronger than the direct effect of hiring skilled workers to perform the actual R&D. This should relieve the worry that the observed increase in the college share is driven only by research personnel.

3.5 Other firm level outcomes

In addition to examining the outcomes for R&D and skill demand, I can check whether other firm outcomes seem to respond to the reform. In Table 13 in the Appendix, I check whether output, capital or the overall size of firms change as a result of increased R&D, in a triple difference model as the one given in equation (6). As is evident from the table, there seems to be no significant change in either of these variables.

4 Estimating Skill-Biased Productivity

The results from the reduced form exercises show that treated firms increased their skill demand, and that for the subset of treated firms that responded to the reform, this increase was substantial in magnitude. This indicates that there is complementarity between skills and R&D, beyond the direct effect of hiring more workers to do the actual research activities. This direct impact of R&D on skill demand is one contributor to skill-biased technological change.

[22] The data does not specify the actual education level of R&D workers, and it might be the case that not all of them are college educated. However, removing more than the true amount of workers from the number of college educated workers would only bias the results downward, unless there is a systematic difference in the skill level of R&D workers between treated and control firms.

However, there are sources to technological differences beyond the firm's own research activities that will determine differences in skill demand across firms. There might be spillovers from innovation across firms, differences in adoption rates, and so forth. Also, the extent to which offshoring, capital intensity and other firm characteristics matter for within-firm increases in relative skill demand are not easily evaluated in the reduced form framework. To this end, I employ an estimation technique based on tools and concepts commonly used in the productivity estimation literature. In this section, I describe the theoretical framework and strategy for this structural estimation.

4.1 Theoretical Framework

In this section, I lay out a dynamic model of firm behavior to guide the empirical analysis. The theoretical framework builds on Doraszelski and Jaumandreu (2014), which is extended to allow for two different skill groups.

The firm is producing a single output Y_{jt} and operates in monopolistically competitive markets. Assume the firm has a nested CES production function of the form:

$$Y_{jt} = \beta_0 \left[\beta_K K_{jt}^{\frac{\sigma_Y - 1}{\sigma_Y}} + \beta_L (L^*_{jt})^{\frac{\sigma_Y - 1}{\sigma_Y}}{}_{jt} + \beta_M M_{jt}^{\frac{\sigma_Y - 1}{\sigma_Y}} \right]^{\frac{\sigma_Y}{\sigma_Y - 1}} e^{\omega_{Njt}} e^{\epsilon_{jt}}, \tag{9}$$

where Y_{jt} is gross output, K_{jt} is capital, M_{jt} is intermediate inputs, and L^*_{jt} denotes an aggregate of skilled and unskilled labor. Without loss of generalization, in what follows I set $\beta_0 = 1$.[23] ω_{Njt} is a Hicks-neutral productivity term, and ϵ_{jt} is an mean zero random shock to production that is uncorrelated both over time and across firms. σ_Y is the elasticity of substitution of the aggregate production function. Depending on this parameter, the production function nests three particular focal cases: Leontieff ($\sigma_Y \to 0$), Cobb-Douglas ($\sigma_Y = 1$) and linear ($\sigma_Y \to \infty$).

The labor inputs are combined using another CES function:

$$L^*_{jt} = [(e^{\omega_{Sjt}} L_{Sjt})^{\frac{\sigma_L - 1}{\sigma_L}} + L_{Ujt}^{\frac{\sigma_L - 1}{\sigma_L}}]^{\frac{\sigma_L}{\sigma_L - 1}}, \tag{10}$$

where L_{Sjt} denotes skilled labor, and L_{Ujt} is unskilled labor. σ_L is the elasticity of substitution between the two types of labor input, which might differ from the elasticity of substitution between the aggregate inputs, σ_Y. Again, the labor inputs can be combined in fixed proportions ($\sigma_L \to 0$),

[23] This is done since the constant term in the production function cannot be identified separately from the constant term in the Markov process for Hicks-neutral productivity.

Cobb-Douglas ($\sigma_L = 1$), or linearly ($\sigma_L \to \infty$). Which of these forms the production function takes will determine whether skill-augmenting technological change is skill-biased or unskill-biased. Finally, ω_{Sjt} is a skilled-augmenting productivity term.[24] This skill-augmenting productivity term evolves endogenously according to a controlled first-order Markov process, detailed below.[25]

Firms operate in monopolistically competitive markets and face an inverse demand function of the standard Dixit-Stiglitz form:

$$P_{jt} = P_{st}(Y_{st}; \eta) = P_{st}\left(\frac{Y_{jt}}{Y_{st}}\right)^{\frac{1}{\eta}},$$

where P_{st} and Y_{st} are industry-level output quantity and price, and $\eta < -1$ is the elasticity of residual demand. Capital accumulates according to $K_{jt+1} = (1-\delta)K_{jt} + I_{jt}$, where I_{jt} is investment and δ is the rate of depreciation of capital. As is standard in this literature, I assume that investments in capital take one period to become productive.

The firm maximizes the expected net present value of future profits. The Bellman equation for its dynamic programming problem is given by

$$\begin{gathered} V_t(\Omega_{jt}) = \max_{K,L_S,L_U,M,R} P_{jt}(X_{jt}^{\frac{\sigma_Y}{\sigma_Y - 1}} e^{\omega_{Njt}}, \eta) X_{jt}^{\frac{\sigma_Y}{\sigma_Y - 1}} e^{\omega_{Njt}} \mu \\ -C_I(K_{jt+1} - (1-\delta)K_{jt}) - w_{Sjt}L_{Sjt} - w_{Ujt}L_{Ujt} \\ -P_{Mjt}M_{jt} - C_R(R_{jt}) + \tfrac{1}{1+\rho}\mathbb{E}_t(V_{t+1}(\Omega_{jt+1})|(\Omega_{jt}), R_{jt}), \end{gathered}$$

where $X_{jt} = \beta_K K_{jt}^{\frac{\sigma_Y - 1}{\sigma_Y}} + \beta_L (L^*_{jt})^{\frac{\sigma_Y - 1}{\sigma_Y}}{}_{jt} + \beta_M (M^*_{jt})^{\frac{\sigma_Y - 1}{\sigma_Y}}$ and $\mu = \mathbb{E}_t[e^{\epsilon_{jt}}]$. ρ is the discount rate, and $\Omega_{jt} = (K_{jt}, \omega_{Sjt}, \omega_{Njt}, w_{Sjt}, w_{Ujt}, P_{Mjt})$ is the vector of state variables. w_{Sjt} and w_{Ujt} are the wages of skilled and unskilled workers, respectively, and P_{Mjt} is the price of materials. C_I and C_R are the costs of investment and R&D, respectively. I follow Levinsohn and Petrin (2003) and Doraszelski and Jaumandreu (2013) in assuming that both types of labor inputs and materials are flexibly adjusted to maximize short-run profits.[26]

[24]Alternatively, the function could have been written as

$$L^*_{jt} = [\beta_{LS} L_{Sjt}^{\frac{\sigma_L - 1}{\sigma_L}} + \beta_{LU} L_{Ujt}^{\frac{\sigma_L - 1}{\sigma_L}}]^{\frac{\sigma_L}{\sigma_L - 1}}.$$

In this case, technological change can be seen as altering these parameters of the production function instead of changing the efficiency of one of the factors. To the extent that there are inter-industry differences in the factor shares β_{LS} and β_{LU}, these will be absorbed by industry fixed effects in the empirical specification.

[25]Hicks-neutral productivity ω_{Njt} can also be modeled as following a separate controlled first-order Markov process, following Doraszelski and Jaumandreu (2014). Since the focus of this current paper is not on estimating the dynamics of this productivity process, ω_{Njt} is simply backed out using the estimated parameters.

[26]The Cobb-Douglas estimation approach described in the Appendix allows the two types of labor to be subject

The first order condition of the aggregate labor input is

$$\mu\beta_L X_{jt}^{\frac{\sigma_Y}{\sigma_Y-1}} e^{\omega_{Njt}} (L_{jt}^*)^{-\frac{1}{\sigma_Y}} = \frac{w_{jt}}{P_{jt}(1-\frac{1}{|\eta|})}, \tag{11}$$

where w_{jt} is the average firm wage, and $|\eta|$ is the absolute value of the elasticity of residual demand faced by the firm as described above.

The first order conditions for materials is

$$\mu\beta_M X_{jt}^{\frac{\sigma_Y}{\sigma_Y-1}} e^{\omega_{Njt}} (M_{jt})^{-\frac{1}{\sigma_Y}} = \frac{P_{Mjt}}{P_{jt}(1-\frac{1}{|\eta|})}. \tag{12}$$

The ratio of these first order conditions of materials and aggregate labor is

$$\frac{\beta_L}{\beta_M}\left(\frac{L_{jt}^*}{M_{jt}}\right)^{-\frac{1}{\sigma_Y}} = \frac{w_{jt}}{P_{Mjt}}.$$

Multiplying both sides by $\frac{L_{jt}^*}{M_{jt}}$ gives the following relationship between expenditure shares and input shares

$$\frac{\beta_L}{\beta_M}\left(\frac{L_{jt}^*}{M_{jt}}\right)^{\frac{\sigma_Y-1}{\sigma_Y}} = \frac{E_{L_{jt}}}{E_{M_{jt}}}, \tag{13}$$

where $E_{L_{jt}^*} = w_{jt}L_{jt}^*$ and $E_{M_{jt}} = P_{Mjt}M_{jt}$ are the expenditure on materials and labor, respectively. This ratio will be crucial for disentangling input prices and input quantities from the materials bill, as explained below in Section 4.2.

If the labor aggregate L_{jt}^* is inserted into the overall production function, it it straightforward to calculate the first order conditions for skilled and unskilled labor. They are given by:

$$\mu X_{jt}^{\frac{\sigma_Y}{\sigma_Y-1}} e^{\omega_{Njt}} (L_{jt}^*)^{-\frac{1}{\sigma_Y}} \frac{\partial L_{jt}^*}{\partial L_{Sjt}} = \frac{w_{Sjt}}{P_{jt}(1-\frac{1}{|\eta|})},$$

and

$$\mu X_{jt}^{\frac{\sigma_Y}{\sigma_Y-1}} e^{\omega_{Njt}} (L_{jt}^*)^{-\frac{1}{\sigma_Y}} \frac{\partial L_{jt}^*}{\partial L_{Ujt}} = \frac{w_{Ujt}}{P_{jt}(1-\frac{1}{|\eta|})}.$$

Solving for $\frac{\partial L_{jt}^*}{\partial L_{Sjt}}$ and $\frac{\partial L_{jt}^*}{\partial L_{Ujt}}$ and taking the ratio of these first order conditions, gives the following simple expression for the relationship between the share of skilled workers, the skill premium

to adjustment costs, to reflect hiring and firing costs. In this approach, however, input prices are assumed to be homogeneous.

and skill-biased technological change:

$$\ln \frac{w_{Sjt}}{w_{Ujt}} = \frac{\sigma_L - 1}{\sigma_L}\omega_{Sjt} - \frac{1}{\sigma_L}\ln \frac{L_{Sjt}}{L_{Ujt}}. \tag{14}$$

This equation forms the basis of the first step of my estimation procedure, and is closely related to the expression at the heart of the literature on skill biased technological change. Disregard the firm-specific subscripts j, and replace the unobserved productivity term ω_{Sjt} with the ratio of factor-augmenting technology terms for skilled and unskilled workers, and the resulting equation is identical to the specification estimated in the seminal paper by Katz and Murphy (1992), and replicated in several other papers.[27] Viewing technological change as operating by changing the efficiencies of skilled versus unskilled labor is therefore equivalent to viewing it as changing these parameters of the production function.

