Bargaining, Sorting, and the Gender Wage Gap: Quantifying the Impact of Firms on the Relative Pay of Women*

David Card, UC Berkeley and NBER
Ana Rute Cardoso, IAE-CSIC and Barcelona GSE
Patrick Kline, UC Berkeley and NBER

May 2015

Abstract

There is growing evidence that firm-specific pay premiums are an important source of wage inequality. These premiums will contribute to the gender wage gap if women are less likely to work at high-paying firms or if women negotiate (or are offered) worse wage bargains with their employers than men. Using longitudinal data on the hourly wages of Portuguese workers matched with income statement information for firms, we show that the wages of both men and women contain firm-specific premiums that are strongly correlated with simple measures of the potential bargaining surplus at each firm. We then show how the impact of these firm-specific pay differentials on the gender wage gap can be decomposed into a combination of sorting and bargaining effects. We find that women are less likely to work at firms that pay higher premiums to either gender, with sorting effects being most important for low- and middle-skilled workers. We also find that women receive only 90% of the firm-specific pay premiums earned by men. Importantly, we find the same gender gap in the responses of wages to changes in potential surplus over time. Taken together, the combination of sorting and bargaining effects explain about one-fifth of the cross-sectional gender wage gap in Portugal.

JEL Codes: J16, J31, J71

*We are grateful to five anonymous referees, and to Laura Giuliano, Michael Ransom, Jesse Rothstein, Andrea Weber, seminar participants at California Polytechnic State University, Harvard University, RAND, University College Dublin, the University of Mannheim, the University of Potsdam and the University of Venice for many helpful comments and suggestions. We are also grateful to Alex Fahey for her expert assistance. We thank the Spanish Ministry of the Economy and Competitiveness (grant CO2012-38460) and the Severo Ochoa Programme for Centres of Excellence in R&D (SEV-2011-0075) as well as the Center for Equitable Growth and the Center for Labor Economics at UC Berkeley for generous funding support. An earlier version of this paper circulated under the title "Bargaining and the Gender Wage Gap: A Direct Assessment."

Despite rapid advances in the educational attainment and job experience of women, there is still a substantial gender wage gap in most countries (e.g., OECD, 2015a). Though some analysts argue that the gap is primarily driven by male-female differences in effective productivity (e.g., Mulligan and Rubinstein, 2008), a more expansive view, consistent with models of frictional labor markets (e.g., Manning, 2011) is that equally productive men and women also face different job prospects and strike different wage bargains with their employers. Concern for such possibilities permeates the legal system in both the U.S. and the E.U., where laws require equal access to job openings for men and women and equal treatment of male and female employees within a firm.

Two long-established strands of research suggest that firm-specific pay policies may in fact be important for understanding the gender wage gap. One focuses on potential differences in the fractions of men and women employed at different firms (Blau, 1977; Groshen 1991; Petersen and Morgan, 1995), and in the rates that men and women move to higher-paying jobs (e.g., Loprest, 1992; Hospido, 2009; Del Bono and Vuri, 2011). The other emphasizes the wage-setting power of firms and the possibility that women are offered (or negotiate) systematically lower wages at a given firm.[1] These studies point to two complementary channels for generating gender disparities: a *sorting* channel that arises if women are less likely to be employed at higher-wage firms, and a *bargaining* channel that arises if women obtain a smaller share of the surplus associated with their job.

In this paper we provide the first comprehensive analysis of the impact of firm-specific pay premiums on the gender wage gap, using matched worker-firm data from Portugal merged with financial information for employers.[2] Building on a simple rent-sharing model, we develop an approach to measuring the sorting and bargaining channels via an Oaxaca-style decomposition (Oaxaca, 1973; Fortin, Lemieux and Firpo, 2011) of gender-specific firm wage effects. Like Abowd, Kramarz and Margolis (1999) – hereafter, AKM – our model includes fixed effects for individual workers and fixed effects for employers that measure the wage premium paid by each firm relative to some reference firm or group of firms. A key issue for assessing the contribution of the bargaining channel is the need to define the relevant reference group for each gender (Oaxaca and Ransom, 1999). We use the mapping between measures of the potential surplus at each firm and the estimated wage premiums to define a normalization that yields a lower-bound estimate of the differential bargaining power of women. We verify our results using an a priori assumption on the degree of rents available in the hotel and restaurant sector – a traditional low-wage industry.

Since our analysis builds directly on AKM's assumption that different firms pay different wage premiums relative to the overall labor market, we begin our empirical analysis by providing some descriptive evidence on the stability of these premiums, and on the exogenous mobility assumptions needed to measure them via ordinary least squares (OLS) methods. Corroborating earlier exercises by

[1]Robinson's (1933) monopsonistic wage-setting model was motivated in part by trying to explain why a firm might pay lower wages to women than men. As pointed out by Barth and Dale-Olsen (2009), this framework has been largely ignored in the gender wage literature. Lang and Lehmann (2012) discuss models of employer wage setting in a racial discrimination context. Bertrand (2011) presents a review of recent work emphasizing the relative negotiating abilities of men and women.

[2]Nekby (2003) relates male and female wages to measured profitability in a cross section of Swedish firms, but does not address the potential selectivity issues caused by non-random sorting of men and women with different unobserved skill characteristics to more profitable firms. Barth and Dale-Olsen (2009) examine firm-specific gender wage differences in a monopsony framework.

Card, Heining and Kline (2013) with German data, and by Macis and Schivardi (2013) with Italian data, we find that these assumptions are approximately satisfied for both men and women in Portugal. Comparing the average wage gains and losses for men and women who move between matched sets of firms we also show that women benefit less from firm-to-firm mobility than men.

We then estimate separate AKM models for male and female workers in Portugal. We find that firm-specific pay premiums explain about 20% of wage variation among both men and women, while positive assortative matching (i.e., the positive covariation between worker and firm effects) explains another 10%. The pay premiums offered to men and women are highly correlated across firms. We use a simple decomposition method to assess the contribution of firm-specific wage setting to the overall gender wage gap, and to the wage gap among workers with different levels of age and education, and in different occupations and industries. Overall we find that the under-representation of women at firms that offer higher wage premiums for both gender groups – the sorting effect – explains about 15% of the overall gender gap in Portugal. Another 5% is attributable to the fact that women gain less than men from higher-wage firms – the bargaining effect. We find that sorting effects rise with age, and are more important among less educated workers, while bargaining effects are larger for highly-educated workers. Both components vary by occupation, with the largest contribution of sorting for traditional skilled and semi-skilled blue collar jobs and clerical jobs. To verify that the relative bargaining effect is really due to gender, and not to differential wage setting for traditionally female occupations, we fit separate AKM models for workers in mainly female and mainly male occupations. We also conduct a separate analysis of workers with high school or greater education to avoid the potentially confounding effects of the minimum wage.

We then narrow our focus to the component of the firm-specific wage premiums paid to men and women that is directly related to a simple proxy for the average bargaining surplus available at each firm. We find that women's wages are only 90% as responsive to observable measures of the surplus per worker as men's, and we can easily reject the hypothesis of equal responsiveness, thus confirming that women gain a smaller share of firm-wide rents than their male co-workers. We also confirm that women are under-represented at firms with higher measured surplus levels. Bargaining and sorting based on the observable component of surplus account for about 80% of the overall impact of firm-specific pay premiums on the gender wage gap.

As a final step in our analysis we examine the effects of changes in the average surplus per worker on the wages of men and women who remain with the firm over a multi-year period. This approach, which mirrors the design employed in the modern rent-sharing literature (e.g., Guiso, Pistaferri, and Schivardi, 2005; Carlsson, Messina and Skans, 2011; Card, Devicienti and Maida, 2014), uses an entirely different component of wage variation than our analysis of firm-specific pay premiums. Reassuringly, we obtain a nearly identical 90% estimate of women's relative bargaining power, suggesting that a simple wage setting model with gender-specific bargaining parameters can successfully explain both the between-firm structure of relative wages for men and women, and the variation over time at a given firm in male and female wages.

1 Firm-specific Determinants of the Gender Wage Gap

In traditional competitive labor market models, wages are determined by market-level supply and demand factors rather than by the wage-setting policies of particular firms.[3] This perspective is central to Becker's (1957) model of employer-based discrimination, which asserts that the *market-wide* discriminatory wage premium depends on the preferences of the marginal employer of women.[4] Building on this framework, most studies of the gender wage gap focus on measured skill differences between men and women and attribute any unexplained component to a combination of discriminatory factors and unobserved skill gaps (see Altonji and Blank, 1999 and Blau and Kahn, 2000, for reviews).[5]

Despite the market-level focus of most economic studies, legislation aimed at eliminating gender discrimination is primarily directed at firms. In the U.S., for example, the Equal Pay Act requires that employers offer equal pay to men and women for "substantially equal" work, while Title VII of the Civil Rights Act prohibits firms from discriminating against women (and other protected groups) in decisions over hiring, layoffs, and promotions. In Portugal, the 1976 version of the Constitution guarantees equality of access to jobs (article 52), while also banning discrimination (article 53) and ensuring the right to equal pay for equal work (article 59).

To the extent that firms have some control over the wages offered to a given worker, the average wages of women relative to men will be affected by two factors. The first is whether firms that tend to pay higher wages are more or less likely to hire women. The second is whether firms offer *different* average wage premiums for men and women relative to the "market" (or a reference employer).

The potential importance of the between-firm sorting channel to the gender wage gap was noted by Blau (1977), who used wage data for white collar workers at different establishments in three cities and concluded that establishments with higher average wages tended to employ fewer women. Subsequent research, including Groshen (1991), Petersen and Morgan (1995), and Bayard et al. (2003), suggests that the differential sorting of females and males to higher and lower paying workplaces explains some fraction of the gender wage gap.[6] A concern with these studies is that they do not control for unobserved characteristics of workers, thus confounding segregation by ability with segregation by gender. This concern is addressed by studies of interfirm mobility (Loprest, 1992; Hospido, 2009; Del Bono and Vuri 2011) which show that women are about as likely to move between firms as men, but experience smaller average wage gains with each move. Nevertheless, these studies cannot distinguish between the hypothesis that women are less likely to find jobs at higher-paying firms and the alternative that the wage gain for a given firm-to-firm transition is smaller for women than men.

[3] Wages can vary across firms if there market-based compensating differentials for firm-wide amenities or disamenities, such as long hours of work (Bertrand, Goldin and Katz, 2010). We examine the correlation between firm-specific wage variation and hours of work later in the paper but find no evidence of hours-related compensating differences.

[4] See Charles and Guryan (2008, 2011) for a recent application to the black-white wage gap and reviews of related work.

