Motherhood and the Gender Productivity Gap*

Yana Gallen†

March 4, 2023

Abstract

Using Danish matched employer-employee data, I compare the relative pay of men and women to their relative productivity as measured by production function estimation. I find that the gender "productivity gap" is 8 percent, implying that almost two thirds of the residual gender wage gap is due to productivity differences between men and women. Motherhood plays an important role, yet it also reveals a puzzle: the pay gap for mothers is entirely explained by productivity, whereas the gap for non-mothers is not. In addition, the decoupling of pay and productivity for women without children happens during their prime-child bearing years. These estimates are robust to a variety of specifications for the impact of observables on productivity, and robust to accounting for endogenous sorting of women into less productive firms using a control-function approach. This paper also provides estimates of the productivity gap across industries and occupations, finding the same general patterns for mothers compared to women without children within these subgroups.

JEL Classification: J71, J31, J24
Keywords: Discrimination, Wage Gap, Labor Productivity

*Thank you to my advisor, Matthew Notowidigdo, and my committee members Lori Beaman and Jon Guryan for their very helpful advice and comments. I also thank David Berger, Stéphane Bonhomme, Matthias Doepke, Steven Durlauf, Martin Eichenbaum, Trevor Gallen, Brent Hickman, Jeff Grogger, Rune Vammen Lesner, John List, Lee Lockwood, Casey Mulligan, Seema Jayachandran, Rune Vejlin and participants at the Northwestern Applied Micro Reading Group and Applied Micro Seminar, ASU, Berkeley, the Chicago Federal Reserve, Conference on Labour Market Models and their Implications, Cornell, CeMENT, Harris School, H2D2 conference, DePaul, Duke, Nordic Data Workshop in Oslo, Purdue University, Stanford, UIC, and UIUC LER for helpful comments and discussions. Thank you to Henning Bunzel and LMDG, Department of Economics and Business, Aarhus University for access to the data. LMDG is a Dale T. Mortensen Visiting Niels Bohr professorship project sponsored by the Danish National Research Foundation. All mistakes are my own.

†Correspondence: University of Chicago Harris School of Public Policy, 1307 E 60th St, Chicago, IL 60637. Tel.: (773) 834-2784. Email: yana@uchicago.edu.

1 Introduction

Women earn less than men and the gender pay gap expands when women become mothers (Angelov et al. (2016), Kleven et al. (2019)). This paper asks how much of the wage gap between men and women, and particularly between men and women with children, is explained by measurable productivity differences. Using Danish administrative data, I study how firm output varies with the gender and parenthood status of employees. I find that about 8 percentage points of the 12 percent (residual) gender pay gap can be explained by productivity differences between men and women. Productivity explains the entire pay gap associated with motherhood.

To measure the productivity gap, I use Danish data that matches worker characteristics with firm accounting information. I estimate a firm-level production function that takes labor, material goods, and capital as inputs and treats male and female labor units as perfect substitutes. The gender productivity gap is the number of efficiency units lost if a worker is female, holding other variables such as age, education, experience, and hours worked constant.

I find a sizable productivity gap of eight percent. However, this average masks differences between women over the lifecycle, particularly differences for mothers compared to non-mothers. For mothers, I find that the earnings gap is approximately equal to the productivity gap, suggesting that there is little or no discrimination in the form of uncompensated output against mothers. This is consistent with the literature arguing that the wage gap occurs primarily for women with children who work fewer and more flexible hours than their male counterparts (see for example Goldin (2014), Gicheva (2013), Angelov et al. (2016), and Kleven et al. (2019)) and that there may be some output loss associated with these work arrangements.

At the same time, there is evidence of uncompensated productivity among women without children. Although the wage gap is smaller for women without children, women without children are actually more productive than similarly aged men. In addition, I find that the disparity between wages and productivity for non-mothers happens especially during their prime child-bearing ages. After age 40, there are no meaningful differences in the relative productivity of mothers and non-mothers. Discrimination in the form of uncompensated productivity is largest in the group with a smaller residual pay gap: young non-mothers.[1]

This age path suggests that the pay gap between men and women without children but of childbearing age may be due to statistical discrimination: if productivity falls with motherhood but employers cannot lower wages when women give birth, then employers may offer lower wages to productive women in anticipation of motherhood. Alternatively, women who will have children soon may be unwilling to ask their employer for a raise when they know they will have children soon. Consistent with this, I find that women who will

[1]Differences between pay and productivity by gender/parenthood could arise from gender differences in worker preferences, from differences in non-wage compensation, for structural reasons, or due to firm-level statistical or taste-based discrimination.

have children soon are more underpaid relative to their productivity than women who have no children and will not have children soon. I also find that women who look, based on observables, the most like future mothers experience the largest gaps between pay and productivity. Mothers do not experience gaps between pay and productivity. Among women without children in their prime childbearing years, the ones with no pay-productivity gap are not those who look like mothers, but rather those who will never have children in the future. An additional piece of evidence consistent with statistical discrimination by employers is the fact that conditional on productivity, pay is lowest for women without children in those occupations with the longest average tenure, measured by the typical years with a given firm among male employees. This suggests that when employers expect workers to stay at the firm for many years, there is larger underpayment of women without children. Taken together, these data suggest that compensation is tied to the expected productivity of workers over a time interval which may include large productivity fluctuations.

I present estimates of the productivity gap in the cross-section, over time, by industry, and accounting for selection of workforce composition based on unobservables using a control function approach which I discuss in Section 3. The control function approach does not change the overall estimate of the relative productivity of men compared to women. Consistent with the small role for selection in estimating relative productivity via the production function, using a wage decomposition as in Card et al. (2016), I also find little evidence that women work in lower wage firms within this subset of relatively large, private sector Danish firms.

This paper contributes to an extensive literature on the sources of the gender pay gap (see Altonji and Blank (1999), Olivetti and Petrongolo (2016), and Blau and Kahn (2017) for an overview).[2] Whether mothers' lower pay reflects preferences or discrimination is an important unanswered question. In the Danish setting, Kleven et al. (2019) use an event study around the time of childbirth and find that mothers (relative to fathers) experience a permanent and substantial decrease in earnings growth explained in equal parts by a reduction in LFP, hours worked, and in wage rates. The authors find that, relative to other observables, the role of motherhood in explaining the wage gap has doubled since the 1980s. I study the importance of productivity difference between mothers and other workers as a potential explanation of the pay gap. The results suggest that while mothers indeed are paid less than similarly skilled men, this difference in pay is completely explained by productivity differences.

The literature linking relative productivity to relative pay was started by Hellerstein et al. (1999) which

[2] Polachek (1981) argued that the constraints of childrearing generated sex segregation of women into occupations where human capital did not depreciate as quickly even before children are born. Mulligan and Rubinstein (2008) study the changing nature of female selection into the labor force from the 1970s to the 1990s finding that most of the apparent narrowing of the gender wage gap resulted from compositional changes in the female labor force. Gayle and Golan (2011) estimate the impact of the changing labor force on wages through the lense of an adverse selection model but find that labor market experience is the dominant determinant of wages. Since the 1990s, however, the gap has stagnated even as women's education relative to men's has increased leading to an interest in understanding what the "final" sources of the pay gap are (Goldin (2014), Blau and Kahn (2017)).

studies the relationship between wage gaps and gaps in marginal product for a variety of observable characteristics in the manufacturing industry in 1990. The authors find that with the exception of gender, differences in wages based on observables are equal to differences in marginal productivity. This paper differs from Hellerstein et al. (1999) primarily by introducing data on motherhood and its relationship to the productivity gap. This paper additionally builds on the work of Hellerstein et al. (1999) by using more recent data which cover a wider set of industries, and incorporates detailed differences in the occupations held by men and women in estimates of the productivity gap. As discussed by Blau (1977), and revisited in the case of the US by Blau and Kahn (2017) and in the case of Denmark by Gallen et al. (2017), an important component of the pay gap is occupational choice.[3] Women may prefer working in low-wage occupations because these jobs allow fewer and more flexible hours (Goldin, 2014; Wasserman, 2017). Conditional on occupational choice, this paper investigates whether and where gaps in productivity emerge. In addition, this paper incorporates the role of sorting based on firm unobservables using the Olley and Pakes (1996) control function approach. Card et al. (2016) find that women work in lower pay-premium firms compared to men in Portugal. This paper investigates the degree to which this is true in Denmark and empirically corrects estimates for the potential role of selection by women into less productive firms.

In the same spirit as this study, Azmat and Ferrer (2017) document the difference between hours billed and new clients brought in to the firm for male vs. female lawyers. They find that the large differences in the earnings of male and female lawyers (particularly mothers) are largely explained by these measures of productivity. Cook et al. (2018), studies the choices and experience profiles of men compared to women which explain the seven percent gender pay gap among ride-share drivers. The pay of ride-share drivers is set by an algorithm, so they are studying a setting in which taste-based discrimination is not possible. Finally, Adams-Prassl (2023) studies the online labor market and finds a gender earnings gap of 20 percent which is primarily explained by the interruptions faced by women with young children. This consistent evidence from a diverse set of occupations—from Uber drivers to online workers to lawyers—sets the stage for the broader, cross-industry analysis in the present paper.

This paper is the first to investigate the role of motherhood in driving the gender productivity gap using production function estimation. This has important implications for a literature which is increasingly

[3] There is a large body of literature documenting the differences between women and men which may explain the gender wage gap, but are more subtle than differences in human capital accumulation, child-rearing, and occupational choice. As reviewed by Niederle and Vesterlund (2011), women have been documented in both the lab and the field to be less competitive than men, conditional on performance. This competitiveness factor has been studied extensively in recent years. See for example Bertrand (2011), Croson and Gneezy (2009), Flory et al. (2015), Buser et al. (2014), Apesteguia et al. (2012), Markussen et al. (2014), Kamas and Preston (2012), Berge et al. (2015), Zhang (2015), and Reuben et al. (2015). Gneezy et al. (2009) argue that this link between gender and competition is reversed in a matrilineal society, implying that most of the link is driven by cultural rather than biological differences between men and women. Ichino and Moretti (2009) and Rockoff and Herrmann (2012) debate whether women are differentially absent from work due to menstruation. Small biological difference may turn into large differences in career pathways when mediated by social norms (see for example Fryer and Levitt (2010) and Bursztyn et al. (2017)).

focused on motherhood as an important turning point in the careers of women (Kleven et al. (2019), Angelov et al. (2016)).[4] The role of motherhood in explaining the gender productivity gap not only confirms that motherhood is associated with changes in the work women do, but also highlights a new problem: why are non-mothers still underpaid compared to men?

Women's wages may be lower than men's both due to statistical discrimination (if they are indeed less productive than men on average) as described in Aigner and Cain (1977) or due to taste-based discrimination as described in Becker (1971).[5] Gunderson (1989) discuss the very different policy implications of statistical and taste-based discrimination. I focus on identifying one particular type of phenomenon: gender differences in pay unexplained by differences in output. Such a gap would occur if, for example, women did not bargain as well as men for raises.[6] Another possibility would be that firms did not pass improvements in productivity on to female employees as much as male employees, as Black and Strahan (2001) demonstrate is the case for banks during a period with heavy barriers to entry and as suggested in Card et al. (2016) in the case of Portugal. This leaves out many important potential drivers of the labor market differences between men and women. For example, if women are not offered jobs at high productivity firms, or if women are not invested in or offered promotions despite being equally able to work in more demanding jobs (see for example Thomas (2015), Stearns (2017), Albrecht et al. (2015), and Albrecht et al. (2003)). This "mommy tracking" is difficult to distinguish from preferences, but may occur if many firms have a distaste for hiring women, or if firms are sufficiently risk averse and the distribution of female productivity differs from that of male productivity. Hiring/promotion based discrimination, both interesting and important, is not addressed here. Instead, I focus exclusively on the link between realized output and pay.

The paper proceeds as follows: Section 2 describes the data used in estimation. Section 3 provides the model and estimating equations. Section 4 presents results, and Section 5 concludes.

2 Data

The data used in this paper come from three primary sources: a relatively new Danish register on employees called eIncome, a more commonly used Danish register on employees called IDA, and a detailed survey of firm accounts, called Regnskabsstatistikken (abbreviated FIRE). eIncome is register data covering all employees

[4] The discussion of why motherhood might be associated with different career choices for men relative to women has a long history. Decades ago, Becker (1985) hypothesized that differences in demands on or abilities of women in the home production sector translate to differences in career choice. Hersch and Stratton (1997) provide early evidence of the effect of housework on market wages. The importance of finding flexible work arrangements to accommodate childrearing is emphasized in Goldin (2014) and Blau and Kahn (2013).

[5] For evidence of taste-based discrimination, see Weber and Zulehner (2014), Hellerstein et al. (2002), Heyman et al. (2013), Pan (2015), and Goldin and Rouse (2000).

[6] See a survey by Babcock and Laschever (2003) who find that women are 50 percentage points less likely than men to negotiate their starting salaries or Leibbrandt and List (2015) for field evidence.

working in Denmark, from 2008. The data is reported monthly, by all employers to the Danish Customs and Tax Administration, who pass the data to Statistics Denmark's eIncome Register to be used in calculation of national statistics at the monthly level. The primary advantage of this dataset is that it reports work by all employees in a given firm: their occupation, total pay for that month, and total hours worked in the month (as well as the dates of employment).

This dataset is distinct from the commonly used Danish IDA dataset, which is annual and has data on payments and hours based on the worker's status in November of that year. The hugely improved hours variables in eIncome shrink the wage gap considerably. In particular, the gap falls from 16 percent to ten percent. The main improvement is better tracking of workers who are not continuously employed in a firm and enter and exit the data only for a few months. In the IDA dataset, using bracketed hours worked and a November employment measure, it is difficult to properly assign total hours worked at a given firm. Bracketing alone accounts for a two percentage point increase in the pay gap. Of course, there is some non-response even in the eIncome dataset—about 15 percent of the hours data is imputed. All employees are included regardless of hours worked. Main jobs and side jobs are included. Employees who are not residents of Denmark are included. If an employee doesn't have pay for up to 45 days at a job, but subsequently returns to the same employer for pay (for example for training), he or she is included in the data for the months without pay.

The hours worked measure in IDA is based on employee contributions to retirement benefits. The brackets are four bins of weekly hours (0-8, 9-17, 18-27, 27+) or four bins of monthly hours, (0-38, 39-77, 78-116, 117+). The data also measure the fraction of the year worked. There is a large fraction of workers whose hours are not distinguished from one another but may in reality differ substantially. eIncome is not completely immune to this problem, though it is certainly less severe. In eIncome, salaried workers would be listed as working 37 hours a week, unless they clocked in overtime hours. Many likely do not, and work slightly less or slightly more than 37 hours per week. There is no reason to think this is orthogonal to gender—women work fewer hours on the margins we can measure, they may also work fewer hours on the margins we have more difficulty measuring. Indeed US time-use data (the American Time Use Survey) suggests that conditional on working full-time, mothers of older children work about one hour less per day than fathers and mothers of young children work about 40 minutes less per day than fathers.

The eIncome register can be linked with data on firm value added using a dataset called FIRE which contains information on firm accounts. The FIRE employer data is the basis for national accounts. As in Mortensen et al. (2010), I follow the methodology for constructing value added and capital stock used in national accounting. The details of this procedure exactly follow Mortensen et al. (2010) and are discussed in the Data Appendix. FIRE includes information on firms from tax records (such as revenue and the value of capital) and also contains detailed accounting measures from surveys. Firms are surveyed based on size. Firms with more than 50 employees are surveyed annually, firms with 20-49 employees are surveyed every other year, firms with 10-19 employees are surveyed every 5th year, and firms with 5-9 employees are surveyed every 10th year.

Firms which are not in the survey in a given year have some of their information imputed into the dataset, though much of the imputation comes from tax records. Detailed information on the cost of intermediate goods, however, is completely imputed for a large fraction of firms in the data. My measure of value added is revenue less the cost of these intermediate inputs so the measurement error generated by using imputed values is on the left hand side and does not systematically bias results. When information is imputed, it is based on industry-level averages weighted by employment and revenue. In the results reported, I use all data which was not completely imputed (that is, data taken from tax records combined with survey results). About 9,000 firms are actually surveyed in each year.[7]

To supplement the worker-level information available in the income registers merged with FIRE—which is essentially occupation, hours worked, wages, and industry—I add demographic information on workers from IDA. This includes information on the birthdate of all children, year of marriage, education including major, age, and experience (which is constructed as the sum of hours worked in the labor market according to IDA).

2.1 Summary statistics

The earnings gap in Denmark is surprisingly similar to the gap in the US. Table 1 below provides estimates of the earnings gap in the US from Goldin (2014) compared to a similar population in Denmark and compared to my restricted sample of large industries in the FIRE database. The raw earnings gap is smaller in Denmark than in the US but it also is less explained by controlling for hours, education, and occupation. The smaller raw gap is consistent with Blau and Kahn (2003) who find that countries with more compressed wage distributions (such as Denmark) have smaller wage gaps.