Again disregarding the firm-specific subscripts j, equation (14) shows that the aggregate skill premium is driven by the relative supply of skills in the economy, as well as by skill-augmenting technology. It also clarifies the role played by the elasticity of substitution between the two factors. The larger this is, the less effect changes in relative supply will have on the relative wages.

How relative improvements in skill-augmenting productivity affects the skill premium, also depends crucially on the elasticity of substitution. From the expression above it is easy to see that if $\sigma_L < 1$, the factors are gross complements, and an increase in skill-augmenting productivity is *unskilled-biased* and leads to a reduction in the skill premium. If $\sigma_L = 1$, the skill premium perfectly reflects the demand for skilled workers and there is no role for skill-biased technological change in the traditional sense. If $\sigma_L > 1$, the factors are gross substitutes, and an increase in skill-augmenting productivity is *skilled-biased* and leads to an increase in the skill premium.

The standard approach in this literature is to use variation in the share of skilled workers and the skill premium to estimate σ_L by OLS. However, skill-augmenting technology is not observed and has to be approximated. Katz and Murphy (1992) use a linear time trend to this end, and estimate $\sigma_L = 1.4$ on US census data for 1963 to 1987. Acemoglu and Autor (2011) look at a longer time frame, and include a flexible time trend. Their estimates lie in the range of 1.6-1.8. Indeed, most estimates place σ_L between 1 and 2 when the relevant factors are skilled and unskilled labor.

An equivalent way of formulating this relationship is the following

$$\ln \frac{L_{Sjt}}{L_{Ujt}} = (\sigma_L - 1)\omega_{Sjt} - \sigma_L \frac{w_{Sjt}}{w_{Ujt}}, \tag{15}$$

[27] See for instance Card and DiNardo (2002), Violante (2008) and Acemoglu and Autor (2011).

which eases interpretation when thinking about how to estimate these parameters using firm-level data. Equation (15) makes it cleat that the firms' relative demand for skill depends on the relative price of skills, and the skill-augmenting productivity term ω_{Sjt}. It is also clear that estimating equation (15) by OLS using firm-level data and some proxy for skill-augmenting productivity might lead to biased results. First of all, a time trend can not be expected to capture the process of technological change, which particularly at the firm-level is characterized by nonlinearity and uncertainty. In addition, this unobserved skill-augmenting productivity is likely to be correlated over time, and also to be correlated with the skill premium. If skilled labor becomes relatively more productive, the relative price of skills will rise. This problem is unlikely to be solved by using a firm-specific measure of innovation by itself.

Finally, there is the worry that ω_{Sjt} and $\frac{w_{Sjt}}{w_{Ujt}}$ might co-vary across firms in a non-random fashion. In a perfectly competitive labor market with homogeneous workers that only differ in the level of educational attainment, the skill premium is an equilibrium outcome determined by the relative supply and demand for skills in the local labor market. However, if workers are heterogeneous in ability, differences in the skill premium at the firm level might pick up differences in worker quality, that might also be correlated with the technological advancement of the firm. If technologically more advanced firms employ higher quality skilled workers that are paid a higher wage, this might lead to biased estimates.

The estimation strategy employed in this paper aims to deal with these issues in various ways. Details on this strategy is given in the following section.

4.2 Estimation Strategy

In this section, I describe the estimation strategy. It proceeds in two steps. From the first step procedure, I obtain an estimate of the elasticity of substitution between skilled and unskilled labor σ_L, and I can back out firm-specific, time-varying estimates of unobserved skill-augmenting productivity ω_{Sjt}. These estimates are then employed in a second step, in which I estimate the full production function from equation (9) to obtain the aggregate elasticity of substitution σ_Y, the production function coefficients, and unobserved Hicks-neutral productivity ω_{Hjt}.

To estimate the gross output production function, I follow the approach of Grieco et al. (2016).[28] This estimation technique is well-suited for firm level data where the materials bill is not split into quantity and prices, which is the case for my data. In addition, it has the advantage that firms can face heterogeneous input prices due to for instance quality differences.[29]

[28]See Grieco et al. (2016) for details on this procedure.

[29]Note that quantity discounts, on the other hand, are assumed away in this framework.

This method exploits the first order conditions of the firms' maximization problem that are given in the preceeding section. The use of first order conditions solves the problem of identifying the elasticity of substitution and the bias of technological change simultaneously, which is known as Diamond's Impossibility Theorem (Diamond et al., 1978).

4.2.1 First step

In the estimation strategy proposed in this paper, I treat skilled-augmenting productivity in a manner similar to how Hicks-neutral productivity is often modeled in the productivity estimation literature. In this literature, Hicks-neutral productivity is viewed as an unobservable residual in the production function, that is either controlled for using a proxy function or backed out using structural assumptions on the production function.

In most of the papers in this literature, this unobserved Hicks-neutral productivity is treated as evolving exogenously (Olley and Pakes, 1996; Levinsohn and Petrin, 2003; De Loecker, 2011). More recently, this assumption has been relaxed, and unobserved productivity is in several applications allowed to be endogenously affected by for instance the firm's R&D investments (Doraszelski and Jaumandreu, 2013), exports (De Loecker, 2013), or both (Aw et al., 2011).

I follow this newer strand of the literature and model skilled-augmenting productivity as unobserved, and allow it to evolve over time according to a Markov process that can be endogenously shifted by for instance the firm's R&D investments. For simplicity, I assume that this a linear process.[30] The baseline specification is given by

$$\omega_{Sjt+1} = \mathbb{E}[\omega_{Sjt+1}|\omega_{Sjt}, RD_{jt}] + \xi_{jt+1} = \alpha_0 + \alpha_1\omega_{Sjt} + \alpha_2 RD_{jt} + \xi_{jt+1}. \tag{16}$$

Actual skill-augmenting productivity at time $t+1$ consists of (to the firm) expected skill-augmenting productivity at time t and a random shock ξ_{jt+1}. The expected part of skill-augmenting productivity depends on lagged values, as well as lagged R&D investments. The baseline specification uses a dummy variable for having positive R&D expenditures, since a large share of firms report zero R&D investments.

To derive the first step estimation equation, first solve equation (14) for the unobserved productivity term ω_{Sjt}, and then insert this into the Markov process in equation (16):

$$\ln\frac{L_{Sjt}}{L_{Ujt}} = \tilde{\alpha_0} - \sigma_L \ln\frac{w_{Sjt}}{w_{Ujt}} + \alpha_1\left(\ln\frac{L_{Sjt-1}}{L_{Ujt-1}} + \sigma_L\ln\frac{w_{Sjt-1}}{w_{Ujt-1}}\right) + \tilde{\alpha}_2 RD_{jt-1} + \tau_t + \psi_s + \tilde{\xi}_{jt}, \tag{17}$$

[30] The results reported in the paper are robust to a simple, non-linear extension, namely introducing a quadratic term in the Markov process.

where superscript ˜ indicates that the variable is multiplied by $(\sigma_L - 1)$. Year fixed effects (τ_t) are added to absorb common time shocks, and common industry level shocks are controlled for by industry dummies (ψ_s). Equation (17) is estimated using GMM.

This estimation strategy is reminiscent of the one employed in Doraszelski and Jaumandreu (2014) with the purpose of estimating labor-augmenting (capital-biased) technological change. An advantage of this procedure is that instruments have to be uncorrelated with the innovation term in the Markov process $\tilde{\xi}_{jt}$, not necessarily with skill-augmenting productivity ω_{Sjt} itself. Lagged values of the skill premium and skill shares are by definiton uncorrelated with $\tilde{\xi}_{jt}$, the unanticipated shock to skill-biased productivity in period t. The current skill premium might, on the other hand, be correlated with the current shock to skill-biased productivity. The degree to which it does depends on how much the skill premium reflects exogenous factors such as geographical and temporal differences in labor supply, versus endogenous factors such as the firm-specific quality of labor.

The previous literature (Olley and Pakes, 1996; Doraszelski and Jaumandreu, 2014), posits that lagged values are less susceptible to endogeneity and hence are useful as instruments for current values. In the baseline specification, I follow this strategy. I also generate a firm-level skill premium measure that is purged of unobserved, time-varying individual characteristics such as ability, and show that the results are robust to using this as an instrument for the firm-specific wage premium. Finally, I include the relevant instruments squared.

Another advantage of this framework is that I can easily investigate to what extent other firm activities influence the skill-augmenting productivity process. As explained above, offshoring is expected to have similar observational effects as skill-biased technological change. Hence, I augment the Markov process to take the firm's offshoring activities into account. The resulting specification is given by:

$$\ln \frac{L_{Sjt}}{L_{Ujt}} = \tilde{\alpha_0} - \sigma_L \ln \frac{w_{Sjt}}{w_{Ujt}} + \alpha_1 \left(\ln \frac{L_{Sjt-1}}{L_{Ujt-1}} + \sigma_L \ln \frac{w_{Sjt-1}}{w_{Ujt-1}} \right) + \tilde{\alpha_2} RD_{jt-1} + \tilde{\alpha_3} Off_{jt-1} + \tau_t + \psi_s + \tilde{\xi}_{jt}, \tag{18}$$

where Off_{jt} is a measure of the offshoring intensity of the firm, either the value of imported intermediate inputs divided by operating costs, or the log of this offshoring share. This specification is reminiscent of the one employed in Kasahara et al. (2013), who investigate to what extent offshoring affects skill demand. However, their specification does not control for lagged values of skill-augmenting productivity, and they do not control for investments in R&D or technology.

Lastly, there might be complementarity between skilled workers and capital (Krusell et al., 2000). To evaluate whether this can be driving my results, I extend the specification to include the firm-specific capital-skill ratio. In both instances, the added variables are instrumented using lagged values and their squares.

4.2.2 Second step

From the first step estimation, I retrieve an estimate of the elasticity of substitution σ_L, and I can back out the skill-biased technology term ω_{Sjt}. With these estimates in hand, I proceed to estimating the full production function. I use a gross output production function following the recent critiques by Gandhi et al. (2013) related to the identification of production functions in the presence of variable inputs. They show that the standard methods employed in the productivity estimation literature are not identified, and that the commonly used strategy of using a value added production function does not solve the identification problem.

The intuition is straightforward: variable inputs have no dynamic implications and can be flexibly adjusted each period. This means that lagged input choices do not affect current choices and are not useful as instruments for current choices.

Instead of proxying for unobserved Hicks-neutral productivity by inverting the investment or input demand, like much of the literature has done to date, this procedure exploits the first order conditions of the firms' maximization problem. It is widely known that CES production functions have to be normalized in order to provide meaningful parameters.