[5] An interesting exception is audit-based studies of potential discrimination (e.g., Heckman and Siegelman, 1993; Neumark, Bank and Van Nort, 1996; Bertrand and Mullainathan, 2004), which focus on the hiring practices of individual employers.

[6] One piece of evidence suggesting that the exclusion of women is driven by employer preferences comes from studies of banking (Ashenfelter and Hannan, 1986; Black and Strahan, 2001) which find that deregulation led to a rise in the share of female employees in the industry. Neumark et al. (1996) also report that higher-wage restaurtants are less likely to interview female applicants.

Cardoso, Guimarães and Portugal (2012) also focus on differential sorting using an AKM style model but imposing the assumption that the firm effects are the same for men and women.

The possibility that equally productive women and men are paid differently by firms with some wage-setting power was suggested by Robinson (1933) in her seminal analysis of imperfect labor markets, and arises in wage posting models in which women's and men's turnover rates are differentially responsive to firm-specific wage premiums, or in search and matching models in women and men have different relative bargaining power.[7] The relative bargaining power interpretation is also emphasized in the social psychology literature, which argues that women are less likely to initiate negotiations with their employers (Babcock and Laschever, 2003; Bowles, Babcock and Lai, 2007), and are on average less successful negotiators.[8]

2 Institutional Setting and Data Overview

Our analysis relies on an annual census of employees in Portugal that includes data on earnings and hours of work, as well as firm-specific information that allows us to link workers to the income statements of their employers. Although our focus on Portugal is driven by the quality of these data, three features suggest that our findings may be broadly generalizable to other settings. First, Portuguese women have relatively high labor force participation rates. Fifty eight percent of adult women in the country were in the labor force in 2010 (ILO, 2012), comparable to the rates in the U.S. and Northern Europe. Second, the vast majority of women in Portugal (over 90% of those in private sector jobs) work full time, reducing concerns that the gender wage gap is confounded by differences between full-time and part-time jobs. Third, the gender wage gap in Portugal is within a few percentage points of the gaps in the U.S. and U.K., and is very close to the OECD-wide average.[9]

Most private sector jobs in Portugal are covered by sector-wide collective agreements negotiated by employer associations and trade unions (ETUI, 2015). Bargaining is synchronized and most wage clauses are renegotiated annually in January. Since these contracts set pay on a gender-neutral basis they arguably exert some equalizing effect on the relative pay of women (Blau and Kahn, 2003). On the other hand, firms have wide latitude in assigning employees to job categories, and most workers also earn substantial wage premiums over the base pay rates for their job category (Cardoso and Portugal, 2005). The minimum wage is also relatively high in Portugal, potentially raising women's wages relative to men's – an issue we address in detail below. Nevertheless, Portugal has very high levels of overall wage inequality, suggesting that wage setting is relatively unconstrained by institutional forces.[10]

[7] Manning (2011, section 3) shows that these two alternatives are observationally equivalent.

[8] Stuhlmacher and Walters (1999) present a meta-analysis of lab-based studies of the effect of gender in bargaining, and conclude that on average women obtain a smaller share of the surplus than men. Save-Soderbergh (2007) found that female college graduates who were asked to submit a salary demand at the start of their first job job tended to ask for lower salaries and ended up receiving lower salaries than men.

[9] The gender gap in median full time earnings was 16% in Portugal, 19% in the U.S., and 16% on average across 26 OECD countries (OECD, 2012).

[10] Martins and Pereira (2004) tabulate the 90/10 gap in hourly wages and in the returns to education for 16 countries, including the U.S., the U.K. and other European countries, and find that Portugal is highest in both measures of inequality.

2.1 Data Sources

Our main data source is Quadros de Pessoal (QP), a census of private sector employees conducted each October by the Portuguese Ministry of Employment. Firms with at least one paid employee are required to submit information on their full workforce as of the survey reference week. Government employees and independent contractors are excluded from coverage, as are people who are unemployed or out of the labor force in the survey week.[11] Over our 2002-2009 sample period we have information on roughly 4 million workers who are observed between 1 and 8 times, with firm and establishment identifiers for their jobs in the survey week. Since our financial data are firm-based, we aggregate establishments to the firm-level for the small fraction (4%) of multi-plant firms.

The QP asks employers to report each employee's gender, education, occupation, date of hire, regular monthly salary, regular wage supplements, and hours of work. Information is also collected on the industry, location, and founding date of the firm, as well as gross sales in the preceding calendar year. We construct hourly wages by dividing the sum of a worker's base salary plus any regular earnings supplements by his or her normal hours of work, yielding a "straight time" hourly wage.[12] The availability of hours information is a unique strength of the QP and allows us to address concerns that the gender wage gap is driven in part by differences in hours of work by men and women (e.g., Wood et al., 1993; Bertrand et al, 2010).

We augment this information with financial data from the "SABI" (Sistema de Analisis de Balances Ibericos) database. Businesses in Portugal are required to file income (or profit and loss) statements and balance sheet information annually with the Conservatoria do Registo Comercial.[13] These reports are publicly accessible and are collected by financial service firms and assembled into the SABI database by Bureau van Dijk. Information in SABI includes the firm's name, address, industry, founding date, and total employment, as well as income statement and balance sheet items. SABI data are available from 2000 onward, but coverage of the database was limited before 2006.

Since the QP does not include firm names or tax identifiers we use a combination of variables that are reported in both QP and SABI to match the data sets. Specifically, we use location, industry, firm creation date, annual sales, and year-end shareholder equity as matching variables. As described in the Data Appendix, we successfully match about 53% of firms that appear in the QP between 2002 and 2009 to a firm with at least one year's information in SABI. More information on our matching procedure and the match rates for various subgroups is presented in the Data Appendix. Overall we have current-year employer financial data for about 66% of the person-year observations in our QP sample from 2006-2009.

[11] Firm owners and employees on temporary leave are included in the data set but do not report wages, and so are excluded from our analysis.

[12] Normal hours are the hours specified by the prevailing sectoral or firm-specific contract. Regular earnings supplements are payments such as meal allowances that are received regularly.

[13] Based on informal discussions with firm owners we believe that the penalties for non-filing are small, presumably accounting for missing data for many firms.

2.2 Descriptive Overview

We begin with a brief overview of the differences between male and female employees in Portugal. We focus on individuals who are between 19 and 65 years of age, have more than one year of potential labor market experience, and worked as a paid employee in the QP reference week. Our primary analysis sample – described in columns 1 and 2 of Table 1 – contains annual wage observations from 2002 to 2009 for 2.1 million men and 1.7 million women.[14]

Comparisons between the two columns show that female workers in Portugal are slightly younger than their male counterparts but are more likely to have completed secondary or tertiary education. Despite the education gap, women earn about 18% less per hour than men – very similar to the gender gap in median hourly wages in the U.S. in 2007 (EPI, 2010).[15] Women also work slightly fewer hours per month, though the 3% difference is small by international standards.[16] The dispersion in monthly hours is larger for women than men while the dispersion in hourly wages is smaller for women. Thirty-five percent of both male and female employees work in the Lisbon area, another 13 percent work in the Porto area, and the remainder work in smaller cities and rural areas.

Comparing average workplace characteristics, women work at slightly larger firms than men (858 employees vs. 730), a feature that is also true in the U.S. and the U.K.[17] More striking is the difference in the share of female employees at women's and men's workplaces – 70% vs. 24%. This gap indicates that there is significant gender segregation across firms.[18] Indeed, about 21% of men work at all-male firms, while 19% of women work at all-female firms.[19] The presence of single-gender firms poses a problem for assessing the role of firms in the gender wage gap, since we cannot observe the wages that would be offered to women at all-male firms, or to men at all-female firms. For most of our analysis below we therefore limit attention to firms that hire at least one worker of each gender at some point in our sample period. The wage gap between men and women at single-gender firms is relatively small (9%), so eliminating employees at these firms leads to a slightly larger gender gap in the remaining subsample than in the labor market as a whole.

An important issue for an analysis of between-firm wage differentials is the rate of job mobility, since these differentials are identified by the wage changes of job movers. Appendix Table B1 shows

[14]See the Data Appendix for details on the derivation of this sample, and comparisons with the overall population of 16-65 year old workers in the QP. In the small number of cases where an individual is employed at two or more firms in the reference week, we assign them to their highest earnings job.

[15]The wage gap narrowed over our sample period, falling from 21% in 2002 to 16% in 2009 – see Appendix Figure B1.

[16]Data reported by the OECD (2012) for Portugal (based on labor force survey data that include government and independent contract workers excluded from QP) show part-time employment rates for men and women of 8% and 14%, respectively. The same source shows part-time employment rates for men and women in the US of 8% and 17%.

[17]Papps (2012) and Mumford and Smith (2008) report roughly 10% larger workplace sizes for women than men in the U.S. and U.K., respectively.

[18]Hellerstein et al. (2008) report that in 2000, the average fractions of female co-workers for female and male workers at larger establishments in the U.S. were 61% and 41%, respectively. Mumford and Smith (2008, online Appendix Table A2) report that in the U.K. in 2004 the average fraction of female employees at the workplace was 70% for women and 34% for men. These comparisons suggest that Portuguese firms may be more segregated by gender than those in the U.S. or U.K.

[19]About 20% of men and women at single-gender firms are the only employee at their workplace. Construction and trade account for 43% and 20%, respectively, of the person-year observations at all-male jobs. All-female workplaces are prevalent in trade (23% of person-years at all female firms), health services (17%), hotels (14%), and textiles (13%). Mean log wages of workers at single gender firms are relatively low: 1.28 for men and 1.19 for women.

that the distributions of the number of jobs held by men and women in our overall QP sample are very similar. Approximately 73% of men and 74% of women hold only 1 job during our sample period; 19% of both groups have 2 jobs; and 6% have 3 jobs. The remaining 2% of men and 1% of women hold 4-8 jobs. We also examined survival rates of new jobs that are observed starting during our sample period, and found that these are very similar. As shown in Appendix Figure B2, about 40% of new jobs last less than 1 year for both groups. We conclude that job mobility rates are very similar for women and men in Portugal.

3 Modeling Framework

In this section we present a very simple model that allows us to evaluate the effect of firm-specific pay premiums on the observed wages of women and men. Assume that we observe point-in-time wages for workers (indexed by $i \in \{1, ..., N\}$) in multiple periods (indexed by $t \in \{1, ..., T\}$). We denote worker i's gender by $G(i)$ which takes on values in the set $\{F, M\}$, and the identity of his or her employer in a given year by $J(i,t)$ which takes on values in the set $\{1, ..., J\}$. We refer to a particular gender as g and a particular firm as j.