The Denmark and US samples are restricted to ages 25-64. In Denmark, the raw gap is 27.7 log points, compared with 32 log points in the US. Controlling for age, hours worked, education, and occupation, the gap falls to 17.2 log points, compared with 19.1 log points in the US. The R-squared from the last wage regression in the US is about twenty percentage points lower than in Denmark. The lower R-squared in the US may reflect noise expected from survey data. In addition, unions and collective bargaining determine

[7] Restricting only to the set of firms surveyed in detail about their accounts, I estimate the relative productivity of women is 0.94, which is about 1-2 percentage points higher than my baseline estimates and not a statistically significant difference. The cost of using only firms that were actually surveyed is that I would not have power to study differences across industries, across occupation, by age, etc.

Table 1: The pay gap in Denmark vs. US

Sample	Variables included	Coefficient on female	Standard error	R^2
US	Basic	-0.320	0.0010	0.102
US	Basic, time	-0.196	0.0009	0.353
US	Basic, time, education	-0.245	0.0008	0.475
US	Basic, time, education, occupation	-0.191	0.0010	0.563
Denmark (FIRE)	Basic	-0.277	0.0011	0.095
Denmark (FIRE)	Basic, time	-0.193	0.0006	0.727
Denmark (FIRE)	Basic, time, education	-0.200	0.0006	0.750
Denmark (FIRE)	Basic, time, education, occupation	-0.172	0.0006	0.781

Dependent variable is log earnings. The sample is 2009 to 2011 All regressions include a quadratic in age and year dummies. US regressions also include race. Hours controls are added in the second regressions. Hours are bracketed in Denmark (see see the data appendix) and indicate hours per week and weeks per year in the US. Education indicates primary, high school, or more advanced schooling in Denmark, and similar groups in the US, and is added in the third row. Occupation dummies at the 3 digit level are added in the final row. Goldin's ACS sample includes only individuals ages 25-64. For comparison, I restrict to only these ages in the FIRE sample. The number of observation is 3,291,168 in the US, and 2,879,216 in the restricted FIRE sample.

wages to a greater extent in Denmark than the US. However, collective bargaining has become increasingly decentralized and wage dispersion has commensurately increased (Dahl et al. (2013)). In 2000, only 15 percent of the population has wages determined completely by collective bargaining, with no firm-level negotiation. In contrast, over the 1990s, the fraction of workers whose wages are set completely outside the collective bargaining system grew from 5 to 20 percent. For the remainder of workers, collective bargaining sets minimum wages which are binding only for relatively low-skilled workers. For these workers, wage increases resulting from collective bargaining are determined by tenure and education (see Dahl et al. (2013) for a detailed description the these patterns and of wage bargaining in Denmark).

One advantage of the Danish register data compared with the American Community Survey in the US is that it provides information on the experience of a worker and also on the firm ID of the worker. Earnings may depend on experience (and women who take time off work to have children may have a different level of experience than men on the same age). Manning and Robinson (2004) argue that this difference in experience explains much of the gender pay gap in the British Household Panel Survey. Earnings may also vary by firm for observationally identical workers. This may reflect differences in non-wage compensation at different firms and in the presence of gender sorting may explain some of the earnings gap. In Table 2 below, I report the results of a regression of log earnings on hours, a quadratic in age, and sequentially add controls for (1) a quadratic in experience and education level dummies, (2) occupation and industry fixed effects (at the

3-digit ISCO level and the 2-digit NACE level, respectively), and (3) the interaction of firm fixed effects and occupation fixed effects.

Table 2: Conditional pay gap 2000-2011

	(1)	(2)	(3)	(4)
Female	-0.1740	-0.1628	-0.1393	-0.1184
	(0.0002)	(0.0002)	(0.0003)	(0.0003)
Experience	N	Y	Y	Y
Occupation, industry FE	N	N	Y	Y
Firm× occ FE	N	N	N	Y
R-squared	0.8430	0.8442	0.8562	0.8924
N	15613056	15613056	15613056	15613056

Dependent variable is log earnings. All regressions include hours and year controls, a quadratic in age, and education level dummies. Experience indicates a quadratic in experience (measured as hours of employment). Occupation is at the 3 digit ISCO level. Standard errors in parentheses.

Adding controls available with the rich Danish data, such as experience and occupation only causes the wage gap to fall slightly. Adding firm and occupation interactions and identifying the earnings gap using differences in the pay of women and men within a firm in a given occupation narrows the earnings gap to just under 12 percent on average in the years 2000-2011.

I focus my analysis on the six industries (measured at the two digit level) which have the largest number of firm-year observations in the FIRE database: Accommodation and food services, Construction, Manufacturing, Wholesale and retail trade, Other community, social and personal services, and Real estate, renting, and business activities. These make up more than 50 percent of the Danish economy and 98 percent of non-imputed firm-year observations.[8] Table 3 below provides some summary statistics for the firms in each industry and the dataset overall.

The wage gap varies by industry, ranging from 11 percent to 19 percent. The fraction of the workforce in a given industry which is male also varies. In construction, a very large fraction of the labor force is male, while in other services, only half of workers are male. Notably, this study of productivity differences is focused on industries with relatively more men than average. Because there are no accounting statistics for public sector firms, this large portion of the Danish economy (and place of employment for women, disproportionately) is omitted from the analysis. The potential biases from this omission will be discussed in the next section.

[8]Measured by 2010 gross value added by industry tables available from Statistics-Denmark.

Table 3: Cross-industry summary statistics

	Accom./food	Constr.	Manuf.	W/R trade	Other serv.	Real est.
w^f/w^m	0.8808	0.8083	0.8331	0.8418	0.8737	0.8485
Fraction male	0.5911	0.8992	0.6919	0.6782	0.5152	0.5935
Firm size	6.50	5.65	10.04	7.15	7.18	5.28
N	298370	135808	72000	69215	11799	18920

This table provides summary statistics on variables of interest across industries. w^f/w^m is the average wage gap control for quadratics in age and experience, education level, occupation fixed effects at the 3-digit ISCO level, hours worked, and year. Wage regressions are run by industry. The fraction men are averages measured at the person level. Firm size is measured treating the firm as the unit of observation.

3 Model and estimation

The goal of this paper is to understand whether differences in the productivity of men and women, conditional on a set of observable characteristics, explain differences in the wages of men and women, conditional on the same set of observable characteristics. To answer this question, I estimate production functions assuming output is affected by the quantity of capital, the quantity of intermediate inputs, and the quantity of labor. Labor, denoted $\mathcal{L}$, will be the sum of male (L^M) and female labor (L^F), where I allow a unit of female labor to be more or less productive than a unit of male labor

$$\mathcal{L} = \beta L^F + L^M$$

Firm j in year t takes these inputs and produce output, Y_{jt}, using some function F:

$$Y_{jt} = A_{jt} F(\mathcal{L}_{jt}, K_{jt}, M_{jt})$$

where A_{jt} is total factor productivity and β is the object of interest in this paper, measuring the labor-preserving tradeoff between men and women: a $\beta < 1$ implies that women are less productive than men and $\beta > 1$ implies that women are more productive than men. I discuss the details of this function F in subsection 4.2. First, I discuss the measurement of labor.

3.1 Measuring labor inputs

As first noted in Griliches (1957) and more recently Fox and Smeets (2011), there is a difficulty in measuring the quantity of labor a firm has: an individual with a college degree produces more than an individual with primary school; an executive produces more than a janitor. When measuring the labor a firm has access to, one must account for the quality of that labor. There are a number of possible ways to account for the quality of labor and I present estimates using all common definitions.

My baseline specification relies on constructing efficiency units of labor from the market-wide male wage equation. In constructing the efficiency units, I assume that a woman with the same characteristics as a man would have the same returns to those characteristics, and attribute any deviation from this to gender-based productivity differences. I control for the quality of various characteristics by running an efficiency units regression on the subsample of male workers in the data of the form

$$\begin{aligned} \ln e_{it}^M = \alpha^M &+ \gamma_1^M Age_{it} + \gamma_2^M Age_{it}^2 + \gamma_3^M Exp_{it} + \gamma_4^M Exp_{it}^2 + \gamma_5^M HS_{it} + \gamma_6^M Col_{it} + \gamma_7^M BA_{it} + \\ &+ \sum_{j=1}^{225} \delta_j^M H_{it}^j + \sum_{o=1}^{NOCC} \omega_o^M \mathbf{1}\{OCC = o\}_{it} + \sum_{t=2000}^{2010} \phi_t^M Year_t + \varepsilon_{it} \end{aligned} \tag{1}$$

where e_{it}^M is male worker i's labor market earnings in year t. Age, Exp, HS, Col, BA, measure a worker's age, hours of experience in the labor market, whether the worker has a high school or less, some college or trade school, or further education, respectively. H^j are indicators for (binned) hours worked in the year. The efficiency units regression also includes indicators of a worker's occupation at the three digit ISCO level, and year fixed effects. In this baseline specification, the amount of male labor in the firm J at time t is

$$L_{J(t)}^M = \sum_{i \in J(t), M} \widehat{e_{it}^M}$$

Like male labor in the firm, female labor is also measured using the returns to age, experience, occupation, education, and hours from the male wage equation (1), so that the amount of female labor in firm J at time t is

$$L_{J(t)}^F = \sum_{i \in J(t), F} \widehat{e_{it}^M}$$

Whether these characteristics are truly paid their marginal product is an interesting and potentially important question. Men in a given firm will on average be older than women in that firm. If younger workers are systematically underpaid, that will affect the interpretation of the gender productivity gap. The literature is not conclusive on this issue, but I estimate the gender productivity gap by age bins which should alleviate this concern.[9]

I also consider more detailed estimates of efficiency units—using eIncome data gives a finer measure of hours, and I allow for the interaction of occupation with industry and year in the calculation of efficiency

[9] For example, Hellerstein et al. (1999) find a discrepancy between wage and marginal product only for gender. Dostie (2011) uses more age categories and finds on average concave wage and productivity profiles, where wages do not deviate significantly from productivity. However, Hellerstein and Neumark (2007) finds some evidence that wages are deferred over the life cycle.

units. Finally, eIncome includes monthly earnings based on both "take-home" pay (the narrow earnings definition) and earnings including benefits such as contributions to retirement accounts.[10] I use this broad definition of earnings when constructing efficiency units as a robustness check. If high-skilled men preferred to be paid retirement contributions rather than having money in the bank, then my efficiency units calculations would understate the returns to skill specifically for men. More generally, differences between the wage gaps using the two measures of income are potentially important for understanding the potential role for non-pecuniary benefits in generating wage differences.

An alternative specification of the labor units in a firm follows the Griliches (1957) method which has

$$\begin{aligned}\mathcal{L} = L\left[1+(\phi_F-1)\frac{F}{L}\right]\cdot\left[1+(\phi_R-1)\frac{R}{L}\right]\cdot\left[1+(\phi_p-1)\frac{P}{L}+(\phi_o-1)\frac{O}{L}\right]\\ \cdot\left[1+(\phi_N-1)\frac{N}{L}+(\phi_S-1)\frac{S}{L}+(\phi_C-1)\frac{C}{L}\right]\end{aligned} \tag{2}$$

where F is the number of female workers, R is the number of married workers, P is the number of 35-54 year old workers, O is the number of workers 55 and older, N is the number of unskilled laborers, S is the white collar, technical, and sales workers, and C are the number of high skilled workers. This categorization is exactly used in Hellerstein et al. (1999) to capture the quality of a firm's labor force, with the exception that I eliminate the category of Black workers to fit the context of Denmark. ϕ_F is then equivalent to β as a measure of the productivity difference between a unit of female labor and male labor, accounting for differences in age, marital status, education, and occupation. Both methods assume that workers of different ages, occupations, etc., are perfect substitutes and that observable characteristics factor multiplicatively into productivity. The advantage of equation (1) is that I can account more flexibly for returns to age, experience, hours worked, and occupation (rather than using large discrete bins), and this flexibility ties estimates directly to traditional estimates of the residual wage gap—I am able to measure the residual productivity gap accounting for the same differences in observables commonly used when measuring the residual wage gap. Since there are advantages to both methods, I present estimates of the productivity gap using both (1) to predict the labor units in a firm and estimating the production function using labor as in (2).

Fox and Smeets (2011) directly tackle the problem of measuring labor quality in production functions by experimenting with a variety of approaches in Danish data similar to the data I use in this paper. They find that despite the biases associated with measuring the quality of labor using the wage bill, these estimates perform as well as estimates using Griliches-type specifications. In principle, any firm-specific innovations

[10] Other included benefits are the value of a free full-year of residence, the value of free summer residence, the value of a free pleasure boat, the value of a free TV license, the value of a free phone, the value of a PC, anniversary and severance pay, bonus income, and the value of "other" employee benefits as reported by the firm.

to productivity get passed on to workers in a bargaining setting, so using the wage bill to stand in for labor quality correlates covariates and error terms. Using the wage-bill method, labor in the firm is given by $\mathcal{L} = \tilde{\beta}W^F + W^M$ where W^F is the sum of earnings of women employed at the firm and W^M is the sum of earnings of men employed at the firm. The benefit of using the firm's wage bill method is that one asks whether a dollar paid to a man is as productive as a dollar paid to a woman, and one does not need to take a stand on how to model differences in returns to observables—they may differ and this is all summarized in wage differences. Here I distinguish between β in the baseline specification, which does not include wage discrimination by construction, and $\tilde{\beta}$ in this specification. A discriminatory firm which pays a woman less for the same output would result in a coefficient $\tilde{\beta}$ greater than one.

A final possibility for measuring labor quality (also presented in Fox and Smeets (2011)) is to take an ability-perspective: an Abowd et al. (1999) (hereafter, AKM) decomposition of wages models log earnings of a worker i at a particular firms j, as a linear function of a time-invariant person-specific component, such as ability, a time-invariant firm effect for firm j, representing wage premiums paid by some firms to all workers (for example due to productivity differences), and a residual match effect. The model can be expanded to include age effects, year effects, and various interactions of year or age profiles with education. The model is identified from movements of workers between firms, so long as these movements across the firm-quality distribution are not generated by innovations or permanent changes in worker productivity.[11] The worker fixed effects measure the ability of workers as rewarded by firms but take out everything that is constant within a firm over time (including firm productivity). Worker effects capture a worker's underlying ability regardless of the firm employing them. Another benefit of using person effects as a stand-in for labor quality is that if women prefer front-loaded pay over their lifecycle, while men prefer back-loaded pay, using the current wage bill or efficiency units regressions from male wages will bias estimates of the productivity gap relative to the pay gap. A life-cycle measure of worker ability gives an average measure of worker productivity. The form of the AKM decomposition used in this paper to predict person-effects is:

$$\ln e_{ijt} = \alpha_i + \theta_j + \phi_t + \Phi_1 HS_{it} + \Phi_2 Col_{it} + \Phi_3 BA_{it} + \epsilon_{ijt} \tag{3}$$

where $\ln e_{ijt}$ is the log earnings of worker i in year t in firm j, α_i is a person fixed effect, θ_j is a firm fixed effect, and ϕ_t are year fixed effects. Firm j's labor in period t is given by $\mathcal{L}_{jt} = \tilde{\beta}\sum_{i(t)\in j,F}\hat{\alpha}_i + \sum_{i(t)\in j,M}\hat{\alpha}_i$.

[11] This assumption is not innocuous and is violated some models of worker-firm matching. Nonetheless, Card et al. (2016) present some evidence consistent with these assumptions, such as the fact that wage changes for workers moving from top quartiles of the firm-effect distribution to lower-quartiles of the firm-effect distribution are the same size (but opposite direction) as wage changes for workers making the opposite transitions.

3.2 Production function

In the baseline, I model firm value added (revenue minus the cost of intermediate inputs) as a translog function of labor and capital:

$$\begin{aligned} &\log(Y_{jt} - M_{jt}) = a_{jt} + \theta_i \\ &+ \sum_{i \in I} \mathbf{1}\{j \in i\} \cdot \big(\alpha_{1,i} \log(\mathcal{L}_{jt}) + \alpha_{2,i} \log(K_{jt}) + \alpha_{3,i} \log(\mathcal{L}_{jt})^2 + \alpha_{4,i} \log(\mathcal{L}_{jt}) \log(K_{jt}) + \alpha_{5,i} \log(K_{jt})^2\big) \end{aligned} \tag{4}$$

where $\mathcal{L}_{jt}$ is a measure of the firm's labor force which is the sum of male and female efficiency units as described above, K_{jt} is firm j's value of capital stock, and a_{jt} is the log of revenue TFP, excluding industry fixed effects. I allow $\alpha_1, ..., \alpha_5$ to vary by industry and include industry fixed effects (θ_i) at the NACE 2-digit level.

Profit maximizing firms which take wages as given will set the ratio of the price of labor equal to the ratio of marginal revenue product. In this case, a marginal unit of male or female labor has the same effect on revenue, up to a constant β, so that

$$\frac{w_{jt}^f}{w_{jt}^m} = \beta$$

where w_{jt}^f is the average cost to firm j of hiring an additional unit of female labor at time t and w_{jt}^m is the cost of hiring an additional unit of male labor, controlling for observable differences in the quality of labor which enter the efficiency units calculation (1).