I follow Grieco et al. (2016) in using the geometric mean for normalization:

$$\frac{Y_{jt}}{\bar{Y}} = \left(\beta_K \left(\frac{K_{jt}}{\bar{K}}\right)^{\frac{\sigma_Y - 1}{\sigma_Y}} + \beta_L \left(\frac{L^*_{jt}}{\bar{L}^*}\right)^{\frac{\sigma_Y - 1}{\sigma_Y}} + \beta_M \left(\frac{M_{jt}}{\bar{M}}\right)^{\frac{\sigma_Y - 1}{\sigma_Y}}\right)^{\frac{\sigma_Y}{\sigma_Y - 1}} e^{\omega_{Njt}} e^{\epsilon_{jt}}.$$

The distributional parameters β_K, β_L and β_M are restricted to sum to 1.[31]

Recall the ratio of the first order conditions of materials and aggregate labor found above in equation (13). It can be shown that the (normalized) level of materials associated with this equation is given by

$$\frac{M_{jt}}{\bar{M}} = \left(\frac{\beta_L}{\beta_M} \frac{E_{M_{jt}}}{E_{L_{jt}}}\right)^{\frac{\sigma_Y}{\sigma_Y - 1}} \frac{L^*_{jt}}{\bar{L}^*}. \tag{19}$$

Based on this expression, variation in expenditure ratios and optimal input demands can be used to separate the materials quantities from the prices.[32]

[31] This assumption of constant returns to scale can be easily relaxed by adding a scale parameter. However, this scale parameter cannot be indentified separately from the demand elasticity without additional assumptions.

[32] This only holds if $\sigma_Y \neq 1$, so the Cobb-Douglas functional form can not be applied. For $\sigma_Y = 1$, the expenditure ratio is constant and hence it is not possible to infer prices and quantities from equation (19). See the Appendix for an estimation procedure in the Cobb-Douglas case.

Since output quantitites are not observed, the revenue equation is used to estimate both demand and production parameters, following Klette and Griliches (1996). Revenue is given by

$$R_{jt} = P_{st}(Y_{jt}; \eta) Y_{jt}.$$

Inserting for the materials quantity using equation (19) and taking logs gives the following simple second step estimating equation:

$$\ln R_{jt} = \ln \frac{\eta}{1+\eta} + \ln \left[E_{M_{jt}} + E_{L_{jt}} \left(1 + \frac{\beta_K}{\beta_L} \left(\frac{K_{jt}/\bar{K}}{L^*_{jt}/\bar{L}^*} \right)^{\gamma} \right) \right] + \Upsilon_t + \zeta_s + \epsilon_{jt}, \tag{20}$$

where Υ_t and ζ_s are year and industry fixed effects, respectively, added to absorb common shocks.

Two additional restrictions are needed to identify all parameters:

$$\frac{\beta_M}{\beta_L} = \frac{\bar{E}_{M_{jt}}}{\bar{E}_{L_{jt}}},$$

and

$$\beta_L + \beta_M + \beta_K = 1,$$

The first restriction is implied by profit maximization, and the second follows from the restriction on the sum of the distribution parameters in the production function. This second step equation is estimated using nonlinear least squares (NLS).[33]

Finally, I can back out the Hicks-neutral productivity term ω_{Njt} using the following expression:

$$\omega_{Njt} = \frac{\eta}{1+\eta} \ln \left(\frac{\eta}{1+\eta} \frac{1}{\beta_L} \frac{Y_t^{1/\eta}}{P_t} \left(\frac{L^*_{jt}}{\bar{L}^*} \right)^{-\frac{\sigma_Y}{\sigma_Y - 1}} \frac{E_{L_{jt}}}{\bar{Y}^{\frac{1+\eta}{\eta}}} \right. \tag{21}$$
$$\left. \left[\beta_L \frac{E_{L_{jt}} + E_{M_{jt}}}{E_{L_{jt}}} \left(\frac{L^*_{jt}}{\bar{L}^*} \right)^{\frac{\sigma_Y}{\sigma_Y - 1}} + \beta_K \left(\frac{K_{jt}}{\bar{K}} \right)^{\frac{\sigma_Y}{\sigma_Y - 1}} \right]^{-\frac{\sigma_Y - 1}{\sigma_Y}(1+\frac{1}{\eta})} \right).$$

This will allow me to compare estimates of ω_{Njt} with and without taking ω_{Sjt} and the adjusted labor aggregate L^*_{jt} into account.

To summarize the estimation procedure: from the first step estimation I obtain estimates of $\hat{\sigma}_L$ and $\hat{\omega}_{Sjt}$. These are subsequently plugged into the labor aggregator function (equation (10)) to

[33] It can also be estimated using GMM. See Grieco et al. (2016) for details.

calculate $\hat{L}^*_{jt}$. This estimated labor aggregate is used in the second second step estimating equation, given by equation (20).

5 Structural estimation results

I will first present the main results, from the first step estimation. I will then describe the second step results, which are mainly used to back out the Hicks-neutral productivity term ω_{Njt}. I can show that taking the skill-biased component of firm productivity into account matters for the distribution of this Hicks-neutral productivity term.

5.1 Skill-biased technological change

To evaluate how my estimation procedure relates to the previous literature, I start by running regressions of the log college share on the log wage premium and a time trend. The results are given in Table 7. The OLS produces an unprecisely estimated coefficient of -0.19, with the time trend showing a roughly 4 percent increase per year. However, the OLS might be biased upward if, say, larger firms pay higher wages and have a higher skill share. As expected, if firm fixed effects are added, the coefficient increases (in absolute size) to -0.34 and is statistically significant at the 1 percent level, indicating that a rising skill premium reduces the demand for skilled workers. The coefficient on the time trend is reduced somewhat, to 3 percent.

Since a time trend cannot be expected to successfully proxy for technological change at the firm level, I repeat both exercises replacing the time trend with a dummy for lagged R&D expenditures. As above, the coefficient in the OLS is likely biased upward, reflecting the endogeneity issue. Again, adding firm fixed effects brings the estimate back to a similar magnitude as in the specification with fixed effects and a time trend (column (2)). However, the coefficient might still be biased. On the one hand, adding firm fixed effects solves only part of the endogeneity problem since it controls for time-invariant productivity differences between firms. This implies that if a firm is hit by a positive productivity shock, it might subsequently start doing R&D and change the workforce composition, without R&D itself driving the change in skill demand. This would lead to an upward bias. On the other hand, in the presence of measurement error, the fixed effects regression will suffer from downward bias. A priori, there is no way of telling which of these biases are stronger.[34]

[34] This pattern of differences between OLS and fixed effects estimates is reminiscent of the results from Doms et al. (1997), who evaluate how plant-level productivity, wages, and occupational and educational mix vary with the adoption

All these specifications provide estimates on the elasticity of substitution that can be compared to the previous literature. As mentioned above, Katz and Murphy (1992) estimate an elasticity of substitution equal to 1.4, much higher than the 0.34 my naive results are indicating. This means that at the firm level, it appears to be much harder to substitute skilled and unskilled workers than what is found using different levels of aggregation. That is, we observe firms having a much higher skill demand than what we would expect them to have, given the size of the skill premium, for any true value of the elasticity of substitution higher that 0.34.

The most likely explanation is that the time trend or lagged R&D expenditures are are poor proxies for skill-biased technological change at the firm level. It could also be that the firm-level skill premium is in part reflecting differences in worker quality. If there is positive assortative matching in the labor market, more productive firms will have more able workers, who are paid a higher wage. These firms will then be observed as having a larger share of skilled workers than what we would expect, given the high skill premium they are paying.

The first issue is solved by the estimating strategy I employ, which allows unobserved firm-specific skill-augmenting productivity to evolve over time and be endogenously shifted, but not solely determined, by R&D activities. The second potential explanation I attempt to deal in two ways. First, building on the intuition from the previous literature, lagged values should be less plagued by endogeneity issues than present values (Doraszelski and Jaumandreu, 2014). Hence, I instrument for the current skill premium using lagged values. In addition, I can generate a cleaned wage ratio to use as an instrument for the firm-specific skill premium. To this end, I run wage regressions using worker level data and individual fixed effects, and retrieve predicted wages without the worker-specific component. The results are robust to using a firm-specific wage ratio based on these wage measures purged of worker quality as an instrument for the firm-specific skill premium.

The results from the first step estimation are given in Table 8. Column (1) presents the results from the baseline specification in equation (17). The estimated elasticity of substitution between skilled and unskilled labor is 1.35, very close to the estimates previously found in the literature, although somewhat smaller.[35] This indicates that skilled and unskilled workers are substitutes, not complements, in the production function. This, in turn, asserts that increases in skill-augmenting

of new technologies. Their cross-sectional results indicate that there is a strong positive correlation between technology and workforce characteristics. On the other hand, their time-series results show little correlation between changes in adoption and changes in workforce characteristics – instead they find it is the case that already high-wage plants are quicker to adopt new technologies. This highlights the difficulties with identifying the effects of innovation on firm-level outcomes in the absence of exogenous variation: not controlling for firm fixed effects is likely to overestimate the effects due to endogeneity issues, while adding firm fixed effects generates estimates that are likely to suffer from measurement error and hence be downward biased.

[35] Recall the estimate from Katz and Murphy (1992) of 1.42, and 1.6-1.8 is found in Acemoglu and Autor (2011).

productivity indeed implies skill-biased change.[36]

The estimated coefficient on the lagged term in the Markov process is 0.920, indicating that skill-augmenting productivity is highly persistent at the firm level. The estimated direct impact of lagged R&D on skill demand, conditional on lagged skill-augmenting productivity, is 3.5 percent, and statistically significant at the 1 percent level. In this specification, the lagged R&D term $R\&D_{jt-1}$ is given by a dummy for having positive R&D expenditures. In column (2), this dummy is replaced by a continuous measure, $\ln(1 + \text{R\&D expenditure}_{jt-1})$. The coefficient on this continuous measure is 0.005, again significant at the 1 percent level.

Turning now to the impact of offshoring on skill-augmenting productivity, terms capturing the firm's offshoring activities are included in columns (3) and (4). In the first, Off_{jt-1} is given by the log of the offshoring share, while in the second, it is the level of the offshoring share, to include firms that do not offshore at all.[37] While the estimated coefficient on the offshoring variables have the expected sign, they are not statistically significantly different from zero. This indicates that offshoring does not play a major role in determining skill-augmenting productivity, which is consistent with previous findings in the literature.[38] This is not saying that international trade has no effect on the skill premium. The large coefficient on the lagged skill-augmenting productivity term suggests that there are large, persistent differences between firms. These differences might be driven by for instance differential exposure to import competition, differences in management practices, or technological differences. Note also that the coefficient on lagged R&D remains virtually unchanged, indicating that this R&D effect is indeed coming from the firm's own innovation activities, and not through the potential complementarity with imports.

In column (5), a term controlling for the capital-skill ratio is added to the Markov process. This is reminiscent of the specification in Krusell et al. (2000), and captures the fact that there might be complementarities between skilled workers and capital. The coefficient on this term is positive, reflecting that there indeed seems to be some capital-skill complementarity at the firm level. However, the estimated coefficient is not statistically significantly different from zero, and the other coefficients remain largely unchanged by the inclusion of this term.