We posit a wage-setting model in which the logarithm of the real wage earned by individual i in period t (w_{it}) is given by:

$$w_{it} = a_{it} + \gamma^{G(i)} S_{iJ(i,t)t}. \tag{1}$$

Here, a_{it} represents the outside option available to worker i in period t (e.g. the wage in self employment), $S_{iJ(i,t)t} \geq 0$ is the match surplus between worker i and firm $J(i,t)$ in period t, and $\gamma^g \in [0,1]$ is a gender-specific share of the surplus captured by a worker of gender $g \in \{F, M\}$. We are specifically interested in the question of whether women get a smaller share of the surplus associated with their job (i.e. $\gamma^F < \gamma^M$).

We assume that $S_{iJ(i,t)t}$ can be decomposed into three components:

$$S_{iJ(i,t)t} = \bar{S}_{J(i,t)} + \phi_{J(i,t)t} + m_{iJ(i,t)}. \tag{2}$$

The first term, $\bar{S}_{J(i,t)}$, captures time-invariant factors like market power or brand recognition that raise the average surplus for all employees at the firm. The second component, $\phi_{J(i,t)t}$, represents time-varying factors that raise or lower the average surplus for all employees. The third component, $m_{iJ(i,t)}$, captures a person-specific component of surplus for worker i at his or her current employer, attributable to idiosyncratic skills or characteristics that are particularly valuable at this job.

We assume that the outside option a_{it} can be decomposed into a permanent component α_i (due, for example, to ability or general skills), a time-varying component associated with an observed set of characteristics X_{it} (e.g., labor market experience and changing returns to education), and a transitory component ε_{it}:

$$a_{it} = \alpha_i + X_{it}' \beta^{G(i)} + \varepsilon_{it}, \tag{3}$$

where β^g is a gender specific vector of coefficients.

Equations (1) through (3) imply the wage of worker i in period t can be written as:

$$w_{it} = \alpha_i + \psi^{G(i)}_{J(i,t)} + X'_{it}\beta^{G(i)} + r_{it}, \tag{4}$$

where $\psi^{G(i)}_{J(i,t)} \equiv \gamma^{G(i)}\bar{S}_{J(i,t)}$ and $r_{it} \equiv \gamma^{G(i)}\left(\phi_{J(i,t)t} + m_{iJ(i,t)}\right) + \varepsilon_{it}$ is a composite error. Equation (4) is consistent with an additive "two-way" worker-firm effects model of the type considered by Abowd, Kramarz and Margolis (1999) and many subsequent authors, with person effects, *gender-specific* firm effects, and gender-specific returns to the covariates X_{it}. We use this model as the basis for our main analysis, though as explained below, we also explore the possibility that the share of surplus received by workers varies between occupations – specifically, between "typically female" occupations and "typically male" occupations.

3.1 Exogeneity

We estimate models based on equation (4) by OLS, yielding estimated gender-specific effects for each firm. For these estimates to be unbiased, the following orthogonality conditions must hold:

$$E\left[\left(r_{it} - \bar{r}_i\right)\left(D^j_{it} - \bar{D}^j_i\right)|G\left(i\right)\right] = 0 \;\forall j \in \{1, ..., J\}\,, \tag{5}$$

where $D^j_{it} \equiv 1\left[J\left(i,t\right) = j\right]$ is an indicator for employment at firm j in period t and bars over variables represent time averages. To gain some insight into the restrictions implied by equation (5), it is useful to consider the special case where $T = 2$. With two periods, fixed effects estimation is equivalent to first differences estimation and (5) reduces to:

$$E\left[\left(r_{i2} - r_{i1}\right)\left(D^j_{i2} - D^j_{i1}\right)|G(i)\right] = 0 \;\forall j \in \{1, ..., J\}. \tag{6}$$

Using the fact that $\left(D^j_{i2} - D^j_{i1}\right)$ takes on values of $+1$ for workers who move to firm j in period 2, -1 for those who leave firm j in period 1, and 0 for all others, we can write:

$$\begin{aligned} E\left[\left(r_{i2} - r_{i1}\right)\left(D^j_{i2} - D^j_{i1}\right)|G(i)\right] \;\; = \;\; & E\left[r_{i2} - r_{i1}|D^j_{i2} = 1, D^j_{i1} = 0, G(i)\right] \\ & \times P\left(D^j_{i2} = 1, D^j_{i1} = 0|G\left(i\right)\right) \\ & - E\left[r_{i2} - r_{i1}|D^j_{i2} = 0, D^j_{i1} = 1, G(i)\right] \\ & \times P\left(D^j_{i2} = 0, D^j_{i1} = 1|G\left(i\right)\right). \end{aligned}$$

The term $E\left[r_{i2} - r_{i1}|D^j_{i2} = 1, D^j_{i1} = 0, G(i)\right]$ is the mean change in the unobserved wage determinants for *joiners* of firm j, while the term $E\left[r_{i2} - r_{i1}|D^j_{i2} = 0, D^j_{i1} = 1, G(i)\right]$ is the corresponding change for *leavers* of this firm. Hypothetically, it is possible that these two terms are roughly comparable in magnitude since the decision to leave one firm is a decision to join another. In such a case, the mean bias associated with joiners and leavers would cancel whenever the number of firm joiners and leavers is equal, as would occur when the firm is in an employment steady state. However, while joining and leaving firms may yield similar average biases, the joiner and leaver bias associated with any particular firm may be quite different, which would lead to a violation of (6).

Since

$$r_{i2} - r_{i1} = \gamma^{G(i)} \left[\phi_{J(i,2)2} - \phi_{J(i,1)1} + m_{iJ(i,2)} - m_{iJ(i,1)}\right] + \varepsilon_{i2} - \varepsilon_{i1},$$

there are three channels though which the changes may be related to firm specific mobility. The first is a connection between firm-wide shocks ϕ_{jt} and mobility rates. For example, workers may be more likely to leave firms that are experiencing negative shocks and join firms that are experiencing positive shocks. If this is true, then we would expect to see a systematic "Ashenfelter dip" in the wages of leavers just prior to their exit, and unusual wage growth for recent joiners. We look for such patterns below and find no evidence that they are present in the data.

A second potential channel arises if mobility is related the idiosyncratic match effects (m_{ij}). Many search and matching models *assume* that workers search over jobs that differ by a match effect in pay. An implication is that the wage gains of movers will overstate the gains for a typical worker. For example, suppose that firm A offers a 10% larger *average* wage premium than firm B. If mobility is independent of the match effects, then movers from firm B to firm A will experience a 10% average wage *gain*, while movers from firm A to firm B will experience a 10% average wage *loss.* If instead mobility is based in part on comparative advantage then the expected wage losses associated with moving from A to B will tend to be offset by an improvement in match effects. In the limit, if all firm transitions are voluntary and selection is based solely on the match components, *all* moves will lead to wage gains, as in the dynamic matching model of Eeckhout and Kirchner (2011). In our analysis below we examine workers moving in opposite directions between groups of high and low wage firms, and find that their wage changes exhibit the approximate symmetry (i.e., equal magnitude and opposite sign) predicted by an additive model with exogenous mobility. This symmetry is inconsistent with selection based on the match component of wages.

A third channel arises if the direction of firm-to-firm mobility is correlated with the transitory wage shock ε_{it}. For example, a worker who is performing well and receiving promotions may be more likely to move to a higher wage firm, while workers who are stalled in their job may be more likely to move down the job ladder to a lower-paying firm. Systematic mobility of this form implies that people moving to higher-wage firms will have different trends prior to moving than those who move to lower-wage firms. Again, in our analysis below we find no evidence for any of these predictions.

What drives firm-to-firm mobility if it is *not* related to the elements in r_{it}? The most straightforward explanation is that worker-firm matching is based on a combination of the permanent component of worker ability (the α_i component in equations 3 and 4) and the average wage premiums offered by firms. Skilled workers, for example, are more likely to engage in on-the-job search (Pissarides and Wadsworth, 1994; Hall and Krueger, 2012) suggesting that they will be more likely to move to high-wage firms over time. Skilled workers also may have networks of friends and family members that are more likely to work at high wage firms, leading to network-based sorting (as in Kramarz and Skans, 2013). These forms of sorting create no bias for our estimation strategy because we condition on time-invariant worker and firm characteristics. Finally, sorting based on non-wage dimensions such as the location of the firm or the firm's recruiting effort creates no bias provided that these factors are uncorrelated with the time varying error component in (4).

3.2 Normalization

As explained by Abowd, Creecy and Kramarz (2002) the firm effects in a two-way fixed effects model such as (4) are only identified within a "connected set" of firms linked by worker mobility. In our analysis below we limit attention to workers and firms in the largest connected set for each gender. Even within these sets we still require a linear restriction to normalize the firm effects, since the wage premium for any given firm is only identified relative to a reference firm or set of firms.

According to our model the true firm effects for each gender are non-negative, and will be zero at firms that offer no surplus above an employee's outside option. We therefore normalize the firm effects by setting the average wage premium for a set of "low-surplus" firms to 0. More precisely, letting $\overline{S}_j^o$ denote an observed measure of average surplus per worker at firm j, we assume that:

$$E\left[\psi^g_{J(i,t)}|\overline{S}^o_{J(i,t)} \leq \tau\right] = 0 \ , \ g \in \{F, M\} \tag{7}$$

where τ is a threshold level such that firms with observed surplus per worker below τ pay zero rents on average. If (7) is correct then imposing this condition will yield a set of normalized firm effects that coincide with the true firm effects (apart from sampling errors). Otherwise, the normalized effects for each gender will be equal to the true firm effects, minus the average value of the firm effects for that gender group at firms with $\overline{S}_j^o < \tau$.

As discussed below, we use mean log value added per worker for all years the firm is observed in the SABI data set as our main measure of surplus per worker. Value added is reported for most firms, and is constructed as the sum of wage payments, other labor costs, depreciation, interest costs, taxes, and profits (i.e., the sum of payments to labor and capital, plus taxes).[20] Under standard assumptions, value added will be equal to revenues minus the costs of all intermediate inputs. We also repeat our analysis using mean log sales per worker as an alternative measure of surplus and obtain very similar results.

While the normalized effects could, in principle, be estimated in a single step, we opt instead for a two-step approach. We first estimate the gender specific firm effects via unrestricted OLS, arbitrarily setting the effects for a particular large firm to 0. We then renormalize the effects by subtracting off the average value of the gender-specific firm effects at low surplus firms. We explain how we estimate the threshold τ in Section 5.3, below.

As a check on this procedure we normalize the firm effects by assuming that firms in the hotel and restaurant industry pay zero surplus on average. This assumption is motivated by the extensive literature on industry wage differences (e.g., Dickens and Katz, 1986; Krueger and Summers 1988) which suggests that these differentials are, at least in part, driven by rents. We observe that firms in the hotel and restaurant sector have the smallest wage premiums on average, so we simply assume that rents are on average zero in this sector.