One assumption in the estimation of the relative productivity of female labor, β, is that male and female labor are perfect substitutes. For legal or social reasons, firms may prefer to hire women (or men) in particular when they are scarce. I relax the assumption of perfect substitutes and estimate a CES aggregation of male and female labor. In particular I estimate (4) but using

$$\mathcal{L} = (\beta(L^F)^{\frac{\rho-1}{\rho}} + (L^M)^{\frac{\rho-1}{\rho}})^{\frac{\rho}{\rho-1}}$$

where ρ is the elasticity of substitution for labor by gender and L^F, L^M are estimated both in efficiency units and using the wage bill. As $\rho \longrightarrow \infty$, men and women become perfect substitutes.

3.3 Interpretation of β

Fundamentally, β is the ratio of labor a firm can employ in production when replacing a male employee with a similar female employee. While it is an important and understudied metric, the gender productivity gap does not capture many dimensions of gender discrimination. A useful analogy might be to consider gender disparities in grading. While women are unlikely to take engineering courses, and very likely to take psychology courses in college (Bertrand, 2020), we may still ask, "conditional on the courses they take, are female students awarded different letter grades despite receiving the same scores on tests?" It may be that the female students in engineering courses look quite different from female students overall, and that measuring gender differences only in the form of whether grades differ from test scores is missing important aspects of the gender differences in the college experience. In this paper, I do not attempt to quantify all of the aspects of gender differences in labor market performance. I ask more specific questions: conditional on the jobs workers choose, do we see a gender productivity gap? How does it relate to the gender pay gap?

A difference in the productivity of observably similar male and female labor units can arise for a number of reasons, including unmeasured differences in skills, unmeasured differences in hours worked, unmeasured discrimination by the employer. It is possible that with better measures of hours worked, skills, etc., β would move toward one. Regardless of the level of the productivity gap, differences between productivity and pay shed light on the sources of the gender pay gap. As an example, if workers are compensated for the hours they work and women work fewer hours than men, then we would expect a productivity gap but no difference between pay and productivity. A difference between pay and productivity may arise due to discrimination by employers, or female employees less willing or able to successfully negotiate raises than male employees, as found in Biasi and Sarsons (2021), Babcock and Laschever (2003), Exley et al. (2020), among others.

When thinking about either the pay gap or the productivity gap, it is important to note that no experimenter has randomized the gender of individuals in the labor market. The fact that gender is not randomly assigned also means that the gender productivity gap and the gender pay gap are fundamentally descriptive statistics. The thought experiment cannot be randomly changing a worker's gender, because this would also potentially mean changing all of their life experiences, interactions, constraints, and choices. It may well be that what drives the gender pay gap is something correlated with gender, but not equivalent to it. Including controls in a wage regression, as I do here, similarly does not isolate the causal impact of gender, and may understate or overstate market level differences in outcomes by gender. Given known differences in experience, occupation choice, and hours worked by gender, the question of how remaining differences affect pay and productivity is the focus of this paper.

How does segregation in the labor market affect the interpretation of the gender productivity gap? The modal man works in manufacturing, but a small percentage of women work in manufacturing. One concern with this pattern is that gender differences in occupation and industry choice may imply differences in how representative a female worker at a particular firm is of female workers overall. This paper gives average differences in productivity and wages for various groups of workers. There is a parallel treatment of wages and productivity measures, so that the residual gender pay gap among manufacturing workers can be compared to the residual gender productivity gap among manufacturing workers without requiring male and female manufacturing workers to lie on the same underling ability distribution or be to be identical except in their gender. The latter exercise is potentially interesting but it is beyond the scope of this paper.

A second concern with segregation by gender in the workplace is that if there is a large gap between pay and productivity in manufacturing overall (for example), segregation by gender across occupation may drive my results. I return to this concern in the next section, using heterogeneity analysis to mitigate the concern that sorting patterns are driving the results. Overall, this paper provides measures of average differences between groups, and the results should be interpreted in light of substantial differences in the employment patterns of men and women on a variety of dimensions.

3.4 Sorting

What are the primary threats to identification? The estimate of the productivity gap comes from a comparison of the relative value added of firms with different proportions of female employees, conditional on observable differences between the workers in skill, experience, and hours worked, and conditional on the detailed industry of the firm. Differences in the value added of firms with different gender compositions in workers may arise because female workers are less productive, but may also arise because female workers work in firms with unobservably worse total factor productivity. I discuss this concern in the remainder of this subsection.

Descriptive evidence on firm-level sorting in the data

Do men and women select into firms with different productivities? If this is the case, then female wages would be lower than male wages and female efficiency units would be less productive on average because women work at firms with inferior technologies, however if women moved to better firms we would expect the pay and productivity gap to be eliminated. First, I check for selection using the method outlined in Card et al. (2016). I estimate the average firm effects for men and women in the sample of firms which are connected by workers moving across them in the data used in the production function estimation (private sector firms). In particular, separately for the male and female sample, I regress

$$\ln e_{it} = \alpha_i + \phi^{G(i)}_{j(i,t)} + X'_{it}\beta^{G(i)} + r_{it} \tag{5}$$

where $\ln e_{it}$ are the log earnings of worker i at date t, $j(i,t)$ is the firm employing worker i at date t, $G(i)$ indicates gender, and X_{it} is a vector of controls which includes year dummies interacted with education dummies and quadratic and cubic terms in age interacted with education dummies. The estimated firm effects can be interpreted as average difference in log earnings a worker would receive if he moved to particular firm J. I find no average difference between the firm fixed effects of women compared to men. To the extent that firm effects capture the relative productivity of firms, these results suggest that sorting is unlikely to bias estimates of the relative productivity of men compared to women.

Control function approach

One worry with using firm effects as a measure of firm productivity is that differences in the pay of workers across firms can be generated by factors other than TFP, for example, by compensating differentials. In this section, I describe a methodology for accounting for sorting in the production function estimation directly by using a control function approach developed in Olley and Pakes (1996).

This method for controlling for unobserved productivity differences between firms was developed because, as noted by Marschak and Andrews (1944), labor and capital choices are not exogenously assigned, but may be chosen by firms based on productivity. Any unobserved component of TFP which is known to the firm (such as a firm fixed effect) will affect the optimal choice of labor and capital. This biases estimates of the coefficients on labor and capital $\alpha_1, ..., \alpha_5$ in equation (4). The purpose of this paper is not to estimate labor and capital shares in Denmark, but rather to estimate the relative marginal product of men compared with women. For this purpose, endogeneity of input choice is not necessarily a problem. If firms hire a man or woman randomly, then β will not be correlated with productivity (or firm size). In some industries, this may be a reasonable approximation of hiring practices. Overall, however, it will be important to deal with the endogeneity of hiring choices.

If some portion of A_{jt} is known to firms at the time they make their labor decisions, the labor share coefficients will be biased. If TFP is also correlated with the decision to hire a man relative to a women, this will bias estimates of β. This would be the case, for example, if a firm which anticipated a change in technology which made it more productive preferred to hire men, perhaps because it believed men were better able to work with new technology or because men were more interested in working with the new technology and only men applied for the new jobs. In both cases, if we can can control for the unobservable known to the firm at the time they make hiring decisions, then we can control for the role of sorting by gender in the estimation of β.

Following Olley and Pakes (1996), hereafter, OP, I use investment to control for unobservables known to the firm at the time they choose $\mathcal{L}$. The intuition for this control is straightforward: assuming investment has a monotonic relationship with the unobservable component of TFP known to the firm at the time they make their decisions (conditional on capital), then it will be possible to invert the optimal investment rule and use this inverted rule as a control for the unobserved TFP. I describe the assumptions in more detail in

Appendix 2.

In this model, a_{jt} has a component which is a shock to the firm after they make labor and investment decisions, and also a known component (ω_t) which is unobservable to the econometrician directly. In other words, we can write $a_{jt} = \omega_{jt} + \varepsilon_{jt}$ where ω_{jt} is known by the firm and affects their optimal labor and investment decision. OP assumes that ω_{jt} is a scalar which follows an exogenous first order Markov process—that the distribution $p(\omega_{t+1})$ depends only on the observed ω_{jt}. This assumption allows for simple firm fixed effects $p(\omega_{jt+1}|\omega_{jt}) = p(\omega_{jt+1}|\bar{\omega}_j)$, but is more general (Ackerberg et al., 2007). Conditional on capital, investment is then increasing in the unobservable ω_{jt} so that we can invert the optimal investment rule and write $\omega_{jt} = \phi(i_{jt}, k_{jt})$. I use a 5th-degree polynomial in investment and capital to represent the inverted investment rule. In Appendix 2 I discuss the assumption necessary for the validity of a polynomial in investment and capital as a control function in more detail. Key, of course, to this exercise is the monotonicity of the investment in unobservable productivity. When I estimate an OP version of my main specification, I do so by 2-digit industry since the monotonicity assumption is not plausible when comparing across broad industries.

Characteristics of occupations preferred by women

In a final exercise, I investigate how the productivity gap varies with characteristics of occupations. Women and men are not equally represented across occupations in Denmark, and if women tend to work in occupations where the productivity gap is particularly large, then sorting at the occupation level is likely to be drive the results and we may worry that the estimates reflect the composition of particular occupations rather than gender differences per se. To shed some light on the relative role of occupational sorting in driving my estimates, I extend the model to allow measurement of gender gaps at the three digit ISCO level. For each occupation o, I estimate a model of the form

$$\begin{aligned} &\log(Y_{jt} - M_{jt}) = a^o_{jt} + \theta^o_i \\ &+ \sum_{i \in I} \mathbf{1}\{j \in i\} \cdot \left(\alpha^o_{1,i} \log(\mathcal{L}^o_{jt}) + \alpha^o_{2,i} \log(K_{jt}) + \alpha^o_{3,i} \log(\mathcal{L}^o_{jt})^2 + \alpha^o_{4,i} \log(\mathcal{L}^o_{jt}) \log(K_{jt}) + \alpha^o_{5,i} \log(K_{jt})^2\right) \end{aligned} \tag{6}$$

where $\mathcal{L}^o_{jt} = ((\beta^o L^{F,o} + L^{M,o})^{\frac{\rho-1}{\rho}} + \gamma(\beta^{\mathbf{O}\setminus o} L^{F,\mathbf{O}\setminus o} + L^{M,\mathbf{O}\setminus o})^{\frac{\rho-1}{\rho}})^{\frac{\rho}{\rho-1}}$. Efficiency units L are estimated excluding occupation, and the model has labor as a CES combination of the occupation of interest and all other occupations, denoted $\mathbf{O}\setminus o$. I drop from the data any subindustry which does not use any labor from occupation o. I also restrict to occupations which are over 100 workers, and am unable to disclose estimates of β for some occupations which employ few women.

I then gather data on several occupation attributes. First, I estimate the fraction of an occupation which is female. Next, I borrow the methodology outlined in Goldin (2014) and updated to a larger set of occupations by Bang (2021) to measure how flexible an occupation is in terms of five primary job characteristics measured in O*Net: time pressure, interpersonal relationships, contact with others, structured vs. unstructured work, and freedom to make decisions. The standardized flexibility index gives lawyers who have client-facing relationships, face time pressure, and have to work odd hours a value of 1.1 and in general high numbers indicate an occupation which is inflexible. I also take a measure from Erosa et al. (2022) of which occupations have the longest hours worked (by men), and I construct a measure of the average tenure of workers at a given firm by occupation. I will use these measures to shed light which types of occupations have the largest gaps and how this may affect interpretation of the gender productivity gap.

4 Results

4.1 Estimates of the productivity gap

Table 4 presents my baseline estimates of the relative productivity of women compared to men via the translog production function in (4). I find that one unit of female labor is equal to about 0.92 units of male labor. Column 3 gives changes the efficiency units calculation to include an interaction of occupation with hours, to capture the potentially different impact of long hours on wages across occupations. This does not affect estimates, however. The relative productivity of female labor is closer to 0.94 when using the 2008+ eIncome sample for estimation (columns 4-6). This is driven by the better hours measures available in that data, rather than a broader definition of earnings (comparing column 2 to column 4 relative to column 4 and column 5). Column 4 includes non-wage benefits which use the full definition of income available in eIncome, including payments to retirement accounts and the value of other non-monetary benefits, such as a computer, home, etc. Detailed efficiency in column 6 include the interaction of occupation with industry and year in the calculation of efficiency units and use the finest level of education major choice. Using a more detailed definition of efficiency units, the productivity gap falls to just four percent, indicating very little difference between a unit of male labor and a unit of female labor. Note, however, when making efficiency units categories too fine, the interpretation of the productivity gap becomes more confounded with differences in returns to (for example) majors between men and women.

While the efficiency units method summarized in (1) is my preferred specification, I also report alternative specifications of efficiency units of labor. In Table 5, column 1, $\tilde{\beta}$ is the coefficient on the female wage bill, where total labor is measured as the sum of the male wage bill and $\tilde{\beta}$ times the female wage bill in the firm. The interpretation of $\tilde{\beta}$, then, is the productivity of a dollar spent on female labor relative to a dollar spent

Table 4: Estimates of β (relative female productivity)

	(1)	(2)	(3)	(4)	(5)	(6)
β	0.919	0.928	0.928	0.941	0.940	0.960
	(0.008)	(0.007)	(0.005)	(0.009)	(0.010)	(0.009)
w^f/w^m	[0.861]	[0.861]	[0.901]	[0.900]	[0.920]	
Industry-specific shares	N	Y	Y	Y	Y	Y
Occupation interacted with hours	N	N	Y	N	N	N
Better hours (2008+)	N	N	N	Y	Y	Y
Including non-wage benefits	N	N	N	Y	N	Y
Detailed efficiency units	N	N	N	N	N	Y
R^2	0.8489	0.851	0.851	0.853	0.8523	0.853
N	852,729	852,729	852,729	258,978	258,978	258,978

This table gives estimates of β, the coefficient on female efficiency units in the translog production function regression (4) using (1) to form efficiency units. β can be interpreted as the relative productivity of female labor, controlling for differences in the quality of that labor captured by age, experience, education, and occupation. All regressions include 2-digit industry fixed effects and columns 2-5 allow the coefficients on labor and capital to vary at the 1-digit industry level. Standard errors are bootstrapped (50 samples) at the person level to account for estimation error in forming predicted efficiency units and then, for each estimate of efficiency units, cluster bootstrapped at the firm-level in the production function estimation step. The last row of the table, w^f/w^m is relative female wages, residual of quadratics in age and experience, education level fixed effects, and occupation fixed effects. Columns (3)-(5) use a subset of the data (2008 onward) in estimation because more detailed measures of hours worked are available in that time-period. This does reduce the productivity gap by about two percentage points. Columns (3) and (5) also include non-wage benefits in the estimation of efficiency units. Column (5) estimates efficiency units allowing for the interaction of occupation with industry and year the finest level of education major choice.

on male labor. I find that female labor is more productive per dollar, consistent with the evidence in Table 4 which shows a smaller productivity gap than pay gap. The estimates are three percentage points smaller when using the eIncome dataset (columns 2 and 3) and do not depend on whether non-wage benefits are included in the definition of the wage bill.

Column 4 of Table 5 is the measure of relative productivity using predicted individual fixed-effects from an AKM decomposition of wages as effective labor. Similar to the wage bill measure, the AKM measure of $\tilde{\beta}$ can be interpreted as the relative productivity of a female unit of labor measured in average lifetime wages (rather than current period wages) compared to a male unit of labor. Discrimination in the sense of uncompensated productivity is largest using the AKM method. The benefit of the AKM method is that if men and women have different preferences for backloading pay over the lifecycle, then the efficiency wage regressions in (1) may be biased. Since the predicted individual fixed effects in this estimate of the production function do not vary with TFP, it does not have the same mechanical bias that is induced when

Table 5: Alternative measures of β

	(1) Wage bill	(2) Wage bill (2008 +)	(3) Wage bill (2008+ including non-wage benefits)	(4) AKM	(5) Griliches ($\beta = \phi_F$)
$\tilde{\beta}$	1.0624	1.0332	1.0350	1.0961	
	(0.0035)	(0.0055)	(0.0055)	(0.0101)	
β					0.9519
					(0.0060)
R^2	0.8654	0.8724	0.8725	0.8184	0.6412
N	714,254	258,978	258,978	641,916	852,729

This table gives estimates of β, the coefficient on female efficiency units in the translog production function regression (4) using the wage bill (columns 1-3), person fixed effects from an AKM decomposition (column 4), and equation (2) (column 5) to measure effective labor. A coefficient larger than one on wage bill and AKM estimates is consistent with a productivity gap which is smaller than the wage gap, as in Table 4 above. The interpretation of these coefficients is that one dollar paid to female labor is more productive than one dollar paid to male labor. In contrast, column (5), though less than 1 is *also consistent with Table 4*. This is a measure of the relative productivity of a unit of female labor (not a dollar spent on female labor). All regressions use specification (4) to estimate the production function and include 2-digit industry fixed effects and allow the coefficients on labor and capital to vary at the 1-digit industry level. Standard errors are clustered at the firm-level.

using the wage bill.[12] Column 5 of Table 5 gives the Griliches estimate of the relative productivity of women as estimated in Hellerstein et al. (1999). The gap is quite small (five percent), given the coarse categories used in estimation: 3 age bins, 4 occupation bins, married vs. single, and male vs. female. All regressions use a translog production function with 2-digit industry fixed effects and industry-specific coefficients on labor and capital.