Based on these estimates, I can back out a measure of skill-augmenting productivity at the firm level. Using the estimated coefficients in the baseline specification of equation (17), I use the expression in this equation to generate a measure of ω_{Sjt}, purged of common industry and year

[36]Note that the estimate is not statistically significantly different from one.

[37]A very small fraction of firms in my sample do not import anything. Hence, using a dummy variable for offshoring will not be informative.

[38]Goel (2014) finds that offshoring affects the skill premium, but only through increasing innovation. Controlling for innovation, the separate effect of offshoring is negligible, in line with my results.

fixed effects. Figure 4 plots the distribution of this term for firms investing in R&D and firms that do not invest in R&D separately. The pattern is clear: R&D firms have higher levels of skill-augmenting productivity for all levels of this term. Table 9 shows the average growth rates of ω_S for all firms, as well as for firms with and without positive R&D expenditures separately. The average growth rate of the skill-augmenting productivity term is 10.7 percent per year. This is slightly higher for firms investing in R&D at 11.4 percent compared to 9.7 percent for non-innovating firms. This difference is not statistically significant at any conventional level, however. The correlation between the level and growth rate of skill-augmenting productivity is positive, again indicating that differences between firms are persistent over time.

To get a feel for the magnitude of the estimated skill-biased technological change term, I can compare it to the average growth in the relative supply of skill and the skill premium in the Norwegian manufacturing sector over my sample period. Applying a simple model like the one in equation (14) and comparing the change in the aggregate skill ratio to the change in the skill premium from 1997 to 2010, I can calculate the rate at which skill-biased technological change has had to increase to match the observed increase in the skill premium. Using the estimate of the elasticity of substitution from Katz and Murphy (1992) – which is very close to the one I estimate above – I find that the average growth of skill-biased technological change has been roughly 13 percent per year. Compared to my estimates of 10.7 percent, this suggests that my estimated average yearly within-firm growth rate of skill-augmenting productivity term can account for around 80 percent of the observed increase in the skill premium over my sample period.

The estimated effect of lagged R&D on skill demand is a short term effect. To examine the long-run impact of R&D on skill demand, I can iterate on the Markov process in equation (16). Comparing firms with positive R&D in all periods to firms that never invest in R&D, I find that R&D increases the skill demand by 17.7 percent on average.

5.1.1 Robustness

R&D workers. As with the reduced form results, if this estimated effect of R&D on skill demand is due to skill-biased technological change and not simply due to the fact that firms wanting to invest in R&D need to hire research personnel, the results should hold even if R&D workers are removed from the sample. Column (1) of Table 14 in the Appendix shows the results from this exercise. The estimation procedure delivers a very large estimate of the elasticity of substitution in this "counterfactual" case. However, the estimated impact of R&D on skill demand is roughly similar to the baseline specification. In column (2), I see how results change if I impose an elasticity of substitution from my baseline estimate and re-estimate this counterfactual scenario. The estimated impact of R&D on skill demand is in this case substantially higher than when all workers were included, mirroring the findings from the reduced form exercises. This again indicates that the skill-biased technological change channel is stronger than the effect of hiring research personnel.

Firm-specific wages. Another concern is that the variation in firm-specific wages is coming not only from local labor market conditions, but from factors related to the differences in the quality or composition of labor in the firm. If firms hire workers of different quality, these quality differences might show up as differences in firm-specific wages. Using lagged values of firm-specific skill premium as part of the instrument set instead of current values should partly alleviate these concerns, and is the strategy typically followed in the productivity estimation literature (Doraszelski and Jaumandreu, 2014). The validity of this strategy does, however, rest on assumptions on there being no autocorrelation in the error term.

A first reason why this should be less of a problem in this context than in many others, is that the Norwegian labor market is characterized by a large degree of centralized, industry-wide wage bargaining. This implies that the link between firm productivity and firm-specific wages is weak, relieving some of these concerns.[39] Still, I perform a number of robustness checks to evaluate the extent to which the coefficient estimates might be biased due to this issue.

First, I construct industry-specific skill-premia, and use this to replace the firm-specific wage measure in the instrument set. The result from this exercise can be found in column (3) of Table 14 in the Appendix, and show that the coefficients remain largely unchanged. Second, I generate local labor market-specific skill-premia, using the labor market regions as defined by the Norwegian Institute for Urban and Regional Research (NIBR). Again, I replace the firm-specific wage measure in the instrument set and rerun the estimation. Results are given in column (4) of Table 14 in the Appendix, and are again very close to the baseline estimates.

Finally, given the detailed matched employer-employee data set, I can create a skill premium measure at the firm-level that is purged of unobserved worker heterogeneity. To generate, this I run a standard Mincer type wage regression with individual fixed effects, obtaining a predicted wage measure without the worker-specific component. This effectively removes all variation from wages arising from time-invariant observable and unobservable characteristics, such as ability. Based on this, I generate a new firm-specific skill premium. In column (5) of Table 14 in the Appendix, I re-estimate the model using this cleaned measure of relative skilled wages. The results are very similar to the baseline specification. The elasticity of substitution is somewhat higher, and the effect of R&D somewhat lower.

[39] See for instance Moene and Wallerstein (1997).

Non-constant elasticity of substitution. Another worry is that the elasticity of substitution between skilled and unskilled workers might not be constant between firms and over time, and that this assumption will bias the results. I split the sample in two based on time and R&D status of the firm, and perform the first step estimation on these subsamples separately.[40] Results are presented in Table 15 in the Appendix, and show that the elasticity of substitution between skilled and unskilled workers does not seem to very along these two dimensions.

Production function specification. Lastly, to make sure these results are not driven entirely by the particular CES specification of the production function, I estimate a different specification grounded in a Cobb-Douglas production function framework. A standard Cobb-Douglas production function is augmented with interaction terms between the various inputs (capital, skilled and unskilled labor) and a dummy for R&D, capturing potential complementarities between the input in question and innovation, which alternatively can be thought of as a proxy for technology. I employ a control function approach to account for Hicks-neutral productivity in the spirit of Levinsohn and Petrin (2003). I again follow the approach of Doraszelski and Jaumandreu (2013) by allowing, this time, the Hicks-neutral productivity term to evolve over time and be endogenously affected by R&D.

The results from this exercise complement the ones described above, by showing that innovation increases the output elasticity of skilled workers, while reducing that of unskilled workers. More details on the procedure and results from estimating this specification can be found in Section F.1 in the Appendix.

5.2 Production function results

In the second step, the full production function is estimated as explained in Section 4.2.2, using NLS. The results are given in Table 10. Column (1) takes the skill-augmenting productivity term into account, and uses L^*_{jt} as the measure of aggregate labor. In column (2), skill-augmenting productivity is ignored and the estimates are simply based on the number of skilled and unskilled workers in the firm. That is, I use the same elasticity of substitution to create an aggregate labor measure, ignoring ω_{Sjt}. The coefficients in both columns are in line with previous findings in the literature estimating gross output specifications.[41] The coefficients are not substantially different across the two specifications. To the extent that they differ, the pattern is quite intuitive. Taking

[40]The estimating equation is simplified by disregarding the R&D effect for the latter case.

[41]See for instance Amiti and Konings (2007), Gandhi et al. (2013) and Grieco et al. (2016).

skill-augmenting productivity into account leads to a somewhat higher coefficient on the aggregate labor input, and somewhat lower coefficient on capital.[42]

While not crucial for the main results of interest in this paper, it is interesting to note that the estimated elasticity of substitution between the labor aggregate, capital and materials is somewhat higher than the elasticity of substitution between the two types of labor, not lower, as expected. The estimated coefficients are very similar in size to the ones reported in Grieco et al. (2016), where the estimates range from 1.4 to 2.6 using Colombian firm level data. They claim that the traditional methods underestimate the elasticity of substitution due to unobserved input price heterogeneity. They run Monte Carlo experiments showing that their method performs significantly better than the proxy function approaches often used in the literature. However, their estimates are higher than for instance the ones found by Doraszelski and Jaumandreu (2014), which range from 0.44 to 0.80, or the ones found by Raval (2015), which range from 0.4 to 0.7.

Based on the estimated coefficients from this second step, I can back out a measure of firm-level Hicks-neutral technological change from the expression in equation (21). Figure 5 plots the distribution of this Hicks-neutral productivity term for firms with and without positive R&D expenditures when not taking the skill-augmenting productivity term into account. The labor aggregate L_{jt} is simply the CES combination of skilled and uskilled workers combined using the estimated elasticity of substitution.

Figure 6, on the other hand, uses the first step estimation results to take skill-augmenting productivity into account. It is clear that the resulting distribution of Hicks-neutral productivity is less dispersed than what is found if not accounting for the skill-augmenting component of productivity. This highlights that caution has to be exercised when interpreting estimates of Hicks-neutral productivity in the presence of non-neutral technological change.

Finally, the results suggest that firms with higher levels of Hicks-neutral productivity also have higher levels of skill-biased productivity. The correlation between the two measures is 0.59. This implies that policy changes that encourage growth of the most productive firms, like trade liberalization, will potentially also increase skill demand.

6 Conclusions

In this paper, I estimate the extent to which innovation affects skill-biased productivity and the firm's demand for skill. To this end, I combine structural estimation with reduced form evidence, using rich Norwegian firm-level data. For the reduced form estimates, I exploit the introduction

[42]Recall that the distributional parameters are restricted to sum to one in this specification.

of a tax credit, which provides a way of breaking the correlation between unobserved productivity and R&D that otherwise makes it difficult to identify the causal impact of R&D on the firms' skill demand without structural assumptions. The results suggest that firms induced to start innovating also increase the shares and ratios of high-skilled workers, both in a difference-in-difference framework, and by exploiting the variation from the reform as an instrumental variable.

I combine this reduced form evidence with structural production function estimation to evaluate not only the contribution of R&D to skill demand, but the more general question of to what extent productivity is skill-biased at the firm level. I provide a novel strategy to estimate skill-biased productivity using firm-level data, taking the firm's innovation activities into account. The modeling of the evolution of unobserved skill-biased productivity closely follows the way unobserved Hicks-neutral productivity is modeled in the productivity estimation literature. The results confirm that technological change exhibits skill-bias, and is affected by firms' investments in R&D.

I find that the within-firm growth of skill-biased productivity can account for the majority of the observed rise in the skill premium in Norway over the sample period. Consistent with the reduced form results, R&D firms have higher levels and growth rates of skill-biased technological change. The estimated direct short-run (long-run) impact of own R&D investments on skill demand is roughly 3.5 percent (17.7 percent). These results are robust to extending the estimation in order to take the firms' offshoring activities into account. They are also not driven by firms starting to hire more R&D workers. Lastly, I show that the results are robust to a different specification of the production function and a different estimation strategy.

Determining the consequences of innovation not only for aggregate productivity, but for how it affects the production processes at the firm level and the firm's demand for skill, are important questions both in economics and for policy making. If technological change is skill biased at the firm level, policies promoting R&D investments will not only affect aggregate productivity, but also the firms' demand for skills. Hence, such policies can drive up the skill premium and, by extension, income inequality.

References

Abowd, J. M., J. Haltiwanger, J. Lane, K. L. McKinney, and K. Sandusky (2007, April). Technology and the Demand for Skill: An Analysis of Within and Between Firm Differences. NBER Working Papers 13043, National Bureau of Economic Research, Inc.