[20]This is the standard national accounts definition (see e.g., Stassner and Moyer, 2002).

3.3 Decomposing the Effect of Firm-Level Pay Premiums

Equation (4) provides a simple framework for measuring the impact of firm-level pay premiums on the gender wage gap. Using *male* and *female* as shorthand for the respective conditioning events that $G(i) = M$ and $G(i) = F$, we can denote the average pay premium received by men as $E[\psi^M_{J(i,t)}|male]$ and the average premium received by women as $E[\psi^F_{J(i,t)}|female]$. As in the traditional Oaxaca wage decomposition (see e.g., Oaxaca, 1973; Fortin, Lemieux and Firpo, 2011), we can decompose the difference in pay premiums into a combination of bargaining power and sorting effects in either of two ways:

$$
\begin{aligned}
E[\psi^M_{J(i,t)}|male] - E[\psi^F_{J(i,t)}|female] &= E[\psi^M_{J(i,t)} - \psi^F_{J(i,t)}|male] \\
&\quad + E[\psi^F_{J(i,t)}|male] - E[\psi^F_{J(i,t)}|female] \qquad (8)\\
&= E[\psi^M_{J(i,t)} - \psi^F_{J(i,t)}|female] \\
&\quad + E[\psi^M_{J(i,t)}|male] - E[\psi^M_{J(i,t)}|female]. \qquad (9)
\end{aligned}
$$

The first term in equation (8) is the average bargaining power effect, calculated by comparing ψ^M_j and ψ^F_j across the distribution of jobs held by men. The second term in (8) is the average sorting effect, calculated by comparing the average value of ψ^F_j across the jobs held by men versus women. In the alternative decomposition (equation 9) the bargaining power effect is calculated using the distribution of jobs held by women, and the sorting effect is calculated by comparing the average value of the *male* pay premiums across jobs held by men versus women.

It is worth emphasizing that the estimated sorting effects in (8) and (9) are invariant to the particular normalization chosen for the firm effects. In contrast, the estimated bargaining effects depend on the normalization: subtracting different constants from the male and female effects will obviously lead to different values for the first line of either equation (8) or (9). Provided that the rents received by female workers at low surplus firms are no larger than the rents received by male workers at these firms, however, our choice of normalization will yield a lower bound estimate of the bargaining effect, and the overall decomposition will lead to a lower bound estimate of the effect of firm-specific pay premiums on the gender wage gap.

3.4 Relating the Estimated Firm Effects to Measures of the Bargaining Surplus

An alternative approach to measuring the sorting and bargaining components of the gender wage gap is to look directly at how the estimated wage premiums offered by a given firm vary with the measured surplus per worker at the firm. Specifically, building on our normalization approach, we assume that:

$$E\left[\bar{S}_{J(i,t)}|\overline{S}^o_{J(i,t)}\right] = \kappa \max\left\{0, \overline{S}^o_{J(i,t)} - \tau\right\}. \qquad (10)$$

In other words, actual average surplus per worker is linearly related to the deviation of the observed surplus measure from the threshold level τ for firms with $\overline{S}^o_j > \tau$, and is 0 otherwise. For simplicity we refer to the quantity $\max\left\{0, \overline{S}^o_j - \tau\right\}$ as firm j's "net surplus" NS_j. Given a value for τ (which we estimate in a prior step, as explained in Section 5.3) we can write:

$$\psi^{g}_{J(i,t)} \;=\; \pi^{g} NS_{J(i,t)} + \nu^{g}_{J(i,t)} \tag{11}$$

where $\pi^g \equiv \gamma^g \kappa$ and $E\left[\nu^{g}_{J(i,t)} | NS_{J(i,t)}, G(i)\right] = 0$. Notice that $\pi^F/\pi^M = \gamma^F/\gamma^M$. By taking the ratio of the estimated gender specific slopes after estimating equation (11) for male and female workers we obtain a direct estimate of the bargaining power ratio γ^F/γ^M.

Using this setup, we can decompose the difference in the average value of the first term of equation (11) for male relative to female workers as:

$$E[\pi^M NS_{J(i,t)}|male] - E[\pi^F NS_{J(i,t)}|female]$$

$$= (\pi^M - \pi^F)E[NS_{J(i,t)}|male] + \pi^F \begin{pmatrix} E[NS_{J(i,t)}|male] \\ -E[NS_{J(i,t)}|female] \end{pmatrix} \tag{12}$$

$$= (\pi^M - \pi^F)E[NS_{J(i,t)}|female] + \pi^M \begin{pmatrix} E[NS_{J(i,t)}|male] \\ -E[NS_{J(i,t)}|female] \end{pmatrix}. \tag{13}$$

Focusing only on the part of the firm surplus that is explained by our observed measure of net surplus, the contribution of the bargaining channel to the male-female wage gap is simply the difference in coefficients $(\pi^M - \pi^F)$, weighted by the measured net surplus at men's jobs (equation 12) or women's jobs (equation 13). The corresponding contribution of the sorting channel is the difference in average net surplus at men's jobs and women's jobs, weighted by either π^F (equation 12) or π^M (equation 13).

3.5 Within-Firm Changes in Wages Over Time

While our main focus is on gender differences in *between-firm* wage differentials, our model also implies that the wages of male and female employees who are observed working at the same firm over time will respond differently to *changes* in firm surplus. Define $S_{jt} \equiv \bar{S}_j + \phi_{jt}$ as the actual surplus per worker in period t, and S^o_{jt} as the observed surplus measure for firm j in year t. We assume that these are related by:

$$\begin{aligned} S_{jt} \;&=\; \lambda \max\left\{0, S^o_{jt} - \tau\right\} + \varsigma_{jt} \\ &\equiv\; \lambda NS_{jt} + \varsigma_{jt}, \end{aligned} \tag{14}$$

where the error ς_{jt} has mean zero when we condition on the firm's observed net surplus and the characteristics of workers observed working at the firm continuously between an initial period $t = 1$ and a later period $t = T$ (i.e., "stayers"). Using equation (4) we can therefore write:

$$\begin{aligned} &E\left[w_{iT} - w_{i1} | NS_{J(i,1)1}, NS_{J(i,1)T}, X_{i1}, X_{iT}, G(i), stayer\right] \\ &\quad = (X_{iT} - X_{i1})' \beta^{G(i)} + \theta^{G(i)}[NS_{J(i,1)T} - NS_{J(i,1)1}], \end{aligned} \tag{15}$$

where $\theta^g = \gamma^g \lambda$ and *stayer* is shorthand for the conditioning event that worker i is continuously employed at the same firm throughout the sample period. Estimating this equation by OLS separately by gender yields estimates of the slope parameters θ^M and θ^F which can be used to form another estimate of the relative bargaining power ratio γ^F/γ^M, based on the differential reactions of male and female wages to changes in surplus.

To actually estimate the relative bargaining power ratio (and its sampling error) we rely on the insight from our model that:

$$\frac{E\left[w_{iT} - w_{i1} - (X_{iT} - X_{i1})' \beta^F | female, stayer, J(i,1) = j\right]}{E\left[w_{iT} - w_{i1} - (X_{iT} - X_{i1})' \beta^M | male, stayer, J(i,1) = j\right]} = \frac{\gamma^F}{\gamma^M}$$

That is, the covariate-adjusted average wage changes of male and female stayers at the same firm are deterministically related by the gender bargaining power ratio. Given the small size of most firms in our sample, we estimate this relationship using a two-step instrumental variables (IV) procedure. For each gender, we regress the change in wages on covariates and firm dummies to obtain adjusted average firm wage changes by gender. We then regress the adjusted average change in female wages at each firm on the corresponding average male change using the change in measured surplus as an instrument and weighting by the total number of stayers at each firm. Similarity of this estimate, based on within-firm changes in wages and measured surplus, with the estimate from equation (11) based on between-firm variation in wages and surplus, provides support for the simple rent-sharing model specified by equations (1)-(3).

4 Descriptive Evidence on Firm-Specific Pay Premiums

Although the two-way effects model specified in equation (4) has been widely used over the past decade, the additive structure of the model and the restrictive assumptions needed for OLS estimation have been strongly criticized by some authors (e.g., Lopes de Melo, 2009; Eeckhout and Kirchner, 2011). Following Card, Heining and Kline (2013), we present some descriptive evidence on the patterns of wage changes for people who move between jobs with higher- and lower-paid co-workers. We document five basic facts that are all consistent with equation (4) and the exogenous mobility condition (5). First, men and women who move between jobs with higher- and lower-paid co-workers experience systematic wage gains and losses, suggesting that there are significant firm-specific pay premiums for both genders. Second, there is no indication that movers to firms with higher- or lower-paid co-workers experience differential wage trends prior to their move. Third, wage changes for people who move between firms with similarly-paid co-workers experience little or no excess wage growth relative to job stayers. Fourth, the gains and losses from moving between jobs with higher-paid and lower-paid co-workers are approximately symmetric, suggesting that the firm-specific pay premiums are additively separable (in logarithms) from other pay components and that mobility patterns are not driven by comparative advantage in wages. Fifth, women gain less than men from moving to jobs with more highly paid co-workers, as predicted by a rent-sharing model in which women get a smaller share of the rents than men.

We begin by selecting men and women from the overall analysis sample described in columns 1 and 2 of Table 1 who are employed at firms with at least one worker of each gender at some point in our sample period. We construct mean log co-worker wages for each person in each year and divide all jobs for both men and women into four quartiles of co-worker wages, excluding the small number of workers at firms with only one employee. Then we identify job changers who are observed for at least two years at their origin firm and two years at their destination firm. Finally, we construct average wages in the years before and after a move for male and female job-changers, classified into 16 cells based on the co-worker wage quartiles of their origin and destination jobs.

Figures 1 and 2 plot the wage profiles before and after the job change for men and women who moved from jobs in the lowest ($1st$) quartile of co-worker wages, and for those who moved from jobs in the highest ($4th$) quartile. The figures show that men and women who move from jobs with highly paid co-workers to jobs with poorly-paid co-workers experience large average wage losses, while those who move in the opposite direction experience large wage gains.[21] Moving within a quartile group, by comparison, is associated with relatively small wage changes. Moreover, although the *levels* of wages on the old job differ between people from the same origin quartile who move to different destination quartiles, the *trends* prior to moving are very similar across groups. Likewise, then trends after moving are very similar across groups. These observations imply that inter-firm mobility is correlated with the permanent component of individual wages (i.e., the α_i component of equation 4) but not with the transitory error components (i.e., ϕ_{jt} or ε_{it}).