To deal with the possibility of selection of women into less (or more) productive firms, I take two approaches. First, I decompose log worker earnings into individual and firm fixed effects, controlling for education, age, and year fixed effects. I use the method outlined in Card et al. (2016) to do this, but focus on the connected subset of my sample of private-sector firms, as outlined in equation (5). The difference between firm effects $E(\phi^F_{j(i,t)}|m)$ and $E(\phi^F_{j(i,t)}|f)$ summarizes the degree to which men and women work in firms with different average pay. I find only a 0.005 log point difference between these expressions, suggesting the sorting by gender across private sector firms is not a large factor in the gender pay gap.[13]

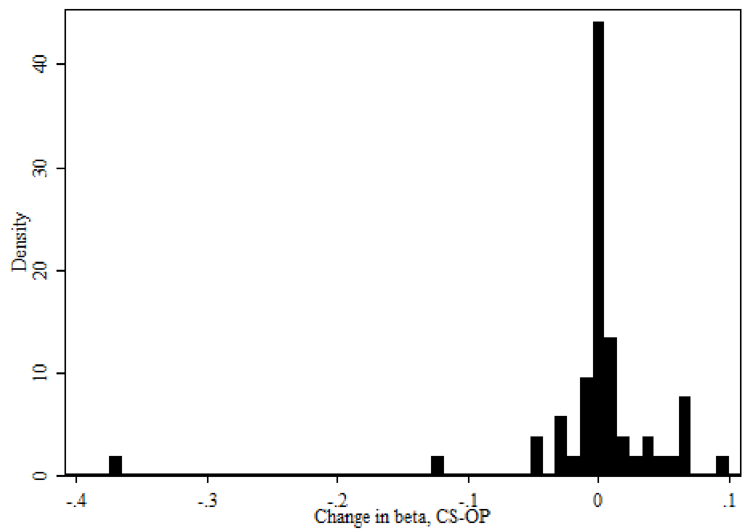
I can also use the Olley-Pakes control function approach described in Section 4.3 to account for TFP-based sorting by gender. Including a 5th-order polynomial in investment and capital in equation (4) does

[12] However, as noted before, Fox and Smeets (2011) argue that the wage bill measure works well for practical purposes, despite this bias.

[13] This contrasts with the role of firm-level sorting in the overall Danish economy. Focusing on an establishment level definition of the firm and including both the private and public sector, Gallen et al. (2017) find that 2.7 percentage points of the log pay gap can be explained by sorting in Denmark in the 2000s. I further discuss the sorting of women into the public sector below.

not change my estimates, but this may be because the assumptions underlying the OP method are not valid when looking across wide industries. Turning to a by-2-digit industry application of Olley-Pakes does result in a slightly larger role for selection in some sub-industries. Figure 1 gives a histogram of the difference between the productivity gap estimated using a translog production function at the 2-digit industry level, and a specification which adds to this production function the OP "control" for unobserved productivity which the firm uses when optimizing its factor choices.[14] The vast majority of my productivity gap estimates do not change when adding the OP control. Interestingly, those that do change move both in the positive and negative direction, suggesting there is some positive sorting of women into more productive firms (this is especially true in manufacturing sub-industries).

Figure 1: Effect of OP on estimates



This figure shows the difference between estimates of the productivity gap (by 2-digit NACE industry) with and without the OP control function. Estimates without an OP control function are the coefficients on female labor efficiency units in a translog (in capital and the sum of male and female labor) production function which include detailed industry fixed effects. Estimates with an OP control add a fifth-order polynomial in investment and capital to this production function in order to approximate unobserved productivity known to the firm at the time it makes its labor and investment choices. The difference between the coefficients on female labor in these two production functions captures the role of selection of women into lower TFP-firms in explaining the gap between male and female productivity. I find that in most industries, there in no meaningful selection (captured by a large mass at 0). If there is selection, it is not always negative (though it is more often negative). There are 54 unique sub-industries with at least 100 observations in the data.

Implicit in the production function estimated thus far is an assumption that male and female labor are perfect substitutes. This need not be the case. If firms prefer more equal shares of employees by gender, then

[14] I also require that the sub-industry have at least 100 observations in the data.

we can model total labor as the CES aggregation of male and female labor in the firm, with a non-infinite elasticity of substitution. Table 6 gives the coefficient on female labor under a CES specification for total labor. There is a decline in the estimated productivity of women relative to men. Intuitively, the first order condition (which is not used in the estimation) would imply that β should fall when the fraction of female labor matters to firms: women are more scarce than men in the labor force, yet they are paid less.[15] The wage bill estimate of discrimination also falls. The elasticity of substitution between male and female labor is between 5 and 10, depending on the specification. These fairly large estimates suggest that perfect substitutes is not an unreasonable assumption. Column 3 or Table 6 imposes the elasticity of substitution between college educated and non-college educated workers found in Autor et al. (2020) but treats male and female workers as perfect substitutes. This reduces relative productivity of women compared to men to 0.907, suggesting that assuming perfect substitutes across skill-types gives somewhat conservative estimates of the productivity gap.

Table 6: Imperfect substitutes: relative productivity of female labor (β) and the elasticity of substitution (ρ)

	(1)	(2)	(3)	
β	0.868	0.872	0.907	
	(0.008)	(0.008)	(0.015)	
$\tilde{\beta}$				1.029
				(0.004)
ρ	5.425	5.496	1.62	9.327
	(0.089)	(0.092)	(-)	(0.264)
N	852,729	852,729	852,729	714,254
Industry specific shares	N	Y	Y	Y
Wage bill	N	N	N	Y

This table gives estimates of β, the coefficient on female efficiency units in the translog production function regression (4) allowing total labor to be a CES combination of male and female efficiency units, where ρ is the elasticity of substitution between male and female labor. Column 3, in contrast to 1 and 2, estimates labor as a CES combination of college and less-than-college educated labor, where the elasticity of substitution is 1.62 as in Autor et al. (2020), while maintaining the assumption of perfect substitutes between male and female labor. All regressions include 2-digit industry fixed effects and columns 2 and 3 allow the coefficients on labor and capital to vary at the 1-digit industry level. Standard errors are bootstrapped (50 samples) at the person level to account for estimation error in forming predicted efficiency units and then, for each estimate of efficiency units cluster bootstrapped at the firm-level in the production function estimation step.

Next, I explore the source of the gap in productivity between men and women. The literature finds that

[15] The FOC is $\beta \left(\frac{L^f_{jt}}{L^m_{jt}}\right)^{-\frac{1}{\rho}} = \frac{w^f_t}{w^m_t}$.

the wage gap increases over a woman's life-cycle, markedly rising when she has children, and falling again only after mid-life (Kleven et al. (2019), Goldin (2014)). If mothers take more time off work to care for children (even in ways not measured by register data on hours worked) then we would expect this group to be driving up the productivity gap. If the productivity gap is instead driven by innate differences between men and women, some other factors correlated with gender, or mis-measurement, it would show up both for mothers and for non-mothers. I find that the productivity gap is driven only by mothers. Women without children are as productive as their male counterparts. I expand on this result in the next section.

4.2 Motherhood

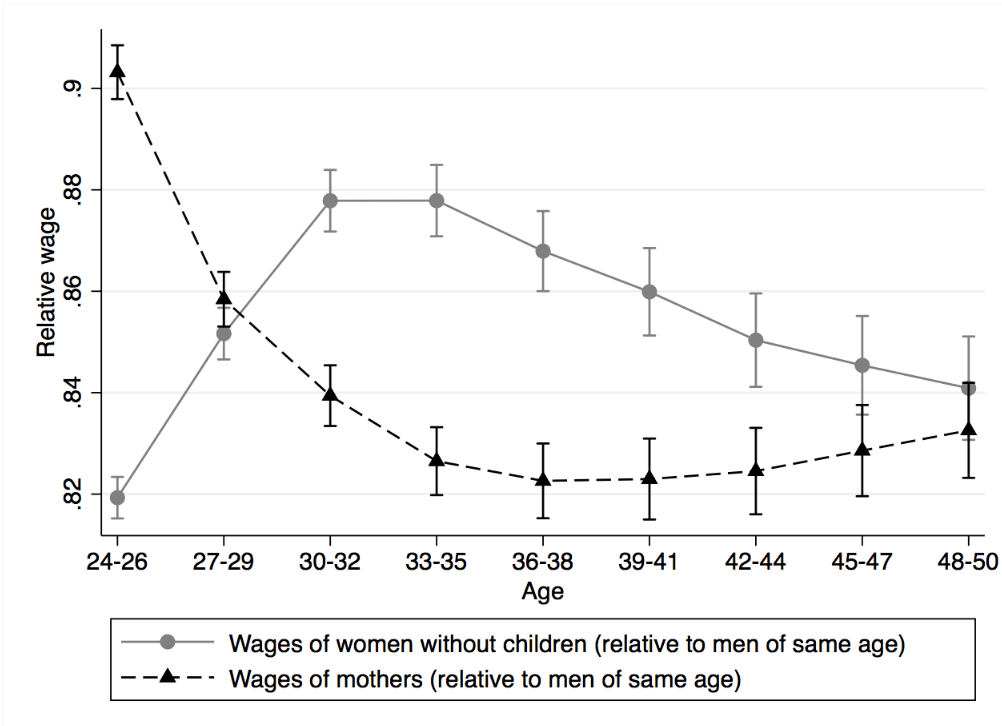
Bertrand et al. (2010) find that in a sample of recent US MBA recipients, the gender gap in career disruptions and female preference for shorter work hours was driven largely by mothers. In Denmark, recent work by Kleven et al. (2019) has argued that much of the Danish wage gap occurs with motherhood. Relative to other potential determinants of the gender wage gap, such as educational attainment, motherhood has explained a larger and larger share of the gender wage gap over time. While the presence of children could explain 40 percent of the gender earnings gap in 1980, children could explain 80 percent of the gap in 2011. The "child-penalty" comes in the form of (roughly equally) lower labor force participation of mothers, fewer hours of work for mothers, and lower wage rates for mothers. In my sample, I consider only mothers who have selected into work and those who are working in industries with good output data, notably excluding the public sector. In these data, motherhood explains less of the earnings gap—women with children are paid 85 cents on the dollar and women without children are paid 90 cents on the dollar compared to men without children. Nonetheless, mothers face the largest earnings gap. This paper is the first to study whether motherhood also affects the difference between earnings and productivity.

Register data makes it possible to incorporate whether or not a worker has a child into the estimates of relative productivity. In Figure 2 I plot the wage gap, measured using a wage regression of log earnings on a quadratic in experience, education dummies, industry, occupation, and hours fixed effects, as well as the interaction of parenthood, gender, and age categories in three year intervals. For each age category, I plot the pay of mothers, and women without children relative to men. As expected, the wage gap is largest for mothers. Figure 3 plots the productivity gap for mothers and women without children relative to men (the productivity analogue of Figure 2). As women move past childbearing age, both the residual productivity and pay of mothers and non-mothers converge.[16] However, when women are in their prime childbearing years, mothers are substantially less productive than non-mothers, while non-mothers and men have similar

[16]Note that there are cohort effects which I do not control for in wage regressions. Older mothers and non-mothers in the sample are from a different cohort than younger mothers and non-mothers. This is not a lifecycle analysis because the level of observation for productivity calculations is the firm, not an individual worker. I discuss composition extensively below.

productivity, and if anything women without children are more productive than men.[17] This is not true of wages: in all age brackets women's wages are lower than men's wages. This exercise raises two important questions: why are mothers less productive? Why are women without children underpaid?

Figure 2: Wage gap by motherhood and age

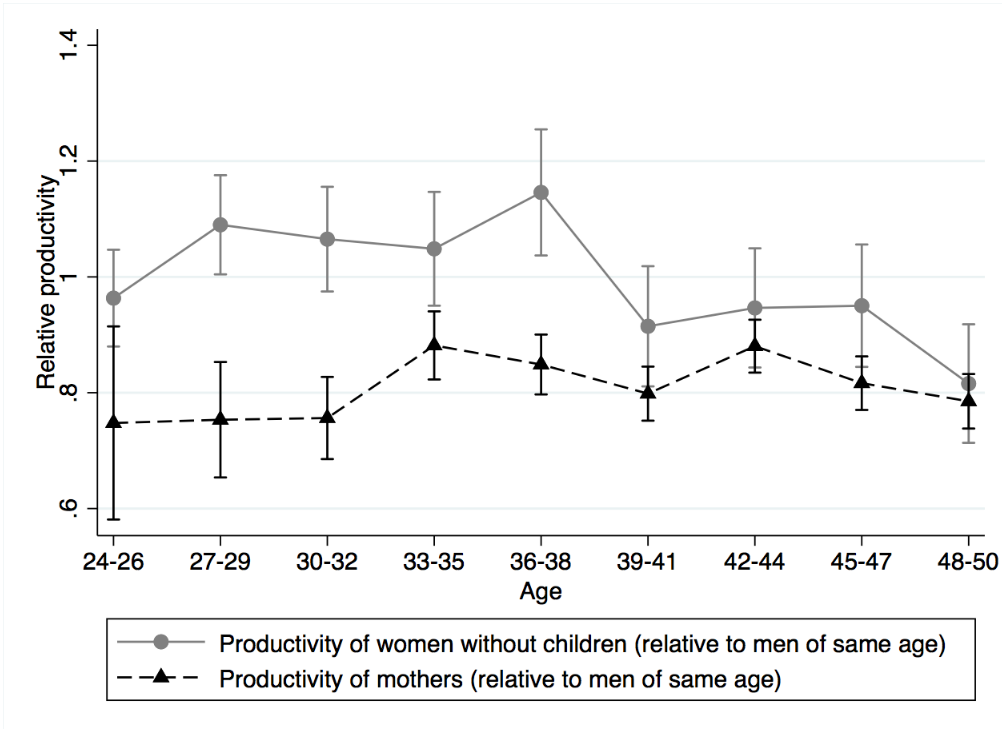


This figure shows the relative wages of women without children compared to men without children of the same age, as well as mothers and fathers compared to men without children of the same age. Relative wages are measured using a wage regression with 2 digit industry fixed effects, 3 education fixed effects, a quadratic in experience, and year fixed effects. For each age category, I normalize the wages relative to those of men without children of the same age.

In the section below, I discuss the potential sources of and interpretation of the motherhood productivity gap. Before doing this, it is important to note the fact that mothers' pay is aligned with their relative productivity does not imply that there is no discrimination. Suppose, for example, that an accounting firm never promotes a woman with children from a junior role in which she examines accounts to a more senior role where she also recruits new clients. In this case we may see that pay and productivity are aligned but there is discrimination in the form of the glass ceiling. This paper focuses on estimating whether women are underpaid for the work that they are doing, however, discrimination may exist in many other forms in the labor market, including the segregation that we see in the glass ceiling example. After discussing the lower

[17] I can also separate men by whether they have children. If I do this, I find that non-mothers are approximately as productive as fathers, and somewhat more productive than men without children (see Appendix Figure C1). This may reflect selection of men into fatherhood and subsequent household specialization.

Figure 3: Productivity gap by motherhood and age



This figure shows the relative productivity of women without children compared to men without children of the same age, as well as mothers and fathers compared to men without children of the same age. Relative productivity is measured using the baseline translog production function with industry specific shares and fixed effects, and the baseline specification for efficiency units but omitting age from the efficiency units. I model efficiency units of the interaction of 12 age bins and 4 gender/parenthood categories as perfect substitutes.

productivity of mothers compared to non-mothers, I turn to the question of why women without children are underpaid. I document sorting around motherhood, but develop a procedure to test the robustness of my estimates to a reasonable (empirically motivated) degree of selection of mothers into lower productivity firms.

Sorting and selection into work after motherhood

One straightforward check of whether mothers sort into different firms than non-mothers is to check whether mothers work in firms with lower average pay premiums, or firm fixed effects from an AKM decomposition (5). I obtain a firm pay premium associated with each firm. When I regress the firm pay premium on an indicator for whether the worker is a mother, within the sample of female employees. I find that on average the firm pay premium difference for mothers vs. non-mothers is 0.021, meaning that the average mother is working in a firm with a higher pay premium (by two percent) than the average non-mother. One worry with this is that it reflects the fact that the typical woman is a mother longer than she is a non-mother in the data and that the average pay premium tends to rise with age. I next study differences in the firm pay premiums around motherhood by estimating event studies around motherhood, controlling for the worker's age and the year.