Acemoglu, D. (2003). Patterns of skill premia. *The Review of Economic Studies 70*(2), 199–230.

Acemoglu, D. and D. Autor (2011). *Handbook of Labor Economics*, Volume 4, Chapter Skills, Tasks and Technologies: Implications for Employment and Earnings, pp. 1043–1171. Amsterdam: Elsevier.

Ackerberg, D. A., K. Caves, and G. Frazer (2006). Structural identification of production functions.

Akerman, A., I. Gaarder, and M. Mogstad (2015). The skill complementarity of broadband internet. *The Quarterly Journal of Economics 130*(4), 1781–1824.

Akhmetova, Z. and S. M. Ferguson (2015). Offshoring, total factor productivity and skill-biased technological change. IFN Working Paper 1074.

Amiti, M. and J. Konings (2007). Trade liberalization, intermediate inputs, and productivity: Evidence from Indonesia. *American Economic Review 97*(5), 1611–1638.

Atkeson, A. and A. T. Burstein (2011). Aggregate implications of innovation policy. *NBER Working Papers* (17493).

Autor, D. H., F. Levy, and R. J. Murnane (2003). The skill content of recent technological change: An empirical exploration. *The Quarterly Journal of Economics 118*(4), 1279–1333.

Aw, B. Y., M. J. Roberts, and D. Y. Xu (2011). R&D investment, exporting, and productivity dynamics. *American Economic Review 101*(4), 1312–1344.

Bernard, A. B. and J. B. Jensen (1995). Exporters, jobs, and wages in US manufacturing: 1976-1987. *Brookings Papers on Economic Activity. Microeconomics 1995*, 67–119.

Bøler, E. A., A. Moxnes, and K.-H. Ulltveit-Moe (2015). R&D, International Sourcing, and the Joint Impact on Firm Performance. *American Economic Review 105*(12), 3704–39.

Bound, J. and G. Johnson (1992). Changes in the structure of wages in the 1980's: An evaluation of alternative explanations. *The American Economic Review 82*(3), pp. 371–392.

Bresnahan, T. F., E. Brynjolfsson, and L. M. Hitt (2002). Information technology, workplace organization, and the demand for skilled labor: Firm-level evidence. *The Quarterly Journal of Economics 117*(1), 339–376.

Burstein, A. and J. Vogel (2010, October). Globalization, Technology, and the Skill Premium: A Quantitative Analysis. NBER Working Papers 16459, National Bureau of Economic Research, Inc.

Bustos, P. (2011). Trade liberalization, exports, and technology upgrading: Evidence on the impact of MERCOSUR on Argentinian firms. *American Economic Review 101*(1), 304–340.

Cappelen, Å., E. Fjærli, F. Foyn, T. Hægeland, J. Møen, A. Raknerud, and M. Rybalka (2010). Evaluation of the norwegian R&D tax credit scheme. Technical Report No. 640, SSB Discussion Papers.

Card, D. and J. E. DiNardo (2002, October). Skill-Biased Technological Change and Rising Wage Inequality: Some Problems and Puzzles. *Journal of Labor Economics 20*(4), 733–783.

Costantini, J. A. and M. J. Melitz (2007). *The Organization of Firms in a Global Economy*, Chapter The Dynamics of Firm-Level Adjustment to Trade Liberalization. Harvard University Press.

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

De Loecker, J. (2013). Detecting learning by exporting. *American Economic Journal: Microeconomics 5*(3), 1–21.

Diamond, P., D. McFadden, and M. Rodriguez (1978). Measurement of the Elasticity of Factor Substitution and Bias of Technical Change. In M. Fuss and D. McFadden (Eds.), *Production Economics: A Dual Approach to Theory and Applications*, Volume 2 of *Histoy of Economic Thought Chapters*, Chapter 5. McMaster University Archive for the History of Economic Thought.

Doms, M., T. Dunne, and K. R. Troske (1997). Workers, wages, and technology. *The Quarterly Journal of Economics 112*(1), 253–290.

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

Doraszelski, U. and J. Jaumandreu (2014, November). Measuring the bias of technological change. CEPR Discussion Papers 10275, C.E.P.R. Discussion Papers.

Feenstra, R. C. and G. H. Hanson (1999, August). The Impact Of Outsourcing And High-Technology Capital On Wages: Estimates For The United States, 1979-1990. *The Quarterly Journal of Economics 114*(3), 907–940.

Fox, J. T. and V. Smeets (2011). Does input quality drive measured differences in firm productivity? *International Economic Review 52*(4), 961–989.

Gandhi, A., S. Navarro, and D. Rivers (2013). On the identification of production functions: How heterogeneous is productivity? Technical report.

Glass, A. J. and K. Saggi (2001). Innovation and wage effects of international outsourcing. *European Economic Review 45*, 67–86.

Goel, M. (2014). Does offshoring lift all boats? The role of induced technology adoption and innovation.

Goldberg, P. K. and N. Pavcnik (2007). Distributional effects of globalization in developing countries. *Journal of Economic Literature 45*(1), 39–82.

Grieco, P., S. Li, and H. Zhang (2016). Production function estimation with unobserved input price dispersion. *International Economic Review 57*, 665–690.

Griliches, Z. (1979). Issues in assessing the contribution of research and development to productivity growth. *The Bell Journal of Economics 10*(1), pp. 92–116.

Hægeland, T. and J. Møen (2007). Input additionality in the Norwegian R&D tax credit scheme, Report 2007/47 Statistics Norway.

Harrigan, J. and A. Reshef (2015). Skill biased heterogeneous firms, trade liberalization, and the skill premium. *Canadian Journal of Economics 48*.

Helpman, E., O. Itskhoki, and S. Redding (2010). Inequality and unemployment in a global economy. *Econometrica 78*(4), 1239–1283.

Kasahara, H., Y. Liang, and J. Rodrigue (2013). Does Importing Intermediates Increase the Demand for Skilled Workers? Plant-level Evidence from Indonesia. CESifo Working Paper Series 4463, CESifo Group Munich.

Katz, L. F. and K. M. Murphy (1992, February). Changes in Relative Wages, 1963-1987: Supply and Demand Factors. *The Quarterly Journal of Economics 107*(1), 35–78.

Klette, T. J. and Z. Griliches (1996, July-Aug.). The Inconsistency of Common Scale Estimators When Output Prices Are Unobserved and Endogenous. *Journal of Applied Econometrics 11*(4), 343–61.

Krusell, P., L. E. Ohanian, J.-V. Rios-Rull, and G. L. Violante (2000). Capital-skill complementarity and inequality: A macroeconomic analysis. *Econometrica 68*(5), 1029–1053.

Levinsohn, J. and A. Petrin (2003). Estimating production functions using inputs to control for unobservables. *Review of Economic Studies 70*(2), 317–342.

Levy, F. and R. J. Murnane (1996). With what skills are computers a complement? *The American Economic Review 86*(2), pp. 258–262.

Lileeva, A. and D. Trefler (2010). Improved access to foreign markets raises plant-level productivity ... for some plants. *The Quarterly Journal of Economics 125*(3), 1051–1099.

Moene, K. O. and M. Wallerstein (1997, July). Pay Inequality. *Journal of Labor Economics 15*(3), 403–30.

Monte, F. (2011, March). Skill bias, trade, and wage dispersion. *Journal of International Economics 83*(2), 202–218.

Mueller, H. M., P. P. Ouimet, and E. Simintzi (2015, January). Wage inequality and firm growth. Working Paper 20876, National Bureau of Economic Research.

OECD (2011). Divided we stand: Why inequality keeps rising. Technical report.

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

Pavcnik, N. (2002, January). Trade Liberalization, Exit, and Productivity Improvement: Evidence from Chilean Plants. *Review of Economic Studies 69*(1), 245–76.

Raknerud, A., D. Rønningen, and T. Skjerpen (2004). Documentation of the Capital Database. *Statistics Norway Documents 2004/16*, 1–13.

Raval, D. (2015). The micro elasticity of substitution and non-neutral technology. Technical report.

Rodriguez-Clare, A. (2010). Offshoring in a Ricardian world. *American Economic Journal: Macroeconomics 2*(2), 227–258.

Santos Silva, J. and S. Tenreyro (2006). The log of gravity. *Review of Economics and Statistics 88*(4), 641–658.

Song, J., D. J. Price, F. Guvenen, N. Bloom, and T. von Wachter (2015, May). Firming up inequality. Working Paper 21199, National Bureau of Economic Research.

Van Biesebroeck, J. (2003). Productivity dynamics with technology choice: An application to automobile assembly. *The Review of Economic Studies 70*(1), pp. 167–198.

Violante, G. (2008). *The New Palgrave Dictionary of Economics* (2nd ed.)., Chapter Skill–biased technical change. Palgrave Macmillan, New York.

Wooldridge, J. M. (1999). Distribution-free estimation of some nonlinear panel data models. *Journal of Econometrics 90*(1), 77 – 97.

Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data*. MIT Press.

Yeaple, S. R. (2005, January). A simple model of firm heterogeneity, international trade, and wages. *Journal of International Economics 65*(1), 1–20.

Zhang, H. (2015). Non-neutral technology, firm heterogeneity, and labor demand. Technical report.

Appendix

A Tables

Table 1: R&D Premia.

	log Size	log Coll Share	Edu	log Wage	Labor Prod	Skill Prem
	(1)	(2)	(3)	(4)	(5)	(6)
R&D dummy	0.39***	0.36***	0.03***	0.04***	0.08***	0.05***
	(0.08)	(0.04)	(0.00)	(0.01)	(0.03)	(0.01)
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	4156	4156	4156	4156	3909	4156

Standard errors in parenthesis are clustered on firm. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 2: Decomposition of changes in the shares of workers, 1997-2010.

	(1)	(2)
	Skilled	Unskilled
Share	0.241	0.759
Δ Share	0.071	-0.071
Δ Between	0.030	0.022
Δ Within	0.040	-0.093

Table 3: Reduced form evidence of increased R&D.

	(1)	(2)
Outcome	log R&D exp	Δ log R&D exp
$\geq 2002 \times$Treated		0.433^{***}
		(0.134)
$1999\times$ Treated	-0.088	
	(0.138)	
$2001\times$ Treated	-0.076	
	(0.138)	
$2003\times$ Treated	0.374^{**}	
	(0.166)	
$2004\times$ Treated	0.431^{***}	
	(0.155)	
$2005\times$ Treated	0.308^{**}	
	(0.153)	
$2006\times$ Treated	0.297^{*}	
	(0.153)	
Observations	3072	1935

Both specifications include firm, industry and year fixed effects. Standard errors in parenthesis are clustered on firm. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 4: Reduced form evidence of increased R&D.

	Linear	Linear	FE-PPML	FE-PPML	FE-PPML
	(1)	(2)	(3)	(4)	(5)
Outcome	Δ log R&D exp	Δ log R&D int	R&D exp (>0)	R&D exp	R&D int
$\geq 2002\times$Treated	**0.433*****	**0.391****	**0.356***	**0.826*****	**1.097*****
	(0.134)	(0.154)	(0.189)	(0.131)	(0.133)
Firm FE	Yes	Yes	Yes	Yes	Yes
Group trend	No	No	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
Observations	1935	1935	1935	3375	3375

Standard errors in parenthesis are clustered on firm. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 5: Reduced form evidence of increased skill demand.