Appendix Table B2 summarizes all 16 groups of men and women, including information on the numbers of observations in each origin/destination group, the fractions of each origin group that move to each of the four possible destination groups, and the average wage change experienced by each group from two years before to two years after the move. The table also reports a regression-adjusted wage change for the job changers, using the coefficients from a model of wage changes fit to the sample of job *stayers* who remain on the same job over a given four-year interval. The average adjusted wage changes for job changers who stay in the same co-worker wage quartile are all relatively small – e.g., 0.1% for male movers from quartile 1 jobs to other quartile 1 jobs, and -1.7% for female movers from quartile 2 jobs to other quartile 2 jobs – suggesting that mobility *per se* has little effect on wage growth. The only exception is for movers among firms in quartile 4, who experience relatively modest wage gains (5% for men and 6% for women) relative to stayers.

Mobility between quartiles, on the other hand, experience relatively large wage gains or losses, even controlling for experience. Moreover, while not precisely symmetric, the mean wage changes for people who move in opposite directions between quartile groups (e.g, from quartile 1 to quartile 2, versus from quartile 2 to quartile 1) are of similar magnitude and uniformly of opposite sign. This is illustrated in Appendix Figures B3 and B4, where we graph the mean adjusted wage changes for downward movers (e.g., from quartile 4 to quartile 3 firms) against the adjusted wage changes for symmetric upward movers (e.g. from quartile 3 to quartile 4). The wage changes of matched upward- and downward movers lie very close to a line with slope -1, consistent with the symmetry implications

[21]The QP does not collect information that allows us to distinguish the reasons for job changes, though we suspect that many transitions to higher-quartile firms are voluntary moves, while many of the transitions to lower-quartile firms arise from layoff and firing events.

of an AKM model with exogenous mobility, though for both men and women we can formally reject the hypothesis of symmetry.[22]

Comparisons between Figures 1 and 2 point to another important fact, which is that the wage changes for female movers in a given origin-destination group tend to be smaller in absolute value than the corresponding changes for men. This is illustrated in Figure 3, where we plot the adjusted wage changes for each of the 16 origin-destination quartiles for women against the corresponding adjusted changes for men. The points lie very tightly clustered around a line with a slope significantly less that 1, confirming that women gain less from moving to jobs with more highly paid co-workers, and lose less from moving in the opposite direction. Equation (4) implies that the expected wage change for men who move from firm j to firm k is $\psi_k^M - \psi_j^M$, while the expected change for women making the same transition is $\psi_k^F - \psi_j^F = (\gamma^F/\gamma^M)(\psi_k^M - \psi_j^M)$. The slope of the line in Figure 3 (0.77) can therefore be interpreted as an estimate of the relative bargaining power ratio.[23]

To summarize, our descriptive analysis confirms that firm-specific wage premiums are an important feature of the wage structure, and that firm mobility is (at least over the horizon we study) related to time invariant person components of wages but not to time varying or match components of wages. Furthermore, moves between matched groups of firms affect the wages of men proportionally more than the wages of women – a pattern we interpret as strong qualitative evidence that men have greater average bargaining power than women.

5 Estimation of Worker-Firm Models

5.1 Estimation Sample

We turn now to a more systematic analysis of the firm-specific pay premiums for men and women. Building on equation (4), we fit models that include person effects, gender-specific firm effects, and a set of time-varying observable covariates with gender-specific coefficients. For simplicity, we restrict our analysis to the largest connected set of firms for each gender. The estimation samples are described in columns 3 and 4 of Table 1. Overall, 91% of all person-year observations for male workers and 88% of all person-year observations for female workers are included in the largest connected sets. The included workers are very similar to those in our overall analysis sample, and in particular have only slightly higher average wages. After estimating the AKM models separately using these samples, we then narrow our focus to workers who are employed at firms that are in the connected sets for both men and women. This dual-connected sample of men and women – described in columns 5 and 6 of Table 1 – includes just over two thirds of the person-year observations from columns 1 and 2.

[22]The null hypothesis of symmetry is equivalent to the restriction that the sum of each upward and downward change in a quartile-to-quartile pair is zero. To account for the first stage regression adjustment of wage changes, we used a block bootstrap procedure to compute the standard error of the sum of each transition pair allowing for two-way clustering on worker and firm. This was accomplish by running three bootstraps: one resampling workers, one firms, and one worker-firm matches. The three asymptotic variances were then combined according to equation 2.11 of Cameron, Gelbach, and Miller (2011). We then used the estimated covariance matrix of the quartile-to-quartile sums to compute a Wald test of the hypothesis that the six sums were jointly zero. The test statistics were 17.6 for men and 87.9 for women, both of which possess an asymptotic $\chi^2(6)$ distribution (1% critical value is 16.8).

[23]This estimate should be regarded as suggestive rather than definitive, since women and men are not equally distributed across the firms in each quartile group. We present estimates based on firm-specific comparisons below.

Individuals in the dual-connected set have higher education than in the workforce as a whole, and also have somewhat higher average wages. The gender wage gap is also larger in this sample than in our overall sample (23% versus 18%), reflecting the omission of the single-gender firms, which as noted earlier have a relatively small gender gap.

5.2 Estimation Results

Columns 1 and 2 of Table 2 summarize the parameter estimates and fit of our models for men and women in the largest connected sets of workers of each gender.[24] The models include fixed effects for workers and firms as well as year dummies, fully interacted with 4 education dummies (for 6, 9, 12 and 16 years of education) and quadratic and cubic terms in age interacted with education dummies.

We show the standard deviations of the estimated person and firm effects and the covariate indexes $\left(X_{it}'\widehat{\beta}^g\right)$ for each observation, as well as the correlation of the person and firm effects, the residual standard deviation of the model, and the adjusted R^2 statistics. For both males and females, the standard deviations of the person effects are nearly twice as big as the standard deviations of the firm effects, implying that a relatively large share of wage inequality for both genders is attributed to worker characteristics that are equally rewarded at all firms. The correlations between the estimated person and firm effects are both positive, implying that more highly-skilled men and women are disproportionately employed at firms that pay higher wages to all their workers. Such positive assortative matching has been found in many recent studies of wage determination.[25]

The middle panel of Table 2 shows fit statistics for a generalized model that includes dummies for each worker-firm match. This model, which relaxes the additive structure of equation (4), provides only a slight improvement in fit, with about a 1 percentage point rise in the adjusted R^2 statistics. Comparing the residual standard error of the generalized model to the corresponding standard error for the AKM model we can construct an estimate of the standard deviation of the permanent job match effects (the $m_{iJ(i,t)}$) that are absorbed in the job match model but included in the residual of an AKM model. The estimates are 0.062 for men and 0.054 for women – only about one-quarter as big as the standard deviations of the firm effects for the two genders. The component of the wage premium that is shared by all workers at a given firm is considerably more important than the idiosyncratic component.

We have also conducted a series of additional specification checks of the fit of our basic models. In one check, we examine the mean residuals from equation (4) for subgroups of observations classified by the decile of the estimated person effect and the decile of the estimated firm effect. As shown in Appendix Figures B5 and B6, we find that the mean residuals are very small in all 100 cells for both genders, supporting our conclusion that the additive structure of (4) provides a good approximation to the wage-setting process. In a second check, we examined the mean residuals for workers who

[24] Estimates were computed using a preconditioned conjugate gradient algorithm as in Card, Heining and Kline (2013).

[25] See e.g., Card, Heining and Kline (2013) for West Germany, Mare and Hyslop (2006) for New Zealand, Skans, Edin and Holmlund (2008) for Sweden, and Bagger, Sorensen and Vejelin (2012) for Denmark. The sampling errors in the estimated person and firm effects from a model such as (4) are in general negatively correlated (see e.g., Mare and Hyslop, 2006; Andrews, Shank and Upward 2008), implying that the correlations between the estimated effects are downward biased estimates of the degree of assortative matching.

transition between groups of firms, classified by the quartile of the (gender-specific) estimated firm effects. We find that the mean residuals are small in magnitude for all groups of movers.

The bottom rows of Table 2 present the main components of a simple decomposition of the variance of wages across workers implied by the fitted version of equation (4):

$$\begin{aligned} Var\left(w_{it}\right) &= Var\left(\hat{\alpha}_i\right) + Var\left(\hat{\psi}^{G(i)}_{J(i,t)}\right) + 2Cov\left(\hat{\alpha}_i, \hat{\psi}^{G(i)}_{J(i,t)}\right) \\ &\quad + Var\left(X'_{it}\hat{\beta}^{G(i)}\right) + 2Cov\left(\hat{\alpha}_i + \hat{\psi}^{G(i)}_{J(i,t)}, X'_{it}\hat{\beta}^{G(i)}\right) + Var\left(\hat{r}_{it}\right). \end{aligned} \tag{16}$$

Among both male and female workers, person effects accounts for about 60% of overall wage variation, firm effects account for about 20%, and the covariation in worker and firm effects accounts for an additional 10%. The contribution of the measured covariates (including the main effects and the covariances with the person and firm effects) is relatively small, and the residual component is also small (<5%), reflecting the high R^2 coefficients for the underlying models.

5.3 Normalizing the Estimated Firm Effects

The next step in our analysis is to renormalize the estimated firm effects from the models in Table 2. Following the approach outlined in section 3.2, we identify a threshold level for our measure of the size of the surplus available at a firm – value added per worker – such that firms below that threshold are "zero surplus" firms. Figure 4 shows the relationship between average log value-added per worker and the estimated firm effects for men and women (which were normalized for purposes of estimation by setting the effects to zero for the largest firm in the sample). We group firms into percentile bins of value-added and plot the mean estimated firm effects for men and women at the firms in each bin against average log value-added per worker for firms in the bin.

A striking feature of this figure is the piecewise linear nature of the relationship with value-added. Firms in the bottom percentile groups pay very similar average wages, while at higher percentiles the firm-specific wage premiums for both men and women are linearly increasing in log value-added per worker, suggesting a constant elasticity relationship between wages and value-added above a kink point. To identify the value-added threshold more formally, we fit a series of bivariate regression models of the form:

$$\begin{aligned} \hat{\psi}^M_{J(i,t)} &= \pi^M_0 + \pi^M \max\left\{0, \overline{S}^o_{J(i,t)} - \tau\right\} + \nu^M_{J(i,t)} \\ \hat{\psi}^F_{J(i,t)} &= \pi^F_0 + \pi^F \max\left\{0, \overline{S}^o_{J(i,t)} - \tau\right\} + \nu^F_{J(i,t)}. \end{aligned} \tag{17}$$

where (as above) $\overline{S}^o_j$ is the average of log value-added per worker at firm j and τ is a threshold beyond which the firm begins to share rents. We estimated these equations using firm-level data for all firms in the dual connected sample that can be matched to the financial data set.[26] We then selected the value of τ that minimized the mean squared error of the system of two equations. This procedure

[26]We fit the model to firm-level data using the 47,477 dual connected firms with matched financial data, weighting each firm by the total number of person-years of employment at the firm in our data set. These firms account for 63% of the person-year observations at dual-connected firms.

selects a value of $\hat{\tau} = 2.45$, which visually matches the pattern in the Figure. The estimated values of the coefficients π^M and π^F are 0.156 and 0.137, respectively.[27] We show the fitted relationships in Figure 4.