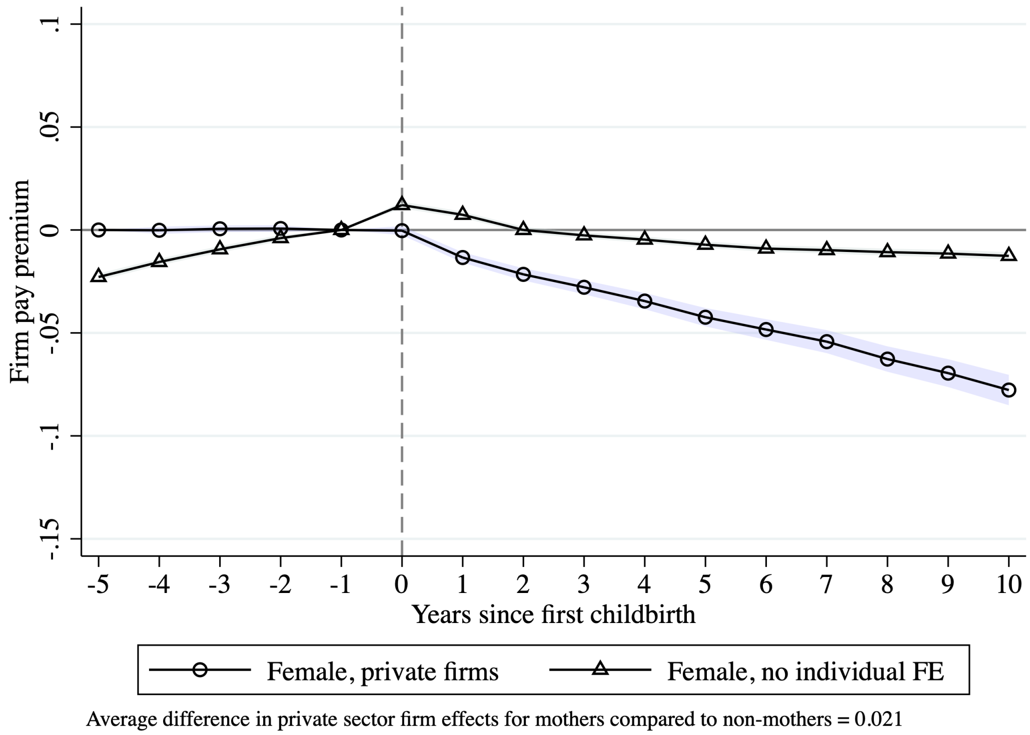
To do this, I begin with an event study as in Kleven et al. (2019) who examine the evolution of wages, employment, and earnings around motherhood for Danish women. This methodology compares women who gave birth at different times in their life, but who are the same age in the same year to identify the impact of motherhood:

$$y_{it} = \sum_{s=-5}^{10} \delta_s \mathbf{1}\{s = t - M_{it}\} + \sum_a \alpha_a \mathbf{1}\{A_{it} = a\} + \gamma_t + \varepsilon_{it} \tag{7}$$

where A_{it} is the age of individual i in year t so that α_a are age fixed effects, γ_t are year fixed effects, and δ_s are event time fixed effects, where the event of interest is the birth of a woman's first child. M_{it} is the year of birth of the first child. The coefficients of interest δ_s measure the difference in the outcome y for women who had children s years ago, relative to a counterfactual of not having children constructed from women of the same age who have not yet had children. The coefficients are interpreted relative to the level of this difference in an omitted year (here, event year $t-1$).

To study the types of firms women work at around motherhood, I use equation (7) where the outcome variable is firm fixed effects from an AKM decomposition (5). Figure 4 plots the coefficients δ_s (the triangles). One worry with the basic event study which regresses firm pay premia on event dummies (years since first birth fixed effects), as well as age fixed effects and year fixed effects, is that a causal interpretation requires that we assume women have children arrive exogenously, conditional on age and year. One might argue that this is implausible, and in fact the pre-trends in the figure above are evidence against such an assumption. To deal with this concern, we might include individual fixed effects. In this case, a causal interpretation requires only that we assume that children arrive randomly within women of a certain type summarized by individual fixed effects, conditional on age and year. However, adding individual fixed effects means that we cannot separately identify age and year trends from event-time trends without additionally assuming something about the event-time trends (Borusyak et al., 2021). In this case, we must impose flat pre-trends (or something equally restrictive). When we do this, we obtain the circle event-study shown in Figure 4. Again, to the extent that pre-trends are not flat and in fact look like the diamond line in the more parsimonious specification, by imposing flat pre-trends we effectively rotate down the post-event estimates. When we do this we see that women do seem to be sorting into lower pay firms after becoming mothers, and increasingly so as their children age. The average implied difference using these estimates is about five percent lower pay for mothers relative to others.

Figure 4: Firm premiums around childbirth

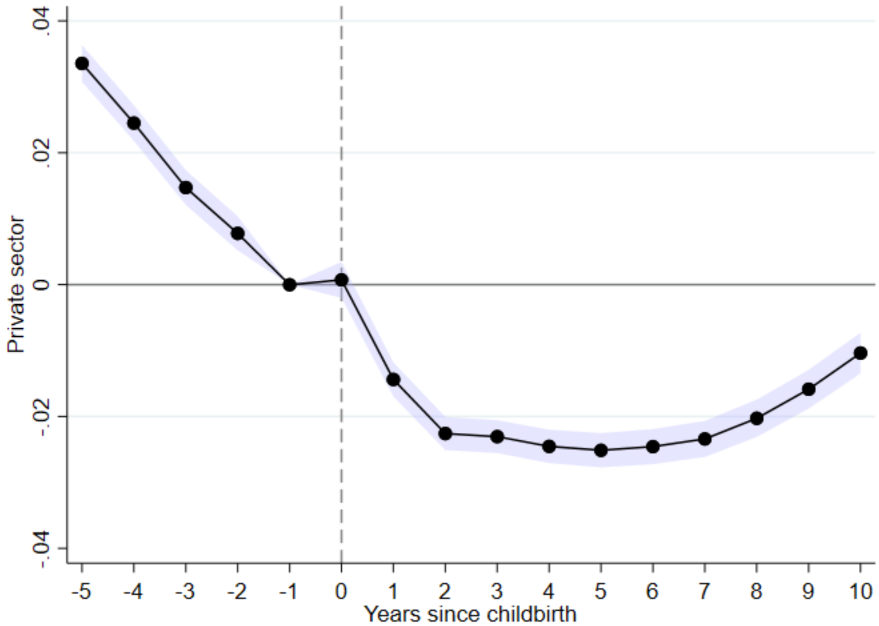


This figure displays the fixed effects δ_s from regression (7) in diamonds, and adding individual fixed effects and additionally constraining the pre-period not to have trends in diamonds.

Another important question arises when thinking about sorting around motherhood: do women move out of the data after becoming mothers? I next change the outcome variable to be an indicator of working in the private sector and in particular, working in the industries used in the production function estimation. The event-time coefficients are plotted in Figure 5. We see that relative to the year before giving birth, women are about two percentage points less likely to be working in the private sector after they become mothers, and are moving into the public sector at a fairly constant rate in the years before giving birth. However, those who shifted from private to public sector within 2 years of child's birth had about one-tenth a standard deviation less education and made about ten thousand dollars less in earnings than those who stayed in the private sector, controlling for age and year, so this selection out of the data is likely negative.

Why do the cross-sectional estimates differ from the event-study estimates with individual fixed effects, and which are most relevant for interpreting the productivity gap estimates? The effect of individual fixed effects on estimates combined with the result that mothers move into the public sector imply that women who stay in the private sector after motherhood are working in relatively higher pay firms, but that motherhood

Figure 5: Probability of working in the private sector around childbirth



This figure displays the fixed effects δ_s from regression (7) where the outcome is an indicator of working in the private sector.

does lead this positively selected group of women into lower-pay firms. These results suggest that had all women who work in the private sector before becoming mothers stayed in the private sector after becoming mothers, we would expect the relative productivity of mothers to be even lower than the estimates in Figure 3.

Unfortunately, understanding the role of sorting around motherhood is challenging because firm productivity is difficult to measure. Firm fixed effects in an AKM decomposition may reflect more than productivity differences and need not be correlated with productivity, due to differences in compensating differentials and amenities across firms, for example. The OP control function approach controls more directly for TFP, without relying on wages. Adding this control function does not change estimates of the productivity gap for mothers compared to non-mothers, but this approach relies on a long series of assumptions which I catalogue in Appendix A. Doubts about the degree of sorting among mothers to lower-productivity firms may remain. To address these doubts, I conduct an exercise to provide reasonable bounds the effect of such sorting on my estimates of the productivity gap.

If mothers sort into less productive firms, how much of the estimated productivity gap reflects this sorting and how much is explained by differences in the effect of mothers on production at a given firm? To answer this question, we must know how much sorting there is—what is the average difference in TFP in firms where mothers work compared to others? Without directly accounting for sorting, the error term in the production function estimation is correlated with $frac_{F,c}$, the fraction of employees who are mothers at the firm. If we are willing to put some structure on this error term and assume that it includes a term $\gamma frac_{F,c}$, then we can re-estimate the production function to include this term $\gamma frac_{F,c}$. We can calibrate the coefficient γ until the residual from the production function (log TFP) differs on average between mothers and other employees by the amount of known sorting. In many situations where unobservables bias estimated coefficients, a similar adjustment is not possible because we also do not have any way of bounding the role of unobservables. In this situation, the AKM decomposition above *does* suggest a bound. If we attribute all of the difference in the firm fixed effects from estimation of AKM equation (5) to differences in log TFP and assumed that there were no compensating differentials generating firm level wage premiums, and additionally assume that these differences are identified using the event-study approach described above, then the estimates from the event-study exercise with firm fixed effects as outcome variables imply that mothers work in firms which are five percent less productive than women without children.[18]

I calibrate γ such that when I assign each individual the log TFP of the firm where they work, the average difference in this variable between mothers and the rest of the population is equal to the five percent difference in firm pay premiums after women become mothers. The results of this exercise are in Table 7. The first column give the baseline estimate of the relative productivity of women without children $\beta_{F,nc}$ and mothers $\beta_{F,c}$ relative to men. The second column gives the estimate controlling for selection using the OP control function approach. The final column gives the estimates using my approach of calibrating unobserved selection to the AKM estimation of firm effects in an event study around motherhood. The OP and baseline results are similar. In contrast, controlling for $\hat{\gamma} frac_{F,c}$ in the production function estimation increases the estimate of the relative productivity of mothers to 0.9, up about nine percentage points from the main estimates. Note that the increase is mechanical—I'm imposing a negative relationship between the unboserved productivity and the fraction of mothers in the firm. What is notable about the results of this exercise is that the productivity gap does not go away, even assuming substantial sorting of mothers into less productive firms. And, with the assumption that mothers sort into less productive firms, the estimates of the productivity gap for non-mothers only slightly changed (even though there is some correlation in

[18]For a number of reasons, this is likely an upper bound on the degree of sorting. First, compensating differentials in amenities like schedule flexibility, parental leave, and commute time may attract mothers to firms which pay lower wages, but are attractive for other reasons (see Sorkin (2017) for some evidence of this). In addition, the assumption of flat pre-trends likely generates a downward rotation of the event study estimates in the post-period because in the estimation without individual fixed-effects, the slope in the pre-period is positive. A positive slope is consistent both with having children in response to positive career shocks, and with a model in which individuals dynamically work harder and save in the years before having children (which acts as a tax on work). Finally, this is an upper bound to the extent that I do not see any evidence of sorting in the cross-section using the AKM decomposition, only within-person over time. However, the cross-sectional correlations are what create bias in the production function estimation.

the types of firms where they work). These results suggest that when a reasonable degree of sorting by mothers into less productive firms is incorporated into estimation, the remaining gap between mothers and men in productivity is ten percent, still a sizable difference in the output generated by one efficiency unit of mother's labor relative to men's labor, or the labor of women without children. Overall, these estimates suggest that an empirically-motivated sorting by mothers into less productive firms does not completely explain the motherhood productivity gap.

Table 7: Effect of sorting of mothers into lower productivity firms on estimates

	Baseline	OP	Simulation with negative sorting of mothers
$\beta_{F,nc}$	1.228	1.239	1.218
	(0.014)	(0.015)	(0.014)
$\beta_{F,c}$	0.811	0.802	0.900
	(0.008)	(0.009)	(0.008)
Model target:	$\overline{TFP}_i^{F,c} - \overline{TFP}_i^{M;F,nc} = -0.05$		

This table gives the results of a simulation of the bias that sorting would induce in the estimates of the motherhood productivity gap, were such sorting summarized by a linear relationship with coefficient γ between the fraction of employees at a firm who are mothers, $frac_{F,c}$, and the error term in the production function estimation of (4). $\hat{\gamma} = -0.075$.

Motherhood and amenities

If mothers have strong preferences for the location of their job (near home) or the hours that they work, we would expect that their wages alone do not reflect their value in production. Instead, their total compensation would include, for example, the ability to set their own schedule or to walk to work. In general, the results in this paper are likely an underestimate of the difference between pay and productivity for mothers. If we are not including some valuable amenities in the calculation of compensation, then mothers may actually be overpaid relative to their productivity. And in some of these amenities are costly for the firm to provide, in terms of productivity, then this may in part explain the motherhood productivity gap. The true gap depends on how much wage women give up in equilibrium in order to access valuable amenities. This is an open question, though some studies have documented that women with children are willing to accept wage cuts in order to access amenities (Wiswall and Zafar, 2017; Mas and Pallais, 2017).

One channel through which motherhood may affect productivity is a flexibility penalty described in Goldin (2014). Adams-Prassl (2023) studies Amazon Mechanical Turk workers and finds that the gender earnings gap of 20 percent is not explained by task selection or worker experience, but rather by work interruptions which are most common among women with young children. Mas and Pallais (2017) find that women working in call centers are willing to receive lower wages in order to obtain schedules which are not set on short notice by employers. Goldin (2014) argues that mothers would like to avoid jobs in which they must be available on short notice to deal with client needs, or jobs with tight deadlines, as these are incompatible with the demands of childrearing for most families. When I estimate equation (6) to obtain occupation-level estimates of the gender productivity gap, point estimates suggest a larger productivity gap in those occupations more attractive to mothers (comparing "Flexible" to "Inflexible" occupation groups in Table 8), though the differences are not significant from one another. This difference is likely driven by sorting of mothers into more flexible occupations: Appendix Table D6 studies the motherhood productivity gap and finds no relationship with flexibility, though again the estimates are noisy. Table 8 also suggest that the productivity gap is larger in those jobs where men typically work more than fulltime hours, while the productivity gap is similar for mostly male and mostly female occupations.[19]

[19] Unfortunately, when zooming in on the three-digit level, estimates of the motherhood productivity gap in some occupations become very noisy. Appendix Table D6 documents the relationship between the motherhood productivity gap and occupation characteristics, as well as the productivity gap for women without children. The latter is not sensitive to outliers, but the motherhood gap is sensitive to outliers. Results should be interpreted very cautiously.

Table 8: Productivity gap by occupation groups

	Mostly male	Mostly female	Less than full time	More than full time	High turnover	Low turnover	Flexible	Inflexible
Average β	0.825	0.837	0.829	0.766	0.838	0.813	0.741	0.883
	(0.060)	(0.034)	(0.037)	(0.127)	(0.048)	(0.064)	(0.097)	(0.043)
Number of occupations	104	37	60	45	36	98	54	51

Note: This table displays the average productivity gap within occupations by groups. The groups are described in the columns, where turnover refers to whether the average tenure of male workers is above four years and flexibility refers to whether the standardized measure of flexibility is above or below zero. Log of average hours worked in the occupation comes from the ACS, replicated from Erosa et al. (2022) and matched to Danish occupation codes following Humlum & Meyer (2020), and the flexibility index based on Goldin (2014) and extended in Bang (2021), applied to Danish occupation codes and population distributions. Regressions are weighted according to the size of the occupation.

Aggregating up to larger industry/occupation groupings and assuming perfect substitutes as in the main specification (4) increases precision. Table 9 reports the results of estimating the productivity gap for mothers by two digit industry. The relationship between the wage gap and the productivity gap for mothers is very strong in all industries other than construction: in real estate and renting, other services, wholesale and retail trade, and manufacturing, the wage gap (generally close to 20 percent) is within two percentage points of the productivity gap.[20] These industries differ substantially in the fraction female, and in the type of work performed by women relative to men. Despite this, the motherhood productivity gap is large, while women without children are everywhere more productive than men. In addition, the difference between pay and productivity of non-mothers is around ten times higher than that for non-mothers.

Table 9: Gender productivity gap by industry and parenthood

	Accom./food	Constr.	Manuf.	W/R trade	Other serv.	Real est.
Female, no children	1.231	1.127	1.043	1.135	1.160	1.075
	(0.015)	(0.042)	(0.038)	(0.031)	(0.074)	(0.082)
	[0.926]	[0.794]	[0.839]	[0.886]	[0.906]	[0.879]
Female, children	0.852	0.881	0.846	0.813	0.829	0.834
	(0.009)	(0.021)	(0.020)	(0.017)	(0.047)	(0.045)
	[0.828]	[0.815]	[0.831]	[0.811]	[0.848]	[0.830]
N	298,370	135,808	72,000	69,215	11,799	18,920
R^2	0.770	0.819	0.848	0.704	0.600	0.7470

This table gives estimates of β, the coefficient on the category of efficiency units listed in the first column in a translog production function where the labor of men, women with children, and women without children are treated as perfect substitutes. Standard errors are in parentheses. Relative wages, residual of the same factors which enter efficiency units estimation, are in brackets. Both relative wages and relative productivity are compared to an omitted category of all men. Across industries, relative wages and relative productivity line up nearly perfectly for mothers, but are unrelated for non-mothers.