	Linear	Linear	Linear
Outcome	Δ College share	Δ log College share	Δ log College ratio
$\geq 2002\times$Treated	0.011***	0.039**	0.046*
	(0.004)	(0.019)	(0.024)
Firm FE	Yes	Yes	Yes
Industry + Year FE	Yes	Yes	Yes
Observations	4184	4156	4156

Standard errors in parenthesis are clustered on firm. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 6: Reduced form evidence of increased skill demand.

Outcome	OLS log College ratio	FE log College ratio	2SLS log College ratio
RD_{t-1}	0.404***	0.030**	0.334**
	(0.046)	(0.014)	(0.154)
Firm FE	No	Yes	Yes
Industry FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Observations	4054	4054	4054
Outcome		First stage RD_{t-1}	Reduced form log College ratio
$\geq 2002 \times \text{Treated}_{t-1}$		0.171***	0.056**
		(0.034)	(0.027)
KP F stat		35.34	
Firm FE		Yes	Yes
Industry FE		Yes	Yes
Year FE		Yes	Yes
Observations		4054	4054

Standard errors in parenthesis are clustered on firm. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 7: College ratio.

Outcome: log College ratio	(1) OLS	(2) FE	(3) OLS	(4) FE
log Skill premium	-0.019	-0.342***	-0.167	-0.314***
	(0.057)	(0.056)	(0.141)	(0.057)
Time trend	0.042***	0.030***		
	(0.003)	(0.003)		
RD_{jt-1}			0.409***	0.039***
			(0.046)	(0.015)
Firm FE	No	Yes	No	Yes
Year FE	No	No	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes
Observations	4156	4156	4156	4156

Standard errors in parenthesis are clustered on firm. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 8: First step estimates.

Outcome: $\log \frac{L_S}{L_U}$	(1) Baseline	(2) Cont. R&D	(3) W/log Off share	(4) W/Off share	(5) W/Capital-Skill
σ_L	1.35***	1.27***	1.24***	1.35***	1.47***
	(0.240)	(0.235)	(0.230)	(0.240)	(0.239)
$\alpha_1(\omega_{Sjt-1})$	0.920***	0.920***	0.926***	0.920***	0.937***
	(0.009)	(0.009)	(0.009)	(0.009)	(0.007)
$\tilde{\alpha_2}$ (RD_{jt-1})	0.035***	0.005***	0.035***	0.034***	0.030***
	(0.009)	(0.001)	(0.009)	(0.009)	(0.008)
$\tilde{\alpha_O}$ (Off_{jt-1})			0.002	0.014	
			(0.003)	(0.021)	
$\tilde{\theta}$ $(\ln \frac{K_{jt-1}}{L_{Sjt-1}})$					0.015
					(0.110)
Industry FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
Observations	3965	3965	3878	3965	3965

In columns 1, 3-5, the R&D variable is a dummy for R&D exp > 0. In column 2, the the R&D variable is log(1+R&D expenditure). Standard errors in parenthesis are clustered on firm. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 9: Skill-biased productivity differences.

	(1)
$\Delta\omega_{Sjt}$	0.107
$\Delta\omega_{Sjt}\|RD = 0$	0.097
$\Delta\omega_{Sjt}\|RD = 1$	0.114
$corr(\omega_{Sjt}\Delta\omega_{Sjt})$	0.11

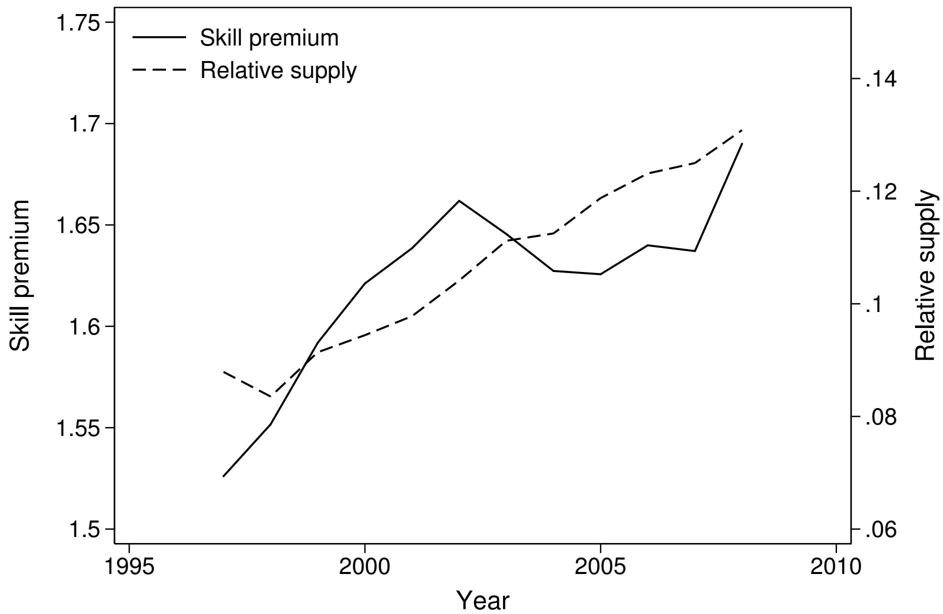
Table 10: Production function estimates.

	(1)	(2)
Outcome: log R	With L^*	W/o L^*
β_M	0.697***	0.678***
	(0.003)	(0.008)
β_L	0.253***	0.247***
	(0.001)	(0.001)
β_K	0.050***	0.075***
	(0.004)	(0.005)
η	-10.584***	-10.171***
	(0.889)	(0.979)
σ_Y	1.635***	1.526***
	(0.0362)	(0.074)
Industry FE	Yes	Yes
Year FE	Yes	Yes
Observations	3965	3965

Bootstrapped standard errors in parenthesis are clustered on firm. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

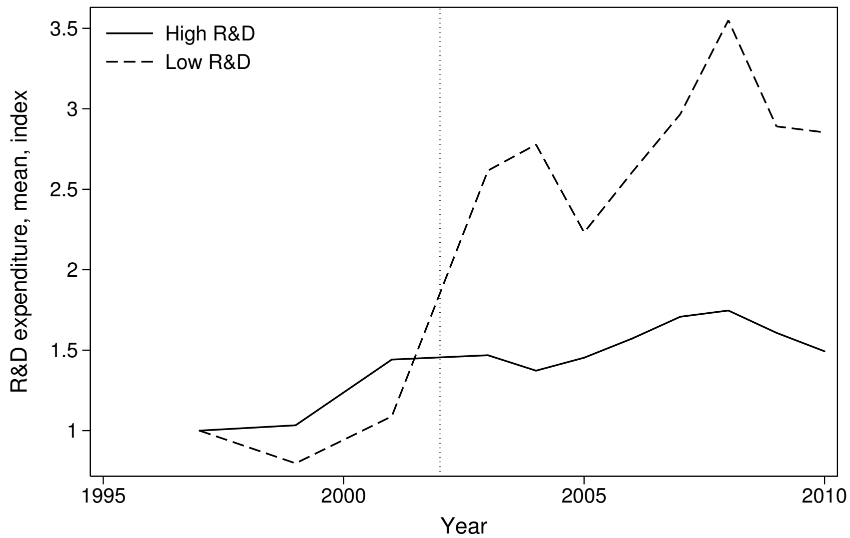
B Figures

Figure 1: Skill premium in Norway.



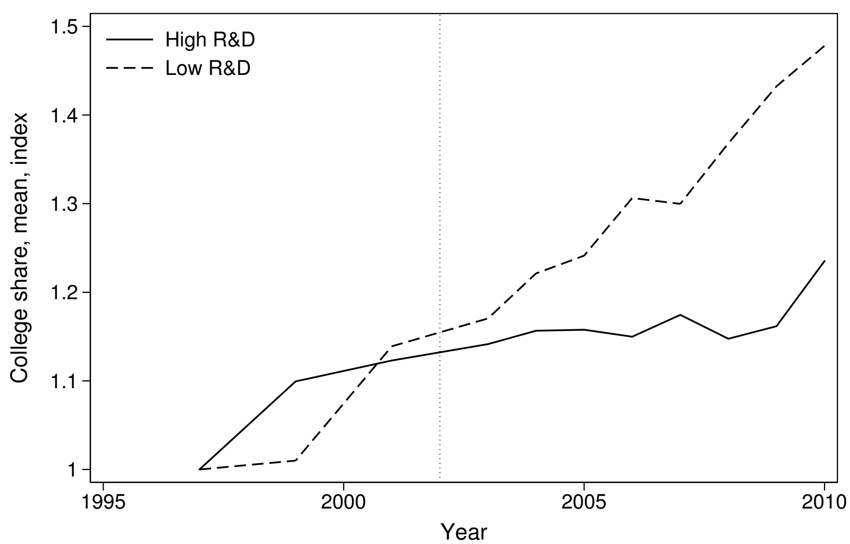
Source: Statistics Norway. Author's own calculations. Skilled workers refer to workers with at least some college eduation, while unskilled workers have less. This definition mirrors the one used in the empirical analysis of the paper. The skill premium is simply average wages of skilled and unskilled workers.

Figure 2: R&D investment. Index, 1997=1.



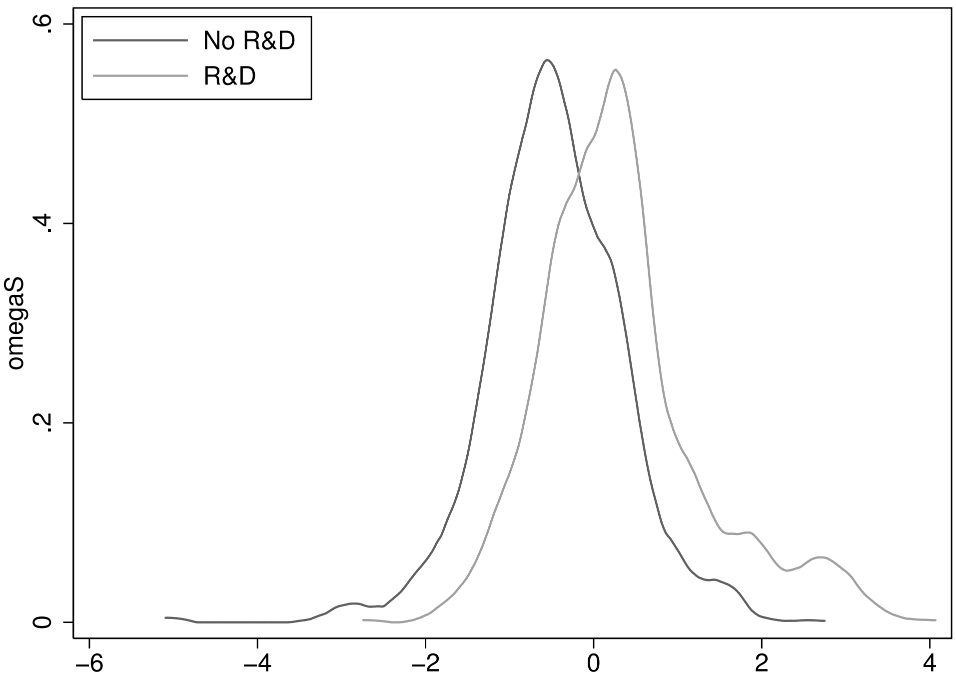
A firm is classified as a treated firm if pre-reform R&D expenditures is less than 4 million NOK, and as a control firm if it exceeds that. The averages have been normalized to 1997 level of R&D expenditures for each group separately.