The implied set of "no surplus" firms (i.e., those with $\overline{S}_j^o < \hat{\tau}$) account for 9% of all person-years at dual-connected firms with financial information. As documented in Appendix Table B3, these firms are relatively small, have relatively low sales per worker, tend to employ more women than women, and are disproportionately concentrated in the hotel and restaurant sector. Given the estimated $\hat{\tau}$ we then re-normalized the estimated firm effects for both genders to have employment-weighted averages of zero across all firms with $\overline{S}_j^o < \hat{\tau}$.[28] To check the sensitivity of our normalization procedures, we used a nonparametric bootstrap procedure to estimate the sampling error of $\hat{\tau}$, which yielded an estimated standard error of 0.09. We then re-calculated the constants used to renormalize the firm effects to have mean zero in the relevant zero surplus sample using the upper and lower bounds of the 95% confidence interval for τ. We obtained normalizing constants that are quite close to the baseline constants for $\hat{\tau} = 2.45$, suggesting that our procedure is relatively insensitive to uncertainty about the location of $\hat{\tau}$. As described below, we also confirm this insensitivity using a normalization that assumes the mean wage premiums paid by firms in the hotel and restaurant sector are zero, and by replicating the procedure from (17) using sales per worker instead of value added per worker as the indicator of surplus.

Figure 5 graphs the normalized firm effects for women against the corresponding effects for men, using the same 100 groups as in Figure 4. There is clearly a strong relationship between the average premiums paid to male workers in each group and the average premiums paid to female workers. The employment-weighted correlation of $\hat{\psi}_j^F$ and $\hat{\psi}_j^M$ is 0.59, and the corresponding regression of $\hat{\psi}_j^F$ on $\hat{\psi}_j^M$ has a slope of 0.56. Given the presence of sampling errors in the estimated firm effects, however, this is a downward-biased estimate of the rent-sharing ratio γ^F/γ^M. Grouping firms into cells based on their average value added per worker averages out the sampling errors and yields a relatively precisely estimated slope coefficient of 0.89.

6 Firm-specific Pay Premiums and the Gender Wage Gap

6.1 Basic Decompositions

Next, we use the normalized firm effects for men and women to quantify the impact of firm-specific pay premiums on the gender wage gap, using the framework of equations (8) and (9). The top row of Table 3 shows the terms involved in these alternative decompositions for all workers in the dual connected sample. As shown in column (1), the gender wage gap for this sample is 0.234. Columns 2 and 3 show the mean values of the estimated male wage premiums among men and women, respectively. These can be interpreted as estimates of the average rents received by men and women relative to jobs at

[27] Appendix Figure B7 shows the adjusted R^2 from the bivariate system for a range of values of τ and the associated estimates of the coefficients (π^M, π^F).

[28] This is essentially the same as subtracting the estimated values of the constants ($\hat{\pi}_0^M$ and $\hat{\pi}_0^F$) in equation (17) from $\hat{\psi}_j^M$ and $\hat{\psi}_j^F$, respectively.

no-surplus firms. The difference in column 4 (0.049) is the overall contribution of firm-specific pay premiums to the gender wage gap, and accounts for 21% of the overall gender wage gap.

The part of this total that is attributable to the sorting channel can be calculated by evaluating the difference in the average of the male wage premiums weighted by the shares of men versus women at each firm, or by calculating a corresponding difference in the average of the female wage premiums. The first of these two estimates is shown in column 5, and amounts to 0.035 (or 15% of the overall gender wage gap), while the second is shown in column 6, and amounts to 0.047 (or 20% of the gap). Likewise, the contribution of the bargaining channel can be calculated either by taking the average difference in the estimated male and female wage premiums, weighted by the fraction of men at each firm (column 7), or by by taking the average difference in the two premiums, weighted by the fraction of women at each firm (column 8). The first method yields a very small estimate of the bargaining effect (0.003 or 1.2% of the wage gap) while the second yields a somewhat larger estimate (0.015 or 6.3% of the gender gap).[29]

To interpret the magnitude of the bargaining effect, note that our estimate of the average rents received by male workers in Portugal is modest (14.8%). If women and men had the same distribution across firms, but women earned only 90% of the wage premiums received by men (i.e., $\gamma^F/\gamma^M = 0.9$), then we would obtain an estimate of the bargaining effect equal to 1.5%. This is about equal to the estimate in column 8 based on the female distribution of workers across firms. The estimate of the bargaining effect based on the male distribution is smaller, implying that men are relatively concentrated at firms where the gap $\hat{\psi}_j^M - \hat{\psi}_j^F$ is small.

The lower rows of Table 3 present a parallel set of decompositions for different age and education subgroups. Comparing across age groups (rows 2-4) the entries in column 1 show that the male-female wage gap in Portugal widens dramatically with age. Firm-specific pay differentials contribute to this pattern, with most of the increase attributable to a rise in the sorting effect with age. A more detailed perspective is provided in Figure 6, which shows the overall gender gap (plotted with triangles) and the components of our decomposition for narrow age bins. Our estimate of the average rents received by men (plotted with squares) shows that these rise with age until the mid-thirties, and then are relatively stable until the mid-fifties, when they begin to fall off. The age profile of average rents for women (plotted with circles) is flatter and peaks earlier. Thus our estimate of the total contribution of firm wage premiums to the gender wage gap (plotted with diamonds) rises until the mid-fifties, peaking at around 7.5 percentage points. As shown by the dotted lines at the bottom of the figure, the sorting component explains between 75% and 95% of the overall contribution.

Comparisons across education groups in the bottom rows of Table 3 show that the gender wage gap is roughly constant across education groups, but the average pay premiums received by both men and women are increasing with years of schooling, confirming that there is positive assortative matching between higher-skilled workers and higher-paying firms. As shown in column 4, the net effect of firm-specific pay premiums on the gender wage gap is about the same for workers with less

[29] In the wage decomposition literature (e.g., Jann, 2008) the sorting effect is often called an "endowment" effect, since it evaluates the differences in the shares of men and women at different firms, using the "returns" to each firm calculated for either men or women. The bargaining effect is often called a "coefficient" effect, since it evaluates the differences in the estimated "returns" to working at a given firm for men versus women using the "endowments" of men or women.

than high school or high school education, but is somewhat smaller for university-educated workers, reflecting a much smaller sorting effect for these workers, coupled with a larger bargaining effect.

6.2 Decompositions by Occupation and Industry

Men and women tend to work in different occupations (see e.g. Manning and Swaffield, 2008, and Goldin, 2014 for recent analyses, and Cardoso, Guimarães and Portugal, 2012 for a discussion in the Portuguese context). This raises the question of whether some of the differences identified in Table 3 are actually due to occupation rather than gender. We investigate this issue in Table 4, assigning each worker to his or her modal occupation. Notice first that the gender wage gap varies widely across occupations, from around 15% for professionals, technicians, clerks, and service workers to 40% for craft occupations. The average size of the firm-specific wage premiums received by male and female workers also varies substantially, with a net contribution to the gender wage gap that ranges from 1-2% for managers and service workers to 6% or more for technicians, clerks, and craft workers. For most occupations the sorting effect is larger than the bargaining effect, though for managers and professionals– the two groups with the highest fraction of university-educated workers – the bargaining effect is relatively large, consistent with the patterns in Table 3.

We have also examined the contributions of the sorting and bargaining channels to the gender wage gap for workers in different major industries. The results, summarized in Appendix Table B4, show that in most industries women tend to be over-represented at firms that pay both men and women relatively lower wages, with particularly large sorting effects in the chemical, non-metallic mineral, business services, and utility industries. Likewise, the firm-specific wage premiums offered to women are smaller than the premiums for men in most sectors, with relatively large bargaining effects in the food products, paper and publishing, and chemical industries. An interesting exception to these patterns is construction, which has the lowest fraction of female workers among the major industries (10%). Females in construction are much better-educated than males, earn higher average wages, and tend to be sorted to firms that pay higher wages to both men and women (i.e., a sorting effect of the "wrong sign").

6.3 An Alternative Normalization

Our estimates of the relative bargaining effect rely on a normalization that allows us to estimate the average rents earned by men and women. To check the robustness of our findings we considered an alternative normalization based on the assumption that firms in the hotel and restaurant industry pay zero rents to workers on average. Firms in this sector pay the lowest average wage premiums of all major industries (see Appendix Table B4). Job turnover rates are also high, suggesting that workers are able to find a job in the industry relatively easily. We therefore normalized the estimated wage premiums for men and women such that the weighted average of both premiums is 0 in the sector (weighting by the total number of workers at each firm). Appendix Table B5 reproduces the decompositions in Table 3 using this alternative normalization assumption. The estimated sorting effects are invariant to normalization and are therefore the same as in Table 3. The estimated bargaining effects, however, are uniformly larger, reflecting the fact that the mean of the estimated male wage premiums for male workers falls from 0.148 under the baseline normalization to 0.146 under the alternative, while the estimated female premium for female workers falls from 0.099 to 0.076. As a result the estimated bargaining power effects are all increased by 0.021 (the difference in these changes) relative to those presented in Table 3.[30] Under this alternative normalization, firm-specific wage premiums explain about 30% of the gender wage gap for workers as a whole, with 15-20% explained by sorting effects and 10-15% explained by relative bargaining effects. This alternative normalization therefore suggests that our baseline procedure leads to a conservative estimate of the bargaining power effect.

6.4 Compensating Differentials for Hours?

A recent literature (e.g., Bertrand et al, 2010; Goldin, 2014) suggests that part of the gender wage gap is due to compensating differentials for long hours of work. If some firms offer packages of high wages and long hours that are relatively unattractive to female workers, we would attribute the resulting pattern of wage and employment outcomes to the sorting channel. To evaluate the potential role of hours differences we calculated average total monthly hours of work for male and female employees at each firm (including regular hours and overtime), and regressed the estimated wage premiums for each gender on the corresponding measures of hours. We fit OLS models with and without controls for major industry, and IV models in which we used hours of the other gender group as an instrument for each group's hours to address the "division bias" problem (Borjas, 1980) that arises because wages are constructed by dividing monthly earnings by average monthly hours. The results, summarized in Appendix Table B6, show no evidence that hours are significant determinants of the wage premiums offered by different firms to either men or women. We therefore conclude that (at least in Portugal) compensating differentials for long hours are not an important driver of the gender wage gap.