It is also possible to disaggregate β by occupation at a more broad level: Table 10 reports the results of production function estimation in (4) when labor is given by

$$\mathcal{L}_{jt} = \sum_{o \in \mathbf{O}} \beta^o_{F,c} \hat{E}^o_{j(t),F,c} + \beta^o_{F,nc} \hat{E}^o_{j(t),F,nc} + \beta^o_M \hat{E}^o_{j(t),M}$$

where $\hat{E}_{j(t),\cdot}$ are the sum of efficiency units in firm j estimated *excluding occupation fixed effects* in category

[20]Gender preferences can translate to differences in wages directly via compensating differentials: when risk-taking, physical and otherwise is rewarded and women shy away from risky jobs, they will on average by paid less than men.

$\cdot$, where these categories are female with children, female without children, and male, respectively.[21] The set of occupations **O** is management; jobs requires knowledge at the highest level (from school teachers to researchers); jobs requires knowledge at the medium level, such as information technology workers; office jobs; sales, service, and care jobs; craftsman jobs; blue collar jobs; military jobs; and agriculture, forestry, fishing requiring basic level knowledge; large heterogenous categories unknown; and other.[22]

[21] The role of discrimination and performance for high-level workers has been studied recently using a Norwegian policy change that forced firms to increase the number of women on their boards to 40 percent. Matsa and Miller (2013) find that firms that increased the number of women on their board in response to this policy change had fewer layoffs than comparison firms in Sweden. This increased their labor costs and reduced short-term profits. This suggests that women at high levels in a company (or the boards that picked them) do have different preferences and management styles than men, on average. It's difficult to generalize the results of this policy change since the women who were newly put on boards differed significantly on observables from women already on boards, as did their companies. It's also not clear that the policy had any long-run effect on the gender wage gap. In recent work, Bertrand et al. (2018) find that this reform did not affect the probability that women enrolled in business programs, changed relative female wages, or affected women's fertility and marital decisions, even though the policy change was tangible and women knew that they would be more likely to be put on a company board in the future.

[22] In estimation, β_M for the category "unknown" is restricted to be one (so all estimates are productivity relative to men in occupations unknown).

Table 10: Gender productivity gap by occupation and parenthood

	M	HS	MS	WC	S	C	LS
$\beta^o_{F,nc}/\beta^o_M$	1.045	1.154	1.240	0.992	1.073	1.129	0.744
	(0.111)	(0.036)	(0.030)	(0.019)	(0.022)	(0.036)	(0.067)
	[0.803]	[0.888]	[0.834]	[0.909]	[0.935]	[0.884]	[0.877]
$\beta^o_{F,c}/\beta^o_M$	0.940	0.816	0.867	0.753	0.552	0.907	0.570
	(0.048)	(0.020)	(0.014)	(0.010)	(0.013)	(0.022)	(0.025)
	[0.752]	[0.858]	[0.791]	[0.865]	[0.827]	[0.874]	[0.870]
N individuals	509,790	1,138,811	2,007,428	1,560,069	1,581,954	2,433,454	1,520,541

This table gives the ratio of relative productivity coefficients by occupation. The first row of coefficients is the relative productivity of women without children relative to men and the second row of coefficients is the relative productivity of women with children compared to men in the same occupation (o). o is one of the 11 occupations described below modeled as perfect substitutes in a translog production function. Bootstrapped standard errors are in parentheses. Relative wages, residual of the same factors which enter efficiency units estimation, are in brackets.
Occupation codes: M = Management at the highest level, HS = Job req. knowledge at the highest level (from school teachers to researchers), MS = Job req. knowledge at the medium level (e.g. information technology workers), WC = Office jobs, S = Sales/Service/Care, C = Craftsman jobs, LS = Blue collar jobs. Occupation data is only available for a person's main job. About 30 thousand person-year observations list the primary job as *military* and another 30 thousand in the category *agriculture, forestry, fishing requiring basic level knowledge*, the estimates for these productivity gaps are in Appendix Table D3, along with large heterogenous categories *unknown* and *other*.

The relative productivity of women without children is everywhere higher than that of women with children, and generally higher than the productivity of men in the same occupation. However, women seem to be substantially less productive than men in low skilled jobs, without commensurately lower pay. These are jobs in which union contracts have the largest influence on wages and retention, which may explain why women appear to be so dramatically over-compensated. Women without children are in every occupation paid less than their relative productivity, while women with children are generally (but not everywhere) over paid.

Why are women without children underpaid?

Why are women without children underpaid? One possibility is that women without children work in different types of jobs or firms than mothers, and this sorting, rather than motherhood itself, drives the relationships we observe. The age-decomposition by parenthood in Figure 3 suggests that women's productivity declines substantially when they have children. Selection is part of this story, as discussed above. Many women leave the private sector when they have children in Denmark (see Figure 5 and Pertold-Gebicka et al.

(2016)). Those who shifted from private to public sector within two years of child's birth had about one-tenth a standard deviation less education and made about ten thousand dollars less in earnings than those who stayed in the private sector, controlling for age and year. Even controlling for a potential preference among mothers for lower-productivity firms, this suggests that the gap in Table 7 is a lower bound of the true productivity gap between mothers and non-mothers, since mothers remaining in the private sector are positively selected.

Another possibility is that when hiring women without children, employers expect these women to become mothers soon.[23] The employer then smooths the worker's wage contract in anticipation of motherhood. This may also arise as a preference of workers: for example, those who anticipate reducing their hours substantially after having children may prefer contracts which smooth wages to contracts reflecting child-related reductions in hours. Or, more simply, they may avoid asking an employer for a raise right before having children. A final possibility which I consider empirically is a type of reverse causality, in which women who are underpaid avoid having children.

I find evidence consistent with the smoothing of contracts around motherhood: women without children who look most like mothers and who will soon become mothers, as well as women working in occupations with longer average within-firm tenure, experience the largest gaps between pay and productivity. I find no evidence for reverse causality by examining the pay and productivity gaps for women who never have children. For these women, pay is aligned with productivity.

Is this type of statistical discrimination a reasonable explanation of the gap between wages and productivity? The probability of having a child at age 30 is 13.8 percent in Denmark. Fertility rates are similarly high for all the prime child-bearing years. Suppose, for the purpose of this example, that the length of a wage contract is 4 years. Then employers would want to pay a 28 year old woman 7 percent less than a man because of risk of childbirth. In other words, taking into account childbearing probabilities, the expected productivity of a 28 year old non-mother over the next four years is seven percent less than her male counterpart's.

I present three indirect tests of this hypothesis. First, I test whether the difference between the wage and productivity gap (underpayment) is larger for women who are on the verge of having children. If employers are able to somehow predict future childbearing and use this prediction to inform wages even for women without children, then we would expect to see a larger gap between pay and productivity for women who

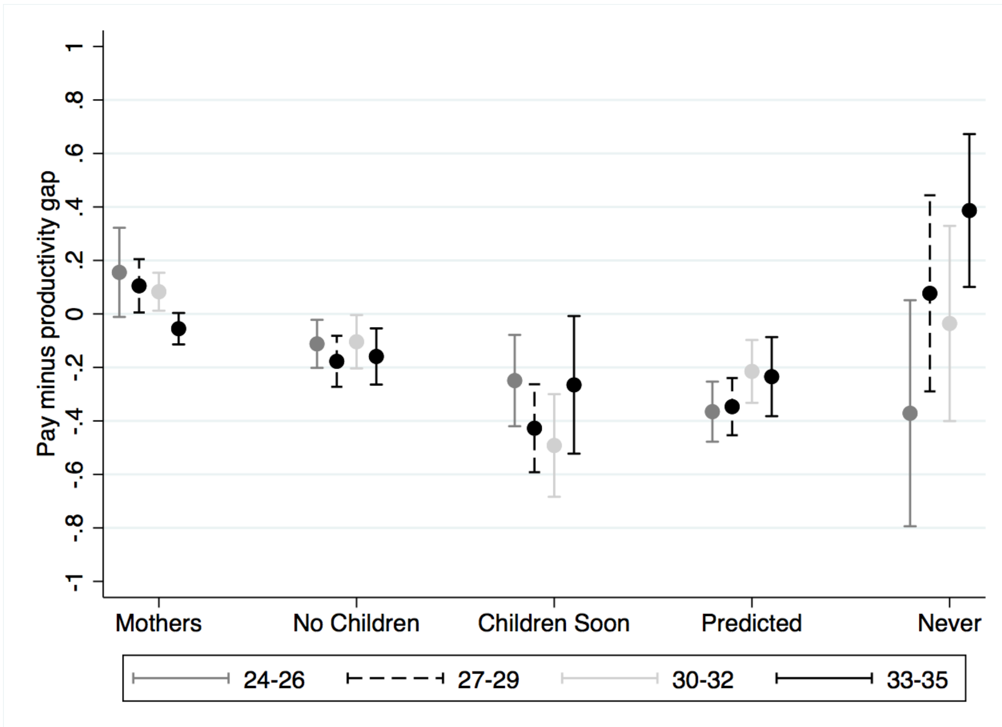
[23]The employer may also anticipate that some women are likely to go on parental leave and parental leave is costly for the firm. Even though in Denmark, the direct costs of leave are paid by the government, not firms, firms may still face re-optimization frictions when a worker takes leave. For some evidence of this see Gallen (2018) in Denmark and Ginja et al. (2023) in Sweden. Mothers, however, are also quite likely to take parental leave (due to having a second or higher parity child) so this is unlikely to be the driver of the underpayment of women without children, since I don't find evidence that mothers in this age range are underpaid.

are going to have children in the near future relative to women who are not going to have children in the near future. Of course, some part of this is endogenous: women who are underpaid or don't feel appreciated by their employer may decide to have children. I next test whether purely based on observables, women who look most like mothers (who employers also might guess are the most likely to soon become mothers) are underpaid. If characteristics of mothers generated differences between pay and productivity, then we'd expect this group of women to have compensation fairly in line with their productivity, as mothers do. However, if these characteristics allow employers to predict their future propensity to have children, and employers use these characteristics to smooth wages, then we would expect large differences between pay and productivity for these women. Finally, this explanation only makes sense if women stay with the same employer for a long period of time. I test whether the gap between pay and productivity for women without children is largest in occupations with typically long same-firm tenure among men.

All of these tests point in the same direction, that the largest gaps between pay and productivity come in prime childbearing years for women who are most likely to have children. Figure 6 plots the difference between the pay gap (relative to same-age men) and the productivity gap (relative to same-age men) for five groups of women: mothers, estimates combining Figures 3 and 2; women without children who will not have children soon ("No Children") and women without children who will have children in the next three years ("Children Soon"), estimated from a production function which takes these women as perfect substitutes with men and mothers and estimates gaps separately for each group; women who are the top 25 percent most likely to have children among childless women based on observables (see figure notes for details), estimated from a production function which takes these women as perfect substitutes with men, women who are not predicted to have children soon, and mothers estimating gaps separately for all of these groups by age; and a final group of women, labeled "Never," who do not have children at any point in the administrative data. The estimates for this last group women who will not go on to have children later come from a production function estimation using only the first year of data, since for this group I have the most information about subsequent fertility. In this data, women who are 24 will be known not to have children by the time they are 35 and women who are 35 will be known not to have children by the time they are 46. This means the younger women in the sample have not quite reached the end of their fertility window and some women in this group may be mis-labeled as never having children. The only group of women for whom zero—indicating no underpayment—is included in the 95 percent confidence interval are mothers and women who will not go on to have children in the future. Women who will have children soon have larger gaps between pay and productivity compared to women who will not have children soon. Finally, women who look most like future mothers based on observables have the gaps between pay and productivity which are similar to other non-mothers, especially those who will have children soon, and dissimilar from mothers.

The fact that women who look like future mothers have pay-productivity gaps as large as actual future mothers seems most consistent with a model in which employers statistically discriminate in wage offers, and not that women who are about to have children avoid asking for pay commensurate with productivity, but standard errors are wide enough that I can't rule out the latter possibility.

Figure 6: Pay relative to productivity gaps across groups



This figure shows the pay gap minus the productivity gap for various groups of women in four age bins (covering the twelve years during which women in the data are most likely to have first children within three years conditional on not having children yet). The first group is mothers, estimated as described in Figures 2 and 3. Estimates for the second and third group, labeled "No Children" and "Children Soon" come from a model similar to what is in Figures 3 but with women without children split further into those who will have children within three years ("Children Soon") and those who will not ("No Children"). The fourth group, labeled "Predicted," are estimates from a model where women without children are split into those who are in the top quartile of those predicted to have children soon based on observables. This prediction is from a probit for the probability of having children within three years on income, age fixed effects, a quadratic in experience, industry fixed effects, marital/cohabitation status, education, and occupation fixed effects. Finally, the last group gives estimates from a model of firm output in 2000 only (the earliest year) in which women without children are split into those who will not have children by the last year in the data and those who will. The 24-26 bin should be interpreted cautiously for this group as fertility is only observed for 11 years, so some women who have children after 35 in this group are mislabeled (and this may explain part of the reason that this observation is an outlier). In all cases, I model efficiency units of the interaction of 12 age bins and 4 gender/parenthood categories as perfect substitutes with one another and with men.

Table 11 next turns to variation in average tenure at the firm across occupations to test whether women who work in jobs which typically involve longer relationships with a single employer experience lower wages before they have children, conditional on productivity. I find that this is indeed the case, women without children who have the same relative productivity have a 2.5 percentage point lower wage when they work at a firm which has one year longer average tenure.[24] This remains true when controlling for other potentially important characteristics of the occupation, such as the fraction male, the average hours worked by men, and the flexibility of the occupation. Interestingly, none of these other characteristics are associated with any additional wage penalty for women without children.

Table 11: Wage gap for women with no children by occupation characteristics

	(1)	(2)	(3)	(4)	(5)
$\beta_{F,nc}$	0.014	-0.023	0.044	0.021	-0.015
	(0.040)	(0.044)	(0.052)	(0.050)	(0.060)
Fraction male	0.000				0.049
	(0.062)				(0.097)
Average tenure		-0.025			-0.027
		(0.010)			(0.014)
Log average hours			-0.104		0.083
			(0.134)		(0.173)
Flexibility index				-0.033	-0.018
				(0.025)	(0.030)
Number of occupations	120	115	98	96	93

Note: This table displays the results of a regression of occupation-specific measures of (1) the fraction of the occupation which is male in the Danish administrative data used in this paper, (2) the average tenure at a given firm of men in the occupation in the Danish administrative data used in this paper (3) the log of average hours worked in the occupation in the ACS, replicated from Erosa et al. (2022) and matched to Danish occupation codes following Humlum & Meyer (2020), (4) the flexibility index based on Goldin (2014) and extended in Bang (2021) and applied to Danish occupation codes/population distributions. Column (5) includes all occupation characteristics simultaneously. Regressions are weighted according to the precision (inverse of the standard error squared) of the estimate of the productivity gap, $\beta_{F,nc}$, which is also included in all regressions.

[24] As in Table 8, lower tunover/high average tenure is associated with larger productivity gaps. Tenure is measured using the sample of male employees. I take average years of employment by men at a given firm, by 3 digit ISCO. The data restrictions imply I can only observe a top-coded measure of tenure for workers who stay at the same firm for the entire observation window, but only 4.7 percent of observations are affected by this restriction. There are meaningful differences in average tenure of men even within broader occupation groups. For example, in low-skilled work and restricting to men, ISCO code 932—manufacturing laborers—have 4.72 years of tenure while 913—domestic and related helpers, cleaners, and launderers—have 2.7 years of average tenure at a given firm.

5 Conclusion

This paper presented estimates of the relative productivity of men and women, accounting for age, education, experience, occupation, and hours worked. Overall, I find that the productivity of women is about 8 percent lower than men, controlling for age, education, experience, and hours worked. This implies that productivity differences explains just under two-thirds of the residual gender pay gap. This productivity difference may arise from differences in the effort, extra (undocumented) hours worked, or effectiveness of men relative to women.

While on average, the pay gap is quite close to the productivity gap, this is not true over all of the lifecycle. In particular, women without children are estimated to be as productive if not more productive than men without children, but they are paid less than these men. Mothers, on the other hand, are substantially less productive than fathers and are paid commensurate with this productivity gap. The data are consistent with a statistical discrimination mechanism: the gap between productivity and pay for women who are going to have children soon relative to women who are not going to have children soon is significantly larger than the same difference for men, and larger in those occupations with longer average tenure of workers.

The results reported above are generally robust to various different specifications of a firms quantity/quality of labor. The baseline estimate uses an efficiency units approach which predicts returns to various observables (potentially correlated with gender) using the relationship between men's wages and the observables. Another method is to use the wage bill to represent labor quality, assuming that higher wages correspond to higher labor quality. To deal with the concern that wages respond to productivity shocks, the AKM-efficiency units method uses person fixed effects from a wage decomposition to represent labor quality. Finally, the Hellerstein et al. (1999) approach estimates the relative productivity of various observables directly as inputs in the production function. All of these methods give similar results, despite making different assumptions about the transformation of observables differences in productive characteristics to wages. The results are also robust to using more detailed wage and hours measures (which exist only from 2008 onward and are robust to controlling for the potential role of sorting by women into less productive firms. Finally, the general pattern that women without children are as productive as men, while mothers are substantially less productive holds across industries and occupations.