Figure 3: College Share. Index, 1997=1.



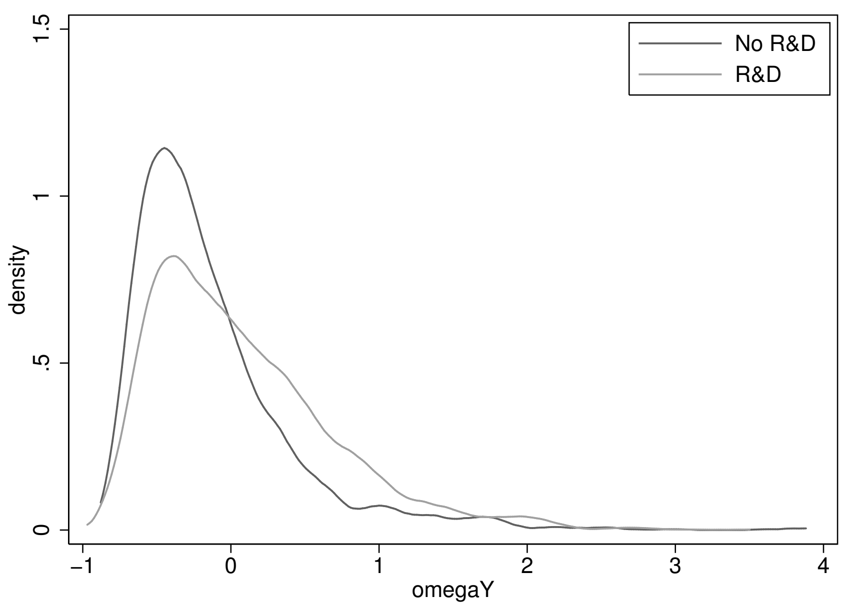
A firm is classified as a treated firm if pre-reform R&D expenditures is less than 4 million NOK, and as a control firm if it exceeds that. The averages have been normalized to 1997 level of college shares for each group separately.

Figure 4: Skill-biased productivity differences.



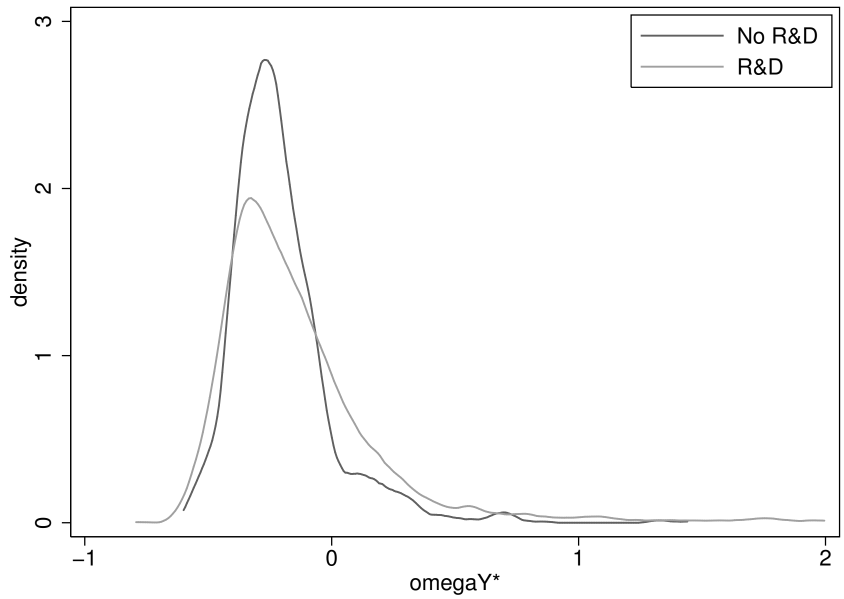
A firm is classified as an R&D firm in period t if has positive R&D expenditures in that year.

Figure 5: Hicks-neutral productivity differences – w/o skill-biased productivity.



A firm is classified as an R&D firm in period t if it has positive R&D expenditures in that year.

Figure 6: Hicks-neutral productivity differences – w/ skill-biased productivity.



A firm is classified as an R&D firm in period t if it has positive R&D expenditures in that year.

C R&D Reform Details

The definition of R&D used for eligibility closely follows the *Frascati Manual*, the internationally recognised standard for collecting R&D statistics. The aim of the projects has to be to "develop a new or improved product, service or production process". The R&D survey includes: (i) all firms with at least 50 employees; (ii) all firms with less than 50 employees and 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; (iii) among other firms with 10-49 employees a random sample was selected within each strata (NACE 2-digit and size class).

Roughly 70 percent of all applications are approved. See Cappelen et al. (2010) for more details and an evaluation of the reform.

D Data

The main source of employment and wage data for the period 1996 to 2010 is the employee register (AT) which holds annual records of worked hours and earned wages at the individual level. Statistics Norway links this register to the tax office database (LTO) to create a correspondence between the wage reported by the employer and those reported to the tax authorities by the individual. This joint file (ATmLTO) is a much cleaner data set and is therefore used instead of the AT register. In addition to wages at the person-firm-year level, the database includes the first and the last date of the employment spell within a given year, the total number of days worked, the municipalities in which the individual lives and works, and an indicator for full-time and part-time employment. The ATmLTO data is then merged with time-varying demographic information about years of education, gender and the number of children, also from Statistics Norway.

D.1 Sample Selection

Starting with the raw employee data, I pick the longest employment spell for each worker-year combination. For that year, I assign to the worker the wage and the firm of that spell.[43] Daily wages are measured as the wage earned in that spell, divided by the number of days worked in that spell. I further restrict the sample to individuals who have worked for at least three months during a year, are between 19 and 67 years old, and have at least one year of potential labor market experience. Only full time employees are used, to avoid measurement error in number of hours worked, as well as biases related to possible part time wage penalties.

To remove outliers, I predict wages based on a simple Mincer regression of log wages on education, experience and experience squared and remove observations that lie outside five times the standard error of the residual. The correlation between the number of workers matched to each firm and the reported number of workers by the firm is 0.88. For some observations, there is a large discrepancy between the number of worker being matched to the firm and the firm's reported number of workers, even before the data is cleaned. This is likely due to data being collected at different times of the year for the different data sources, which results in some firms showing up in the firm (or worker) data despite in reality being shut down. I discard those observations from my sample.

D.2 Variable description

- *Wage.* Individual wages are computed as daily wages, wage in the spell divided by the number of days worked in that spell. See Section D.1 for details.

- *Education.* Education is measured as the number of years of education.

[43] For most workers, this is the only employment they have in that year. For workers who switch firms, they will be counted as switchers only in the next year if they work less than half the year at the new firm.

- *College.* A worker is classified as College educated if the worker has 14 or more years of education.
- *Non-college.* A worker is classified as Non-college educated if the worker has less than 14 years of education.
- *Skilled labor.* The number of workers in a firm that are classified as College.
- *Unskilled labor.* The number of workers in a firm that are classified as Non-college.
- *Experience.* Experience is constructed from pensions data from 1967 and onwards. Experience is calculated as the number of years in which the individual had income above 1000 NOK (roughly 120 USD), including as entrepreneurs.
- *Value added.* Taken from the Capital database, deflated using an industry-specific price index.
- *Labor productivity.* Labor productivity is defined as value added per worker.
- *Size* Firm size is defined as the log of the number of employees.
- *Capital.* Capital is measured as the value of tangible fixed assets excluding buildings and land, in current prices, at replacement rates. It is then deflated using an industry-specific price index.
- *R&D expenditures.* The definition of R&D expenditures closely follows the *Frascati Manual*, the internationally recognized standard for collecting R&D statistics. See Section C for details.
- *R&D intensity.* R&D expenditures over operating income.

E Additional tables

Table 11: Reduced form evidence of increased skill demand – placebo.

Outcome	Linear College share	Linear log College share	Linear log College ratio
$\geq 2002 \times$Treated	0.000	-0.032	-0.021
	(0.007)	(0.033)	(0.044)
Firm FE	Yes	Yes	Yes
Industry + Year FE	Yes	Yes	Yes
Observations	660	660	660

Standard errors in parenthesis are clustered on firm. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 12: Reduced form evidence of increased skill demand – removing R&D workers.

Outcome	OLS log College ratio	FE log College ratio	2SLS log College ratio
RD_{t-1}	0.189^{***}	0.036^{**}	0.960^{***}
	(0.043)	(0.016)	(0.299)
Firm FE	No	Yes	Yes
Industry + Year FE	Yes	Yes	Yes
Observations	3892	3892	3892
Outcome		**First stage RD_{t-1}**	**Reduced form log College ratio**
$\geq 2002 \times \text{Treated}_{t-1}$		0.172^{***}	0.160^{***}
		(0.029)	(0.044)
KP F stat		35.28	
Firm FE		Yes	Yes
Industry + Year FE		Yes	Yes
Observations		3892	3892

Standard errors in parenthesis are clustered on firm. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 13: Reduced form effect – various outcomes.

	(1)	(2)	(3)
Outcome	Δ log Output	Δ log Capital	Δ log Size
$\geq 2002\times$Treated	0.069	-0.024	0.045
	(0.076)	(0.048)	(0.041)
Firm FE	Yes	Yes	Yes
Industry + Year FE	Yes	Yes	Yes
Observations	3967	4156	4156

Standard errors in parenthesis are clustered on firm. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 14: First step estimates – robustness.

	(1)	(2)	(3)	(4)	(5)
	W/o R&D	W/o R&D	Industry Wage	Local Wage	Adj. Wage
σ_L	68.382	.	1.510***	1.431***	1.672***
	(74.200)	.	(0.296)	(0.277)	(0.315)
$\alpha_1(\omega_{Sjt-1})$	0.594***	0.893***	0.950***	0.901***	0.947***
	(0.030)	(0.012)	(0.010)	(0.007)	(0.007)
$\tilde{\alpha_2}$ (RD_{jt-1})	0.029***	0.097***	0.031***	0.032***	0.025**
	(0.006)	(0.011)	(0.010)	(0.011)	(0.011)
Industry FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
Observations	4133	4133	4169	4169	4169

In column (1), the number of workers denoted as R&D workers is subtracted from the college workers. The same exercise is repeated in column (2), imposing an elasticity of substitution of 1.35. In column (3), an adjusted measure of the firm-level skill premium taking out worker fixed effects is used as an instrument. In column (4), the local labor market skill premium is used as an instrument. In column (5), the industry-specific skill premium in used as an instrument. Standard errors in parenthesis are clustered on firm. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 15: Elasticity of substitution – robustness.