6.5 Gender or Occupation?

So far we have assumed that the surplus at different firms is shared differentially by gender, but ignored other possible dimensions of heterogeneity in rent sharing. A potential concern is that some of the differential bargaining power we measure for women is actually due to differences in the extent to which workers in different occupations receive different shares of firm-wide rents.[31] To address this issue we fit separate AKM models for male and female workers who work in mainly male or mainly female occupations, allowing unrestricted firm effects for gender and occupation group. We then investigated whether there is a systematic difference in bargaining power between men and women who work in a given occupation group.

We began by calculating the fraction of female workers in each of the 110 3-digit occupations identified in the QP data set (based on person-year level data). Next, we assigned each individual the average fraction of female workers in his or her occupation in each year. We then classified workers as

[30] Inspection of expressions (8) and (9) shows that if the estimated female premiums are all adjusted downward relative to the male premiums, then the estimated bargaining effect for any subgroup of workers is raised by the difference in the adjustment factors.

[31] Ransom and Oaxaca (2005) present a case study of occupational segregation within a chain of U.S. grocery stores that led to substantially lower wages for female employees at the firm.

having a "mainly female" or "mainly male" occupation depending on whether the average fraction of female workers in his or her occupation(s) is above or below the median across all occupations. We classify 83% of women and 27% of men as having mainly female occupations, and a complementary 17% of women and 73% of men as having mainly male occupations. On average men in mainly male and mainly female occupations earn about the same wages. In contrast, women in mainly male occupations earn substantially more (+20%) than those in mainly female occupations, in part because those in mainly male occupations are 10 percentage points more likely to hold a university degree. Estimated AKM models for men and women in the two occupation groups are similar to the models presented in Table 2 and yield similar conclusions about the relative importance of worker and firm effects in the overall variation of wages.

Table 5 summarizes the results of our comparisons between men and women in mainly female and mainly male occupations. For reference, panel A of the Table reproduces the results from Table 3 for all men and women at dual connected firms. Panel B presents a comparison between men and women with mainly female occupations (confining attention to those who work at firms that employ both groups). The total contribution of firm-specific pay premiums to the gender gap (column 4) is about two-thirds as large as in the overall sample, reflecting a small reduction in the sorting component and a larger reduction in the relative bargaining power component, which is estimated to be close to zero. In contrast, for workers in mainly male occupations (panel C) the bargaining effect is relatively large, explaining 12-21% of the overall gender wage gap. Overall, this analysis suggests that the sorting effect is important for workers in mainly female and mainly male occupations, while the bargaining effect is concentrated among workers in traditionally male occupations, which tend to employ more highly educated women.

6.6 Confounding Effects of the Minimum Wage?

During our sample period the minimum wage in Portugal was typically set at 50% of the median wage of full-time workers (OECD, 2015b). This is higher than in the U.S. or the U.K. and potentially high enough to compress the gender wage gap (see DiNardo, Fortin, and Lemieux, 1996). Indeed, among workers in our dual-connected sample, 7% of men and 18% of women have a wage in at least one year that is within 5% of the minimum wage. Upward pressure from the minimum wage might constrain some firms that would otherwise offer lower wages to female workers, pushing up ψ_j^F relative to ψ_j^M and leading us to conclude that the relative bargaining effect is small. To examine this possibility, we re-estimated our AKM models using workers over age 24 with a high school education or more. Among this subgroup (who represent 24% of all men and 31% of all women in our analysis sample), only 2% of the men and 3% of the women ever have a wage within 5% of the minimum wage. We then repeated the normalization exercise described in Section 5.3 and formed a new set of decompositions. The results are summarized in Appendix C.

Eliminating younger and less educated workers leads to a reduction in the fraction of the remaining workers in the connected sets for each gender group, since many of the links between firms are formed through mobility of these groups. The selective loss of firms that mainly hire less-educated workers leads to an attenuation of impact of between-firm sorting on the gender wage gap, and a reduction in the fraction of the overall gender gap explained by firm-specific wage setting (see Appendix Table C-2). The estimated impact of the bargaining channel, however, is larger in this subsample than in the workforce as a whole, and comparable to the estimate we obtain for university educated workers in our main analysis – around 2 percentage points. Interestingly,the estimated bargaining power effect for university-educated workers is similar whether we use firm effects estimated for all workers (as in our main analysis) or the firm effects estimated only for higher-educated workers. This suggests that the firm effects for women in our main analysis are not significantly attenuated by inclusion of younger and less educated workers whose wages are constrained by the national minimum wage.[32]

7 Firm Wage Premiums and Measured Surplus

7.1 Estimates of Rent Sharing Models

In this section we turn to the question of how the estimated wage premiums for different firms are related to measures of the potential surplus at the firm. As a starting point, we note that the simple correlation coefficients between mean log value added per worker and the firm effects are 0.38 for men and 0.34 for women, while the corresponding correlations with mean log sales per worker are 0.34 and 0.32. Given the presence of sampling errors in the estimated wage premiums, and potential noise in measures of value added and sales (particularly for the small firms that comprise the bulk of our sample), we interpret these correlations are relatively strong.

Table 6 presents estimates of the rent-sharing coefficients π^M and π^F based on equation (11), using three alternative measures of surplus. The models in row 1 use the net surplus measure from our baseline normalization procedure, which is based on mean log value added per worker. The estimation sample includes all firms in the dual connected set that can be linked to SABI and have at least one year of value added data. By construction, the estimates are identical to the estimates obtained from equation (17) at the optimized value for $\hat{\tau}$, and yield elasticities of 0.16 for men and 0.14 for women.[33] To estimate the sampling error for the ratio, note that $\hat{\pi}^F/\hat{\pi}^M$ is the two-stage least squares estimate of the parameter δ_1 from a simple model of the form:

$$\hat{\psi}^F_{J(i,t)} = \delta_0 + \delta_1 \hat{\psi}^M_{J(i,t)} + e_{J(i,t)},$$

using net surplus as an instrumental variable for $\hat{\psi}^M_j$. We therefore use the conventional standard error

[32] To test the attenuation hypothesis directly we regressed the estimated firm effects for women from our main analysis on the estimated effects obtained using only older and more educated workers, instrumenting the right hand side variable with the estimated firm effect for men from our main analysis. The model was estimated at the firm level, weighting by the number of person-year observations for females at the firm. The estimated coefficient is 1.02 (standard error 0.01), suggesting that there is minimal attenuation of the firm effects from our main analysis.

[33] These coefficients are slightly bigger than the IV estimates of the elasticity of wages with respect to value added per worker reported by Card, Devicienti and Maida (2014, hereafter CDM) based on matched worker-firm data for Italy (e.g., 0.09 in Table A4 of CDM). Using the average ratio of quasi-rents to value added per worker estimated by CDM, this implies an elasticity of wages with respect to quasi rents per worker of about 0.07 comparable to or a little bigger than the estimates reported by Arai and Heyman (2009) for Swedish workers, Martins (2009) for Portuguese workers, Guertzen (2009) for German workers, and Guiso et al. (2005) for Italian workers. These elasticities are only about one-quarter as large as elasticities estimated by Abowd and Lemieux (1993) and Van Reenen (1996) using firm-level data without controls for worker quality. See CDM for a more detailed summary of the recent rent sharing literature.

of the two-stage least squares estimator as our estimated standard error for the ratio. The estimated ratio is 0.88 with a standard error of 0.03 (column 4). We can therefore rule out the null hypothesis of equal rent-sharing ($\pi^F = \pi^M$) in favor of the alternative that women receive a smaller share of the component of firm-wide rents that is directly related to excess value-added.

In row 2 we check the robustness of this conclusion using mean log sales per worker as an alternative proxy for the potential surplus at each firm. Since sales (for the previous calendar year) are reported in QP we are able to expand the sample to include all firms in the dual connected set with reported sales in at least one year. Sales are also measured independently of labor costs at the firm, so a finding that our conclusions are robust to using sales per worker provides a check that there is not a measurement-related problem in using value added per worker.

As expected, given that sales per worker are significantly greater than value added per worker, the estimated rent sharing elasticities in row 2 are smaller in magnitude the the elasticities in row 1.[34] Nevertheless their ratio is virtually the same, and is again significantly different from 1. The models in row 3 use a third indicator which is derived from an alternative normalization procedure in which we define net surplus using the excess of sales per worker over a minimum threshold level (see the next section). This choice leads to a slight increase in the rent sharing coefficients relative to the specifications in row 2, but again their ratio is nearly invariant.

To probe the robustness of the results in Table 6 we re-estimated the models including controls for industry (20 dummies), location (dummies for firms located in Lisbon or Oporto) and a quadratic in firm size (based on average total employment in all years). Estimates from these models are presented in Appendix Table B7. In brief, the addition of controls leads to a slight attenuation (on the order of 10-15%) in the estimated rent-sharing coefficients, with a slightly bigger attenuation of the coefficients for women than men. These models therefore reinforce our conclusion that women get a smaller share of rents than men, though in all cases the estimated ratios are within a standard error of the ratios in Table 6.

In summary, we find that the estimated firm-specific wage premiums for men and women are highly correlated with measures of the surplus per worker at the firm. Importantly, the estimated correlations are uniformly smaller for women than men, providing strong support for the view that women get a smaller share of the surplus than men.

7.2 Decompositions of the Gender Wage Gap using Direct Measures of Surplus

The component of the firm-specific wage premium received by male workers that is directly attributed to observable surplus is $\hat{\pi}^M E[NS_{J(i,t)}|male]$, while the corresponding component for female workers is $\hat{\pi}^F E[NS_{J(i,t)}|female]$. Their difference gives the contribution of the observable component of surplus to the gender wage gap, and can be decomposed into bargaining and sorting channels using equations (12) and (13). In Appendix Table B8 we present estimates of the terms in these two alternative decompositions, using the overall sample of dual-connected workers. We find that the component of

[34] The elasticity of wages with respect to sales per worker will be attenuated relative to the elasticity with respect to value added per worker by a fraction that represents the ratio of value added to sales. The the average value of this ratio in our data is 0.43 (weighting each firm by the number of person years of employment at the firm).

the male and female firm effects that is explained by our observed measure of surplus accounts for about 80% of our overall estimates of the size of the rent premiums received by both men and women, and 80% of the impact of firm-specific premiums on the gender wage gap. Applying the decompositions in equations (12) and (13) we find that differential sorting of men to high surplus firms accounts for about two-thirds of the total effect of our surplus indicator, while the lower bargaining power of women accounts for about one-third, or 1-1.5 log points of the overall gender gap. This evidence on the differential sharing of observed surpluses reinforces our conclusion that the lower relative bargaining power of women contributes to the gender wage gap, particularly for subgroups of workers who are most likely to work at high-surplus firms (e.g., higher-educated workers).