Like the wage approach, the productivity approach in this paper implies that motherhood is central to the discussion of the gender pay gap. However, while differences in pay between men and women are largest for mothers, I find that differences in pay which cannot be explained by productivity are largest for women without children. The factors driving the gap between the pay and productivity of women without children (preferences, discrimination, occupation sorting) is an interesting avenue for future research.

References

J. M. Abowd, F. Kramarz, and D. N. Margolis. High wage workers and high wage firms. *Econometrica*, 67 (2):251–333, 1999.

D. Ackerberg, K. Caves, and G. Fraser. Structural Identification of Production Functions. mimeo, UCLA, 2004.

D. Ackerberg, L. Benkard, S. Berry, and A. Pakes. Econometric tools for analyzing market outcomes. In J. Heckman and E. Leamer, editors, *The Handbook of Econometrics*, chapter 63. North-Holland, 2007.

A. Adams-Prassl. The gender wage gap in an online labour market: The cost of interruptions. *Review of Economic Studies*, Forthcoming, 2023.

D. J. Aigner and G. G. Cain. Statistical theories of discrimination in labor markets. *Industrial and Labor Relations Review*, 30(2):175–187, 1977. ISSN 00197939, 2162271X. URL http://www.jstor.org/stable/2522871.

J. Albrecht, A. Björklund, and S. Vroman. Is there a glass ceiling in Sweden. *Journal of Labor Economics*, 21(1):145–177, 2003.

J. Albrecht, P. S. Thoursie, and S. Vroman. Parental leave and the glass ceiling in Sweden. *IFAU working paper 2015:4*, 2015.

J. Altonji and R. Blank. Race and gender in the labor market. In O. Ashenfelter and R. Layard, editors, *Handbook of Labor Economics*, volume 3, chapter 48, pages 3143–3259. Elsevier, 1 edition, 1999.

N. Angelov, P. Johansson, and E. Lindahl. Parenthood and the gender gap in pay. *Journal of Labor Economics*, 34(3):545–579, 2016. doi: 10.1086/684851. URL https://doi.org/10.1086/684851.

J. Apesteguia, G. Azmat, and N. Iriberri. The impact of gender composition on team performance and decision making: Evidence from the field. *Management Science*, 58(1):78–93, 2012.

D. Autor, C. Goldin, and L. F. Katz. Extending the race between education and technology. *AEA Papers and Proceedings*, 110:347–51, May 2020. doi: 10.1257/pandp.20201061. URL https://www.aeaweb.org/articles?id=10.1257/pandp.20201061.

G. Azmat and R. Ferrer. Gender gaps in performance: Evidence from young lawyers. *Journal of Political Economy*, 125(5):1306–1355, 2017.

L. Babcock and S. Laschever. *Women Don't Ask: Negotiation and the Gender Divide*. Princeton University Press, 2003.

M. Bang. Job flexibility and household labor supply: Understanding gender gaps and the child wage penalty. *Mimeo*, 2021.

G. Becker. *The Economics of Discrimination*. University of Chicago Press, 1971.

G. S. Becker. Human capital, effort, and the sexual division of labor. *Journal of Labor Economics*, 3(1): S33–S58, 1985. ISSN 0734306X, 15375307. URL http://www.jstor.org/stable/2534997.

L. I. O. Berge, A. J. G. Pires, K. Bjorvatn, and B. Tungodden. Competitive in the lab, successful in the field? *Journal of Economic Behavior & Organization*, forthcoming, 2015.

M. Bertrand. *New Perspectives on Gender*, volume 4 of *Handbook of Labor Economics*, chapter 17, pages 1543–1590. Elsevier, 2011. URL https://ideas.repec.org/h/eee/labchp/5-17.html.

M. Bertrand. Gender in the twenty-first century. *AEA Papers and Proceedings*, 110:1–24, May 2020. doi: 10.1257/pandp.20201126. URL https://www.aeaweb.org/articles?id=10.1257/pandp.20201126.

M. Bertrand, C. Goldin, and L. F. Katz. Dynamics of the gender gap for young professionals in the financial and corporate sectors. *American Economic Journal: Applied Economics*, 2(3):228–55, July 2010. doi: 10.1257/app.2.3.228. URL http://www.aeaweb.org/articles?id=10.1257/app.2.3.228.

M. Bertrand, S. E. Black, S. Jensen, and A. Lleras-Muney. Breaking the Glass Ceiling? The Effect of Board Quotas on Female Labour Market Outcomes in Norway. *The Review of Economic Studies*, 86(1):191–239, 06 2018. ISSN 0034-6527. doi: 10.1093/restud/rdy032. URL https://doi.org/10.1093/restud/rdy032.

B. Biasi and H. Sarsons. Flexible Wages, Bargaining, and the Gender Gap*. *The Quarterly Journal of Economics*, 137(1):215–266, 08 2021. ISSN 0033-5533. doi: 10.1093/qje/qjab026. URL https://doi.org/10.1093/qje/qjab026.

S. E. Black and P. E. Strahan. The division of spoils: Rent-sharing and discrimination in a regulated industry. *The American Economic Review*, 91(4):814–831, 2001. ISSN 00028282. URL http://www.jstor.org/stable/2677814.

F. D. Blau. *Equal pay in the office*. Lexington Books, 1977.

F. D. Blau and L. M. Kahn. Understanding International Differences in the Gender Pay Gap. *Journal of Labor Economics*, 21(1):106–144, 2003.

F. D. Blau and L. M. Kahn. Female Labor Supply: Why Is the United States Falling Behind? *American Economic Review: Papers & Proceedings*, 103(3):251–256, 2013.

F. D. Blau and L. M. Kahn. The Gender Wage Gap: Extent, Trends, and Explanations. *Journal of Economic Literature*, 55(3):789–865, September 2017.

S. Bond and M. Soderbom. Adjustment Costs and the Identification of Cobb Douglas Production Functions. Economics Papers 2005-W04, Economics Group, Nuffield College, University of Oxford, Jan. 2005.

K. Borusyak, X. Jaravel, and J. Spiess. Revisiting event study designs: Robust and efficient estimation, 2021. URL https://arxiv.org/abs/2108.12419.

L. Bursztyn, T. Fujiwara, and A. Pallais. 'Acting Wife' Marriage Market Incentives and Labor Market Investments. *American Economic Review*, 107(11):3288–3319, 2017.

T. Buser, M. Niederle, and H. Oosterbeek. Gender, Competitiveness, and Career Choices. *The Quarterly Journal of Economics*, 129(3), 2014.

D. Card, A. R. Cardoso, and P. Kline. Bargaining, sorting, and the gender wage gap: Quantifying the impact of firms on the relative pay of women. *The Quarterly Journal of Economics*, 131(2):633–686, 2016. doi: 10.1093/qje/qjv038. URL +http://dx.doi.org/10.1093/qje/qjv038.

C. Cook, R. Diamond, J. Hall, J. List, and P. Oyer. The Gender Earnings Gap in the Gig Economy: Evidence from over a Million Rideshare Drivers. Technical report, 2018.

R. Croson and U. Gneezy. Gender differences in preferences. *Journal of Economic Literature*, 47(2):448–74, June 2009. doi: 10.1257/jel.47.2.448. URL http://www.aeaweb.org/articles?id=10.1257/jel.47.2.448.

C. Dahl, D. le Maire, and J. Munch. Wage Dispersion and Decentralization of Wage Bargaining. *Journal of Labor Economics*, 31(3):501 – 533, 2013.

B. Dostie. Wages, productivity and aging. *De Economist*, 159(2):139–158, 2011.

R. Ericson and A. Pakes. Markov-perfect industry dynamics: a framework for empirical work. *Review of Economic Studies*, 62 (January):53–82, 1995.

A. Erosa, L. Fuster, G. Kambourov, and R. Rogerson. Hours, occupations, and gender differences in labor market outcomes. *American Economic Journal: Macroeconomics*, 14(3):543–90, July 2022. doi: 10.1257/mac.20200318. URL https://www.aeaweb.org/articles?id=10.1257/mac.20200318.

C. L. Exley, M. Niederle, and L. Vesterlund. Knowing when to ask: The cost of leaning in. *Journal of Political Economy*, 128(3):816–854, 2020. doi: 10.1086/704616.

J. A. Flory, A. Leibbrandt, and J. A. List. Do competitive workplaces deter female workers? a large-scale natural field experiment on job entry decisions. *The Review of Economic Studies*, 82(1):122–155, 2015. doi: 10.1093/restud/rdu030. URL +http://dx.doi.org/10.1093/restud/rdu030.

J. T. Fox and V. Smeets. Does input quality drive measured differences in firm productivity?*. *International Economic Review*, 52(4):961–989, 2011. ISSN 1468-2354. doi: 10.1111/j.1468-2354.2011.00656.x. URL http://dx.doi.org/10.1111/j.1468-2354.2011.00656.x.

R. G. Fryer and S. D. Levitt. An empirical analysis of the gender gap in mathematics. *American Economic Journal: Applied Economics*, 2(2):210–40, 2010.

Y. Gallen. The effect of parental leave extensions on firms and coworkers. Working paper, 2018.

Y. Gallen, R. Lesner, and R. Vejlin. The Labor Market Gender Gap in Denmark: Sorting Out the Past 30 Years. *IZA working paper no. 10789*, 2017.

G.-L. Gayle and L. Golan. Estimating a Dynamic Adverse-Selection Model: Labour-Force Experience and the Changing Gender Earnings Gap 1968?1997. *The Review of Economic Studies*, 79(1):227–267, 09 2011. ISSN 0034-6527. doi: 10.1093/restud/rdr019. URL https://doi.org/10.1093/restud/rdr019.

D. Gicheva. Working long hours and early career outcomes in the high-end labor market. *Journal of Labor Economics*, 31(4):785–824, 2013.

R. Ginja, A. Karimi, and P. Xiao. Employer responses to family leave programs. *American Economic Journal: Applied Economics*, 15(1):107–35, January 2023. doi: 10.1257/app.20200448. URL https://www.aeaweb.org/articles?id=10.1257/app.20200448.

U. Gneezy, K. L. Leonard, and J. A. List. Gender differences in competition: evidence from a matrilineal and a patriarchal society. *Econometrica*, 77(5):1637–1664, 2009.

C. Goldin. A Grand Gender Convergence: Its Last Chapter. *American Economic Review*, 104(4):1091–1119, 2014.

C. Goldin and C. Rouse. Orchestrating impartiality: The impact of "blind" auditions on female musicians. *American Economic Review*, 90(4):715–741, September 2000. doi: 10.1257/aer.90.4.715. URL http://www.aeaweb.org/articles?id=10.1257/aer.90.4.715.

Z. Griliches. Specification bias in estimates of production functions. *Journal of Farm Economics*, 39(1): 8–20, 1957. ISSN 10711031. URL http://www.jstor.org/stable/1233881.

M. Gunderson. Male-female wage differentials and policy responses. *Journal of Economic Literature*, 27(1): 46–72, 1989. URL http://EconPapers.repec.org/RePEc:aea:jeclit:v:27:y:1989:i:1:p:46-72.

J. K. Hellerstein and D. Neumark. Production function and wage equation estimation with heterogeneous labor: Evidence from a new matched employer-employee data set. In Berndt and Hulten, editors, *Hard-to-Measure Goods and Services: Essays in Honor of Zvi Griliches*. 2007.

J. K. Hellerstein, D. Neumark, and K. R. Troske. Wages, Productivity, and Worker Characteristics: Evidence from Plant-Level Production. *Journal of Labor Economics*, 17(3):409–446, 1999.

J. K. Hellerstein, D. Neumark, and K. R. Troske. Market Forces and Sex Discrimination. *The Journal of Human Resources*, 37(2):353–380, 2002.

J. Hersch and L. S. Stratton. Housework, fixed effects, and wages of married workers. *The Journal of Human Resources*, 32(2):285–307, 1997. ISSN 0022166X. URL http://www.jstor.org/stable/146216.

F. Heyman, H. Svaleryd, and J. Vlachos. Competition, takeovers, and gender discrimination. *ILR Review*, 66(2):409–432, 2013. doi: 10.1177/001979391306600205. URL https://doi.org/10.1177/001979391306600205.

A. Ichino and E. Moretti. Biological gender differences, absenteeism, and the earnings gap. *American Economic Journal: Applied Economics*, 1(1):183–218, January 2009.

L. Kamas and A. Preston. The importance of being confident; gender, career choice, and willingness to compete. *Journal of Economic Behavior & Organization*, 83(1):82–97, 2012.

H. Kleven, C. Landais, and J. E. Søgaard. Children and gender inequality: Evidence from denmark. *American Economic Journal: Applied Economics*, 11(4):181–209, October 2019. doi: 10.1257/app.20180010. URL https://www.aeaweb.org/articles?id=10.1257/app.20180010.

A. Leibbrandt and J. A. List. Do women avoid salary negotiations? evidence from a large-scale natural field experiment. *Management Science*, 61(9):2016–2024, 2015. doi: 10.1287/mnsc.2014.1994. URL https://doi.org/10.1287/mnsc.2014.1994.

A. Manning and H. Robinson. Something in the way she moves: a fresh look at an old gap. *Oxford Economic Papers*, 56(2):169–188, 2004.

T. Markussen, E. Reuben, and J.-R. Tyran. Competition, Cooperation and Collective Choice. *Economic Journal*, 124:F163–F195, 2014.

J. Marschak and W. Andrews. Random Simultaneous Equations and the Theory of Production. *Econometrica*, 12(3/4):143–205, 1944.

A. Mas and A. Pallais. Valuing alternative work arrangements. *American Economic Review*, 107(12):3722–59, December 2017. doi: 10.1257/aer.20161500. URL https://www.aeaweb.org/articles?id=10.1257/aer.20161500.

D. A. Matsa and A. R. Miller. A Female Style in Corporate Leadership? Evidence from Quotas. *American Economic Journal: Applied Economics*, 5(3):136–169, 2013.

D. Mortensen, B. J. Christensen, and J. Bagger. Wage and productivity dispersion: Labor quality or rent sharing? 2010 Meeting Papers 758, Society for Economic Dynamics, 2010. URL https://EconPapers.repec.org/RePEc:red:sed010:758.

C. Mulligan and Y. Rubinstein. Selection, Investment, and Women's Relative Wages over Time. *The Quarterly Journal of Economics*, 123(3):1061–1110, 2008.

M. Niederle and L. Vesterlund. Gender and Competition. *Annual Review of Economics*, 3:601–630, 2011.

OECD. *OECD Employment Outlook 2010—Moving Beyond the Job Crisis*. 2010.

C. Olivetti and B. Petrongolo. The evolution of gender gaps in industrialized countries. *Annual Review of Economics*, 8(1):405–434, 2016. doi: 10.1146/annurev-economics-080614-115329. URL https://doi.org/10.1146/annurev-economics-080614-115329.

G. S. Olley and A. Pakes. The Dynamics of Productivity in the Telecommunications Equipment Industry . *Econometrica*, 64(6):1263–1297, 1996.

J. Pan. Gender segregation in occupations: The role of tipping and social interactions. *Journal of Labor Economics*, 33(2):365–408, 2015.

B. Pertold-Gebicka, F. Pertold, and N. D. Gupta. Employment Adjustments around Childbirth. *IZA working paper no. 9685*, 2016.

S. W. Polachek. Occupational self-selection: a human capital approach to sex differences in occupational structure. *The Review of Economics and Statistics*, 63:60–69, 1981.

E. Reuben, P. Sapienza, and L. Zingales. Taste for Competition and the Gender Gap Among Young Business Professionals. *NBER Working Paper No. 21695*, 2015.

J. E. Rockoff and M. A. Herrmann. Does menstruation explain gender gaps in work absenteeism? *Journal of Human Resources*, Spring, 2012.

I. Sorkin. The role of firms in gender earnings inequality: Evidence from the united states. *American Economic Review*, 107(5):384–87, May 2017. doi: 10.1257/aer.p20171015. URL https://www.aeaweb.org/articles?id=10.1257/aer.p20171015.

Statistics-Denmark. Production and generation of income (10a3-grouping). www.dst.dk. Data retrieved Aug 6th, 2015.

J. Stearns. The Long-Run Effects of Wage Replacement and Job Protection: Evidence from Two Maternity Leave Reforms in Great Britain. mimeo, 2017.

M. Thomas. The impact of mandated maternity benefits on the gender differential in promotions: Examining the role of adverse selection. mimeo, 2015.

M. Wasserman. Hours Constraints, Occupational Choice, and Gender: Evidence from Medical Residents. mimeo, 2017.

A. Weber and C. Zulehner. Competition and Gender Prejudice: Are Discriminatory Employers Doomed to Fail? *Journal of the European Economics Association*, 12(2):492–521, 2014.