	(1) < 2005	(2) ≥ 2005
σ_L	1.326	1.365
	R&D	No R&D
σ_L	1.422	1.259

F FOR ONLINE PUBLICATION

F.1 Cobb-Douglas Empirical Strategy

In addition to the baseline specification, I estimate a Cobb-Douglas production function. This holds several advantages over other functional forms, as well as some drawbacks. The main limitation is that it restricts the elasticity of substitution to equal one across the factors of production. The main advantage of using a standard Cobb-Douglas production function is that the estimation strategy follows the production function estimation literature. I rely on the Ackerberg et al. (2006) methodology, while at the same time allowing unobserved productivity to be shifted endogenously in the spirit of Doraszelski and Jaumandreu (2013). In the following subsection, I will lay out the estimation strategy in the Cobb-Douglas case.

Consider a standard Cobb-Douglas production function:

$$y_{jt} = x'\beta_0 + RDx'\beta_1 + \omega_{Njt} + \varepsilon_{jt}$$

where y_{jt} is log value added, ω_{Njt} is unobservable Hicks-neutral productivity, and ε_{jt} is an error term.[44] Vector x contains a constant, capital, high and low skilled wage bills (in logs), and I let the input coefficients vary with R&D status of firm by interacting the variables with a dummy that equals one if the firm engages in R&D at time t, and is zero otherwise. I again include the two levels of skill to be able to capture the potential skill complementarity of research activities. This shows where this particular functional form gives rise for concern: it restricts the elasticity of substitution

[44] I observe revenue and the materials bill instead of output and input prices and quantities separately, as usual with firm-level data. In these instances, it is standard in the literature to rely on a value added production function. For a recent critique of this approach, see Gandhi et al. (2013).

to be equal between all inputs, including the two types of labor. On the other hand, an advantage of using this specification is that I do not impose any restrictions on the coefficients in the estimations. In particular, I do not impose constant returns to scale, i.e. the coefficients do not have to sum to one.[45]

I use the approach of Levinsohn and Petrin (2003) of exploiting the demand for intermediate inputs to back out unobserved productivity. In addition, following Ackerberg et al. (2006), I assume that labor is a less freely variable input than materials, so that the demand for low skilled and high skilled labor enter into the intermediate input demand function. This is consistent with the presence of hiring costs, training periods or notice periods before the firm can fire workers. A profit maximizing (cost minimizing) firm will have the following demand for intermediate inputs: $m_{jt} = f_t(\omega_{Njt}, k_{jt}, h_{jt}, l_{jt}, RD_{jt})$.[46] Assuming that this relationship is monotonically increasing in productivity ω_{Njt}, it can be inverted and unobserved productivity can be expressed as:

$$\omega_{Njt} = f_t^{-1}(m_{jt}, k_{jt}, h_{jt}, l_{jt}, RD_{jt}) \tag{22}$$

The estimation procedure is done it two steps, each outlined below.

F.1.1 First step

In the first part of this two-step method, I use equation (22) to insert for ω_{Njt} in the production function:

$$y_{jt} = x'\beta_0 + RDx'\beta_1 + f_t^{-1}(m_{jt}, k_{jt}, h_{jt}, l_{jt}, RD_{jt}) + \varepsilon_{jt} \tag{23}$$

This is the first step equation, which is estimated separately for each 2-digit NACE 2 digit industry using OLS. Year dummies are included to absorb any common time-varying shocks. Again following the literature, the control function term $f_t^{-1}(m_{jt}, k_{jt}, h_{jt}, l_{jt}, RD_{jt})$ is approximated by a second order polynomial expansion in m_{jt}, k_{jt}, h_{jt}, l_{jt} and RD_{jt}. From this I obtain an estimate of the composite function:

$$\Phi_{jt} = x'\beta_0 + RDx'\beta_1 + f_t^{-1}(m_{jt}, k_{jt}, h_{jt}, l_{jt}, RD_{jt}). \tag{24}$$

As in Ackerberg et al. (2006), no coefficients are identified in the first step. It is only used to separate the (to the firm) predictable and observable productivity shock term ω_{Njt} from the unpre-

[45] This is in line with most of the literature estimating Cobb-Douglas production functions. See for instance Pavcnik (2002), Fox and Smeets (2011) and Akerman et al. (2015).

[46] Note that there is an implicit assumption about inputs being homogeneous and traded in a perfectly competitive market.

dictable error term ε_{jt}, and to eliminate this pure error term from the second stage.

F.1.2 Second stage

In the second stage, I again evoke the Markov assumption common to the production function estimation literature, and I allow the level of productivity ω_{Njt} to be endogenously affected by the firm's investment in R&D activities. Unobservable Hicks-neutral productivity then evolves according to a similar process as skill-augmenting productivity does in the main specification:

$$\omega_{Njt} = g(\omega_{jt-1}, RD_{jt-1}) + \xi_{jt},$$

i.e., this period's Hicks-neutral productivity is some function of the previous period's productivity, previous period's R&D investments, and an error term ξ_{jt}. This error term is a random shock to productivity that is not anticipated by the firm and hence is uncorrelated with all the inputs. It represents the uncertainty inherent in the R&D process and any other unexpected shocks to the productivity of a firm, e.g. unanticipated machinery downtime and weather shocks.

I again assume that this relationship is linear:

$$\omega_{Njt} = \alpha_0 + \alpha_1 \omega_{Njt-1} + \alpha_2 RD_{jt-1} + \xi_{jt}. \tag{25}$$

By rewriting equation (24), unobserved productivity can be written as:

$$\begin{aligned}\omega_{Njt} = \hat{\Phi}_{jt} - \beta_{k0}k_{jt} - \beta_{h0}h_{jt} - \beta_{l0}l_{jt}\\ -\beta_{k1}RDk_{jt} - \beta_{h1}RDh_{jt} - \beta_{l1}RDl_{jt},\end{aligned}$$

where β_{l0}, β_{h0} and β_{k0} are the coefficients on low skilled labor, high skilled labor and capital in the absence of R&D activities in the firm, while β_{l1}, β_{h1} and β_{k1} measure the change in these coefficients in the presence of R&D in the firm. If these last three coefficients are significantly different from zero, it is indicative of R&D activities changing the output elasticities of the inputs at the firm level.

Insert for ω_{Njt} and ω_{Njt-1} in Markov process (equation (25)) to derive the second stage estimating equation:

$$\begin{aligned}\hat{\Phi}_{jt} = \alpha_0 + \beta_{k0}k_{jt} + \beta_{h0}h_{jt} + \beta_{l0}l_{jt} + \beta_{k1}RDk_{jt} + \beta_{h1}RDh_{jt} + \beta_{l1}RDl_{jt}\\ +\alpha_1(\hat{\Phi}_{jt-1} - \beta_{k0}k_{jt-1} - \beta_{h0}h_{jt-1} - \beta_{l0}l_{jt-1}\\ -\beta_{k1}RDk_{jt-1} - \beta_{h1}RDh_{jt-1} - \beta_{l1}RDl_{jt-1}) + \alpha_2 RD_{jt-1} + \xi_{jt}\end{aligned} \tag{26}$$

I estimate this equation using GMM. Current period's capital, k_{jt}, is orthogonal to the error term by definition. The same holds for all variables pertaining to period $t-1$: Φ_{jt-1}, l_{jt-1}, h_{jt-1} and RD_{jt-1}. Current period's skilled and unskilled labor are, on the other hand, not orthogonal to this period's unpredictable shock, and are therefore instrumented with lagged values. Importantly, current period R&D activities are also not exogenous. I follow two strategies for instrumentation of these variables. In the first, I follow the traditional approach using lagged values as instruments. Since lagged values are assumed to be orthogonal to this period's random shock ξ_{jt}, they are valid instruments.

As an alternative approach, I use the R&D reform as a source of exogenous variation in R&D activity. This approach breaks the link between productivity and R&D investments that will continue to bias the results if there is serial correlation in the error term ξ_{jt}. The instrument is a dummy variable which equals one for the treated firms in the years after the reform was introduced. Following Wooldridge (2010), I instrument for the interaction between the R&D dummy and the various inputs using the interaction between the input in question and the treatment term $(T_{1j} \times post_t)$.

To summarize the estimation procedure: The first step consists of purging the production function of measurement error and unpredictable shocks to production by eliminating ε_{jt}. In the second stage, I can then compare output elasticities of equally productive firms who differ in R&D behavior. The control function breaks link between productivity and input demand, while the exogenous variation from the reform breaks link between productivity and R&D behavior. Results from estimating the second equation (26) are reported from a one-step GMM estimator using equal weights for every moment, while a two-stage estimation is conducted to perform specification tests.

F.2 Results

Column (1) of Table 16 shows the results of a very basic production function estimation. Note that the coefficients have not been restricted to impose constant returns to scale, yet the coefficients sum up remarkably close to 1. The results are very much in line with results from the literature, indicating that the firms in the sample are similar to those studied previously, which is important for external validity.[47] In addition, last period's R&D has a positive and significant impact on current period's productivity, as expected. The estimate translates into an 18 percent increase in short run productivity for firms from perfoming R&D. In column (2), lagged productivity is added to address the simultaneity bias discussed above. The resulting change in the R&D coefficient is dramatic. Without this inclusion, the coefficient on lagged R&D is vastly overestimated, attributing the vari-

[47] See for instance Ackerberg et al. (2006).

ation coming from differences in productivity to differences in R&D activities. This underscores how the correlation between R&D and productivity makes it crucial to separate these effects.

The results from my baseline specification is given in column (3). Here, the input coefficients are allowed to vary with R&D activity at the firm. In addition, I control for lagged productivity, and the lagged R&D dummy is added to the Markov process for unobserved productivity. Again, the coefficients sum up very close to 1 without any restrictions. The results confirm the finding from the CES specification employed in the main part of the paper that technological change indeed is skill-biased. Innovation increases the output elasticity of high skilled labor, while reducing the output elasticity of low skilled labor. These results are consistent with R&D resulting in production processes that complements high-skilled labor. There is also a positive and significant change in the capital coefficient. This results highly dependent on the type of capital included in the measure, and disappears if buildings are included. In column (4), the variation from the reform is not used. The results mostly have the same sign as in the baseline specification in column (3), although they are somewhat smaller in size.

Table 16: Cobb-Douglas production function estimates.

	Simple	W/ ω_{t-1}	Baseline	W/ reform
	(1)	(2)	(3)	(4)
β_{k0}	0.120***	0.100***	-0.059	0.033
	(0.032)	(0.024)	(0.074)	(0.077)
β_{l0}	0.544***	0.504***	0.750***	0.792***
	(0.123)	(0.044)	(0.107)	(0.095)
β_{h0}	0.366***	0.360***	0.192*	0.205**
	(0.034)	(0.034)	(0.109)	(0.101)
β_{k1}			0.271***	0.239***
			(0.111)	(0.117)
β_{l1}			-0.332***	-0.403***
			(0.118)	(0.109)
β_{h1}			0.244*	0.283**
			(0.144)	(0.140)
α_1 (ω_{t-1})		0.461***	0.414***	0.409***
		(0.111)	(0.071)	(0.071)
α_2 (RD_{t-1})	0.168***	-0.020	-0.024	0.121
	(0.056)	(0.026)	(0.029)	(0.095)
Industry FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Observations	3280	3254	3254	3254

Bootrapped standard errors in parenthesis, clustered on firm. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.