M. Wiswall and B. Zafar. Preference for the Workplace, Investment in Human Capital, and Gender*. *The Quarterly Journal of Economics*, 133(1):457–507, 08 2017. ISSN 0033-5533. doi: 10.1093/qje/qjx035. URL https://doi.org/10.1093/qje/qjx035.

Y. J. Zhang. Culture, Institutions, and the Gender Gap in Competitive Inclination: Evidence from the Communist Experiment in China. *SSRN: http://ssrn.com/abstract=2268874*, 2015.

A Model appendix

Assumptions underlying the Olley-Pakes control function approach to the selection problem:

Assumption 1: Factor prices are constant across firms

The assumption that factor prices are constant across firms allows us to infer that firms which choose different levels of investment do so because they predict that their TFP will differ in the next period. If firms face different labor prices, particularly by gender, then β may still biased due to unobservables (factor prices). In Denmark this assumption is not particularly offensive, since wages are set in no small part by collective bargaining and generally are compressed relative to the US.

Assumption 2: Labor is a non-dynamic input

This assumption *would* be unreasonable in countries where it was difficult to re-adjust the labor force every year. Denmark, however, prides itself on a "Flexicurity" system. This is the combination of a very flexible labor market—it's very easy to fire and hire workers in Denmark—combined with a secure safety net in the case of unemployment. In Denmark and the US, just over 25% of employees are new hires in each year, and about 25% separated from their employer in the same period. In Norway, these rates are closer to 17%. In Italy, they are about 15% (OECD, 2010). See Appendix Figure A1 for a graph of cross-country separation and hiring data.

Figure A1: Separation rates and hiring rates across countries

This figure is directly replicated from OECD data on relative workforce flexibility, OECD (2010) Figure 2.1, see Annex 3.A1. Country averages of reallocation rates expressed in percentage of total dependent employment and adjusted for industry composition. Austria: 2002-07; Belgium: 2000-07; Canada: 2000-06; the Czech Republic: 2001-07; Denmark: 2000-06; Finland: 2000-07; France: 2000-06; Germany: 2000-06; Greece: 2000-05; Hungary: 2000-05; Iceland: 2002-07; Ireland: 2000-05; Italy: 2000-06; the Netherlands: 2000-07; Norway: 2000-04; Poland: 2004-05; Portugal: 2000-06; the Slovak Republic: 2002-06; Slovenia: 2002-07; Spain: 2000-05; Sweden: 2000-06; Switzerland: 2000-07; Turkey: 2007; the United Kingdom: 2000-07; and the United States: 2000-06.

Assumption 3: Conditional on capital, investment is monotonically increasing in the unobservable ω_{jt}

These assumptions rule out, for example, adjustment costs which differ across firms within an industry. Scalar investment is given by $i_{jt} = i_t(\omega_{jt}, k_{jt})$. Pakes (1994, Theorem 27) shows that when $i > 0$, $i_t(\omega_t, k_t)$ is increasing in ω for every k, so that we can invert the investment rule and write $\omega_{jt} = \phi(i_{jt}, k_{jt})$.[25][26] Approximating this investment rule with a flexible, higher-order polynomial in k and I yields the equation

$$\log(Y)_{jt} = a_t + \psi_1 \log(\mathcal{L})_{jt} + \psi_2 k_{jt} + \phi(i_{jt}, k_{jt}) + \varepsilon_{jt} \tag{8}$$

[25] Ericson and Pakes (1995) discuss the conditions for this invertibility in equilibrium in more detail.

[26] The general formation also includes firm age as a state variable, but omitting age does not affect the invertibility in equilibrium and simplifies the problem, since the relationship between firm age and productivity is not of interest in this paper.

where $\phi(i_{jt}, k_{jt})$ is a flexible 3rd degree polynomial in i and k. Since labor does not enter the ϕ polynomial, the labor share and β are identified simply by running this regression.

Ackerberg et al. (2004) (ACF) note that there is a simultaneity problem if investment and labor are truly chosen simultaneously—in this case labor demand can be written $\mathcal{L}(\omega, k)$, problematically. Indeed, if labor can be written as a flexible polynomial in i and k, then there is perfect collinearity between ϕ and inputs in $\mathcal{L}$, making estimated labor coefficients meaningless. ACF suggest a 2-step solution to this problem, as well as a timing assumption which corrects the problem. In the Danish context and with yearly data, this timing is not particularly offensive. More formally:

Assumption 4: Labor is chosen first, then investment is chosen based on an information set correlated but not collinear with the information used to choose labor.

As suggested by Ackerberg et al. (2004) to eliminate the problem posed if i and labor are chosen based on exactly the same information set and factor prices do not vary across firms.[27] In general, all that is needed for identification is different adjustment speeds of various factors (see, Bond and Soderbom (2005)).

To estimate capital share, ψ_2, we can use the knowledge of ψ_1 and β obtained in the first stage to write

$$\log(Y)_{jt} - \psi_1 \log(\mathcal{L})_{jt} = a_{jt} + \omega_{jt} + \varepsilon_{jt}$$

Since ω is a first order Markov process, we can decompose it into its expectation given information at time $t-1$, $g(\omega_{j,t-1})$ and a residual, ξ_{jt}. In addition, we estimate the combination of capital effects in the first stage. Let the first stage coefficient on capital be κ_{jt}. We now have

$$\log(Y)_{jt} - \psi_1 \log(\mathcal{L})_{jt} = a_t + \psi_1 k_{jt} + g(\kappa_{j,t-1} - a_{t-1} - \psi_2 k_{j,t-1}) + \xi_{jt} + \varepsilon_{jt}$$

This paper is focused on the estimation of β, which is identified in the first stage in the case of firm entry and exit, measurement error in investment, and lumpy levels of investment (Ackerberg et al., 2007).

[27] See Ackerberg et al. (2007) for an extensive discussion of OP and alternatives.

B Data appendix

B.1 Measuring value added

Value added is revenue minus the cost of intermediate inputs. There are changes in the definitions of variables and the introduction of new variables which make the calculation year-specific. Following Mortensen et al. (2010):

- From 2000-2001, $Y = (OMS + AUER + ADR + DLG + 0.0079 \times TGT) - (KRH + KENE + KLOE + UDHL + UASI + UDVB + ULOL + ANEU + SEUD)$
- From 2002-2003: $Y = (OMS + AUER + ADR + DLG) - (KRH + KENE + KLOE + UDHL + UASI + UDVB + ULOL + ANEU + SEUD)$
- From 2004 on: $Y = (OMS + AUER + ADR + DLG) - (KVV + KRHE + KENE + KLOE + UDHL + UASI + UDVB + ULOL + ANEU + SEUD)$

Where

OMS = Revenue

$AUER$ = Work conducted at own expense and recorded under assets

ADR = Other operating revenue

DLG = Final inventory minus initial inventory

KRH = Cost of intermediates

$KENE$ = Cost of energy

$KLOE$ = Cost of subcontractors

$UDHL$ = Housing rents

$UASI$ = Purchases of minor equipment

$OEEU$ = Other external costs

$SEUD$ = Secondary costs

TGT = Total credits

$UDVB$ = Purchases of temp. agency services

$ULOL$ = Costs of long-term leasing

$ANEU$ = Other external costs, net of secondary costs

KVV = Purchases of good for resale

$KRHE$ = Cost of intermediates

The capital stock is measured as $K = AADI + GRBY + ATAM + FMAA$ where

$AADI$ = Operating equipment and other equipment and facilities

$GRBY$ = Buildings and sites

$ATAM$ = Technical equipment and machinery

$FMAA$ = Pre-paid material fixed assets and material fixed assets under construction

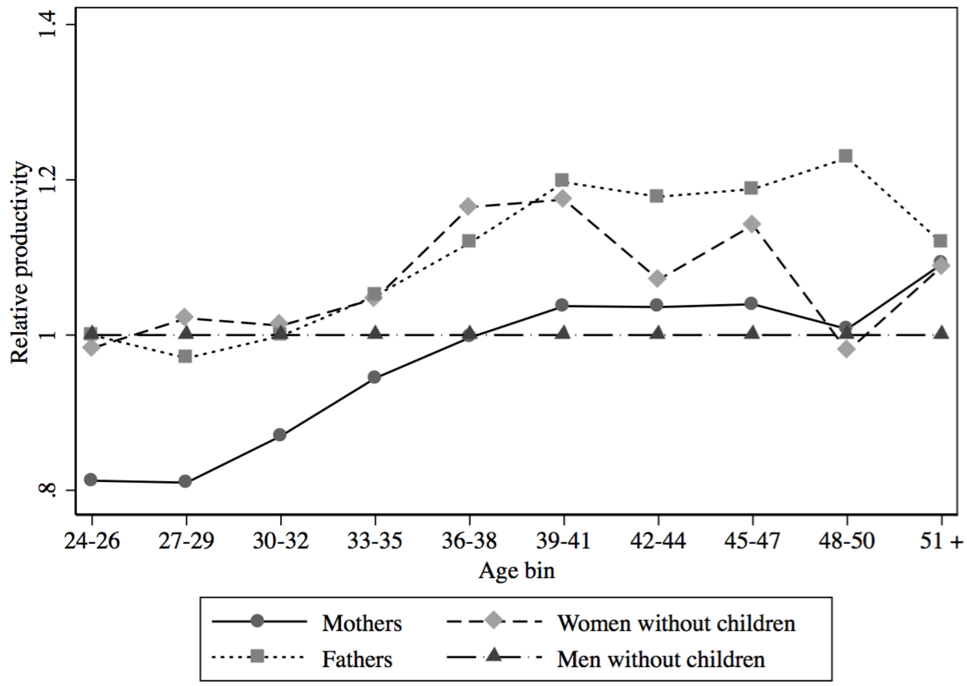
Table B1: Hours and jobs per person-firm-year-month in eIncome

Entries per p-f-y-m	Percent of p-f-y-m observations	25th pctile hours/month	50th pctile hours/month	75th pctile hours/month
1	98.88%	91	157.61	160.33
2	0.98%	6	22	62
3	0.025%	5	17	51
4+	0.0015%	3	8	31

This table describes the distribution of number of separate entries an individual (p) in a given firm (f) in a given year (y) and month (m) has in the eIncome data. The eIncome register is formed from taking monthly payroll statements which include occupation, hours worked, and various compensation breakdowns (take-home pay, adding fringe benefits, adding retirement contributions, etc). Multiple worker observations within a firm in a month (p-f-y-m) may arise because a worker changes occupations/job types in a month or has multiple occupations in a given month in a given firm, or they may arise due to a break in the employment spell in a month (resumed in the same month). There are a total of 135430660 person-firm-year-month observations. 98.88% of these have only one record and virtually all the rest have only two records. These data are used to construct efficiency units when estimates use the 2008+ sample of firms.

C Appendix figures

Figure C1: Productivity gap by parenthood and age



This figure shows the relative productivity of women without children compared to men without children of the same age, as well as mothers and fathers compared to men without children of the same age. Relative productivity is measured using the baseline translog production function with industry specific shares and fixed effects, and the baseline specification for efficiency units but omitting age from the efficiency units. I model efficiency units of the interaction of 12 age bins and 4 gender/parenthood categories as perfect substitutes.

D Appendix tables

Table D1: Efficiency units estimation

More than high school	0.145
	(0.0003)
College	0.434
	(0.0004)
Age	0.061
	(0.0001)
Age^2	-0.001
	(0.0000)
Experience	0.021
	(0.0001)
Experience^2	-0.000
	(0.0000)

This table gives the estimated coefficients on education (omitted category is high school diploma or less), a quadratic in age, and a quadratic in experience. Regressions also include hours bins interacted with the fraction of the year worked, occupation fixed effects at the 3-digit ISCO level, and year fixed effects. Standard errors are in parenthesis.

Table D2: Griliches detail

$\phi_F - 1$ (female)	-0.0482
	(0.006)
$\phi_R - 1$ (married)	0.298
	(0.007)
$\phi_P - 1$ (35-54 year old)	1.038
	(0.015)
$\phi_O - 1$ (55 and older)	0.709
	(0.016)
$\phi_N - 1$ (unskilled laborers)	0.135
	(0.013)
$\phi_S - 1$ (white collar, technical, and sales workers)	0.497
	(0.009)
$\phi_C - 1$ (high skilled workers)	0.565
	(0.016)
α_1	1.192
	(0.011)
α_2	0.421
	(0.006)
α_3	-0.007
	(0.000)
α_4	-0.089
	(0.001)
α_5	0.054
	(0.002)

This table provides details of estimates using specification (2) for constructing $\mathcal{L}$ in the production function (4).

Table D3: Gender productivity gap by occupation and parenthood

	Agriculture	Other	Military	Unknown occupation
$\beta^o_{F,nc}/\beta^o_M$	0.756	1.008	0.535	1.180
	(0.174)	(0.036)	(0.300)	(0.012)
$\beta^o_{F,c}/\beta^o_M$	0.519	0.561	-0.130	0.854
	(0.110)	(0.019)	(0.007)	(0.006)
N	37,326	1,720,432	30,207	4,194,407

This table gives the ratio of relative productivity coefficients by occupation for the three omitted occupation categories in Table 10. These are omitted because the sample size is small and/or the categories are not informative. The first row of coefficients is the relative productivity of women without children relative to men and the second row of coefficients is the relative productivity of women with children compared to men in the same occupation (o). o is the occupation listed in the column heading and modeled as perfect substitutes in a translog production function with the occupations in 10. Bootstrapped standard errors are in parentheses.

Table D4: Means of occupation characteristics

	Fraction male	Average tenure	Log average hours	Flexibility index
Mean	0.652	4.832	7.584	-0.010
Standard Deviation	0.236	1.471	0.117	0.711
N occupations	141	134	105	103

This tables summarizes the means and standard deviations of occupation characteristics. Included are occupation-specific measures of (1) the fraction of the occupation which is male in the Danish administrative data used in this paper, (2) the average tenure at a given firm of men in the occupation in the Danish administrative data used in this paper (3) the log of average hours worked in the occupation in the ACS, replicated from Erosa et al. (2022) and matched to Danish occupation codes following Humlum & Meyer (2020), (4) the flexibility index based on Goldin (2014) and extended in Bang (2022) and applied to Danish occupation codes/population distributions. Observation counts differ when some characteristics cannot be disclosed (cell sizes are too small, as is the case with average tenure among men), or when there is imperfect linking across datasets (the second two categories use ACS data translated to Danish ISCO categories). Regressions using these estimates are weighted by the size of the occupation, as some occupations are extremely small.

Table D5: Productivity gap by occupation characteristics

	(1)	(2)	(3)	(4)	(5)
Fraction male	-0.225				-0.251
	(0.182)				(0.285)
Average tenure		-0.039			-0.022
		(0.038)			(0.061)
Log average hours			0.281		0.479
			(0.575)		(0.833)
Flexibility index				0.107	0.053
				(0.091)	(0.122)
Number of occupations	141	134	105	103	100

Note: This table displays the results of a regression of relative productivity of women (compared to men) within an occupation on occupation-specific measures of (1) the fraction of the occupation which is male in the Danish administrative data used in this paper, (2) the average tenure at a given firm of men in the occupation in the Danish administrative data used in this paper (3) the log of average hours worked in the occupation in the ACS, replicated from Erosa et al. (2022) and matched to Danish occupation codes following Humlum & Meyer (2020), (4) the flexibility index based on Goldin (2014) and extended in Bang (2022) and applied to Danish occupation codes/population distributions. Column (5) includes all occupation characteristics simultaneously. Regressions are weighted according to the size of the occupation.

Table D6: Productivity gap for mothers and women without children, by occupation characteristics

	All occupations		Excluding outliers	
	$\beta_{F,c}$	$\beta_{F,nc}$	$\beta_{F,c}$	$\beta_{F,nc}$
Fraction male	-0.233	-0.298	0.001	-0.265
	(0.175)	(0.196)	(0.135)	(0.177)
Average tenure	0.013	-0.078	-0.003	-0.047
	(0.038)	(0.042)	(0.028)	(0.039)
Log average hours	0.156	0.712	0.311	0.538
	(0.514)	(0.576)	(0.391)	(0.520)
Flexibility index	0.059	0.007	0.048	-0.021
	(0.075)	(0.084)	(0.056)	(0.076)
Number of occupations	94	93	78	76

Note: This table displays the results of a regression of the estimated productivity gap within occupations on occupation-specific measures. The second row of the table specifies the outcome variable in a given regression, where $\beta_{F,c}$ is the productivity gap for women with children estimated at the occupation level and $\beta_{F,nc}$ is the productivity gap for women without children estimated at the occupation level. The first two rows include all observations, while the second two rows drop outliers with extremely high or extremely low estimates of the productivity gap. The results for mothers are more sensitive to the sample. The covariates include the fraction of the occupation which is male in the Danish administrative data used in this paper, the average tenure at a given firm of men in the occupation in the Danish administrative data used in this paper, the log of average hours worked in the occupation in the ACS, replicated from Erosa et al. (2022) and matched to Danish occupation codes following Humlum & Meyer (2020), and the flexibility index based on Goldin (2014) and extended in Bang (2022) and applied to Danish occupation codes/population distributions. Regressions are weighted according to the size of the occupation.