7.3 Sales per Worker as an Alternative Measure of Surplus

We have relied on measures of value added per worker in two ways: in choosing the normalization for the estimated firm-specific wage premiums, and in directly measuring the relative bargaining power of male and female workers. To check the robustness of our procedures, we followed the same steps in developing a normalization and estimating rent sharing equations for male and female workers, using sales per worker as an alternative indicator of surplus. As discussed above, an advantage of sales is that it is reported in QP for most firms (89% of the 84,720 firms in our dual-connected sample).

Appendix D summarizes the results from this alternative approach. Appendix Figure D1 graphs the (unnormalized) firm effects for men and women against mean log sales per worker. As we found using mean log value added per worker as a measure of the surplus, there is a clear visual kink in the relationship between the estimated firm effects and mean log sales per worker. Using the same procedure as we describe Section 5.3, we identified the kink point τ_s. We then define excess mean log sales per worker as $\min[0, \overline{S}_j - \overline{L}_j - \hat{\tau}_s]$ where $\left(\overline{S}_j - \overline{L}_j\right)$ represents mean log sales per worker, calculated using annual sales data in QP for all years it is available for a given firm. Appendix Table D1 presents as series of decompositions parallel to those in Table 3 but using the normalization that the mean firm premiums for workers at firms with $\overline{S}_j - \overline{L}_j \leq \hat{\tau}_s$ are equated to 0. Reassuringly, this alternative normalization yields estimates of the bargaining effect that are essentially identical to the estimates from our baseline procedure. Finally, as reported in row 3 of Table 6, use of excess mean log sales per worker as a measure of surplus yields estimates of the rent sharing coefficients for male and female workers that are about 60% as large as the estimates based on excess mean log value added per worker, consistent with the fact that value added per worker is on average just over one half of sales per worker. Overall, we conclude that our results are highly robust to the use of either value added per worker or sales per worker as an indicator of the surplus available at each firm.

8 Within Firm Changes in Profitability and Wages

As a final step in our analysis, we use observations from the last four years of our analysis sample to measure the effects of *changes* in the measured surplus at each firm on the wages of male and female job stayers. Our base sample includes information on some 280,000 men and 200,000 women who were employed continuously at a firm that can be linked to value added information for 2006 and 2009 in

SABI. These workers have similar age, education and wages as men and women in our overall analysis sample (see Appendix Table B9). Moreover, the wage gap between male and female stayers is 22 log points in both 2006 and 2009 - about the same as in our dual connected set.

Table 7 presents a series of models based on equation (15) that show the relationship between the change in excess log value-added at a firm and the wage changes of male and female stayers. We estimate these models in two steps, first regressing wage changes on a quadratic in age (separately by gender) and firm indicators, then in a second stage regressing the average regression-adjusted wage changes at each firm on the change in excess log value added at the firm, weighted by the number of workers at the firm in the gender group. Given the large variability in measured value-added, our main specifications Winsorize the change in excess value-added at +/- 0.50. As shown in row 1 of the table. the resulting estimates of the rent sharing coefficients for male and female stayers are 0.049 and 0.045, respectively. Their ratio, shown in column 4 of the table, is 0.91 (with a standard error of 0.09). The point estimate of the relative bargaining power of women is therefore quite consistent with estimates based on between firm comparisons (in Table 6), though we cannot rule out a value of 1 for the ratio. Row 2 shows the same specification, estimated without Winsorizing the change in excess value added. Although the rent sharing coefficients become about 30% smaller – as would be expected if some of the very large changes are due to measurement error – their ratio remains very close to 0.9.

Row 3 uses a larger sample of stayers, observed over the period from 2005 to 2008, for which we have data on sales from the QP. (The earlier sample period reflects the fact that the QP data report sales from the previous calendar year, so the latest sales data is for 2008). Using excess log sales per worker (as defined in developed in Section 7.3) as our measure of surplus, we again estimate significant rent sharing coefficients for men and women, with a ratio of 0.88, though the standard error is relatively large.

Compared to the rent-sharing coefficients from the between-firm analysis Table 6, the estimates for stayers are only about 30% as large. There are several plausible explanations for the discrepancy. First, we suspect that our measures of surplus are relatively noisy, and that taking the difference over a 4 year period leads to a decrease in the signal-to-noise ratio relative to the average over the same period. Second, contrary to equation (1), it may be that wages are less responsive to transitory fluctuations in rents than to permanent differences. Guiso et al. (2005), for example, analyze the relationship between wages and firm value added using Social Security earnings record for Italian workers, and find smaller impacts of short run changes in value added than for long-run changes. A third possibility is that the rent sharing coefficients are attenuated because we are focusing on a selected sample of job stayers. To evaluate this possibility we constructed a simple grouped data control function (Gronau, 1974) for selection bias for the male and female stayers at each firm, based on the fractions of workers employed at the firm in 2006 who stayed to 2009, and re-estimated the models including this control.[35] The results, presented in Appendix Table B10, give no indication of selectivity bias.

[35]The control function is $\phi(\Phi^{-1}(p))/p$, where p is the fraction of stayers among the (gender-specific) set who were at the firm in the base period, ϕ is the normal density function, and Φ is the normal cumulative density function. This is an appropriate control function if the individual probability of staying is determined by a latent index with a normally distributed error, and individual wage changes have a normally distributed error.

Overall, the estimates in Table 7, while limited by the relatively short sample period over which we can observe job stayers, are supportive of the hypothesis that female workers gain less than their male co-workers when their employer becomes more profitable. Indeed, our estimate of the ratio γ^F/γ^M is centered around 0.9 for all women and women in mainly female occupations, quite close to the estimated ratio from our previous designs. We have estimated a variety of additional models for other subgroups of male and female stayers, including workers in larger and smaller firms, workers in firms with larger and smaller fractions of female employees, and workers in firms with higher and lower within-firm wage inequality. Unfortunately, as suggested by the standard errors for the estimated ratios in Table 7, our ability to precisely estimate the relative bargaining power of women is limited, and none of the estimates of the relative ratio of female to male bargaining power are significantly different from 0.9 – the average ratio across firms.

9 Conclusions

A growing body of research argues that firm-specific wage premiums are a pervasive and economically important feature of labor market earnings. These premiums will contribute to the gender wage gap if women tend to work at firms that offer smaller premiums, or if female employees tend to earn smaller premiums than their male colleagues when employed at the same firms. Our analysis of Portuguese data finds that female employees receive about 90% of the wage premiums that men earn at equivalent firms. Moreover, women are disproportionately likely to work at low surplus firms paying small premiums to both genders. We conclude that the sorting and bargaining channels together explain about 20% of the gender wage gap in Portugal, with roughly two-thirds of this 20% explained by sorting and one-third by the shortfall in relative bargaining power.

Our approach to combining worker-firm fixed effects models with Oaxaca-style decompositions into sorting and bargaining components is potentially applicable to other classic wage gaps including the black/white wage gap, the immigrant/native wage gap, the experience profile of wages, and wage gaps based on various measures of intelligence – explanations for which have traditionally relied upon a market price perspective. The proliferation of rich employer-employee datasets offers the opportunity to determine the extent to which these heavily studied sources of wage inequality are in fact mediated by heterogeneity across firms.

References

Abowd, John, Robert Creecy, and Francis Kramarz. 2002. "Computing Person and Firm Effects Using Linked Longitudinal Employer-Employee Data." Cornell University Department of Economics Unpublished Working Paper, March 2002.

Abowd, John, Francis Kramarz and David Margolis. 1999. "High Wage Workers and High Wage Firms." *Econometrica* 67(2): 251-333.

Abowd, John and Thomas Lemieux. 1993. "The Effects of Product Market Competition on Collective Bargaining Agreements: The Case of Foreign Competition in Canada." *Quarterly Journal of Economics* 108 (4): 983-1014

Altonji, Joseph and Rebecca Blank. 1999. "Race and Gender in the Labor Market." In Orley Ashenfelter and David Card (editors), *Handbook of Labor Economics* Vol. IIIc. Amsterdam: Elsevier, pp. 3143-3259.

Andrews, M.J., L. Gill, T. Schank and R. Upward. 2008. "High Wage Workers and Low Wage Firms: Negative Assortative Matching or Limited Mobility Bias? *Journal of the Royal Statistical Society* (Series A) 171 (3): 673-697.

Arai, Mahmood, and Fredrik Heyman. 2009. "Microdata Evidence on Rent-Sharing." *Applied Economics* 41 (23): 2965-2976.

Ashenfelter, Orley and Timothy Hannan. 1986. "Sex Discrimination and Product Market Competition: The Case of the Banking Industry." *Quarterly Journal of Economics* 101(1): 149-174.

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

Babcock, Linda, Michele Gelfand, Deborah Small and Heidi Stayn. 2006. "Gender Differences in the Propensity to Initiate Negotiations." In De Cremer, D., M. Zeelenberg, and J.K. Murnighan (editors), *Social Pyschology and Economics*. Mahwah NJ: Erlbaum Press.

Bagger, Jesper, Kenneth Sorensen and Rune Vejelin. 2012. "Wage Sorting Trends." School of Economics and Management University of Aarhus Working Paper 2012-17.

Barth, Erling and Harald Dale-Olsen. 2009. "Monopsonistic Discrimination, Worker Turnover, and the Gender Wage Gap." IZA DP No. 3930. January.

Bayard, Kimberly, Judith Hellerstein, David Neumark, and Kenneth Troske. 2003. "New Evidence on Sex Segregation and Sex Differences in Wages from Mathed Employer-Employee Data." *Journal of Labor Economics* 21 (4): 887-922.

Becker, Gary. 1957. *The Economics of Discrimination*. Chicago: University of Chicago Press. (second edition 1971).

Bertrand, Marianne. 2011. "New Perspectives on Gender." In Orley Ashenfelter and David Card (editors) *Handbook of Labor Economics*, Vol. 4b. Amsterdam: Elsevier. pp. 1543-1590.

Bertrand, Marianne and Sendhil Mullainathan. 2004. "Are Emily and Greg More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination." *American Economic Review* 94(4): 991–1013.

Bertrand, Marianne, Claudia Goldin and Lawrence F. Katz. 2010. "Dynamics of the Gender

Gap for Young Professionals in the Financial and Corporate Sectors." *American Economic Journal: Applied Economics* 2(1): 228-255.

Black, Sandra E. and Philip E. Strahan. 2001. "The Division of Spoils: Rent-Sharing and Discrimination in a Regulated Industry." *American Economic Review* 91(4): 814-831.

Blau, Francine D. 1977. *Equal Pay in the Office*. Lexington MA: Lexington Books, 1977.

Blau, Francine D. and Lawrence M. Kahn. 2000. "Gender Differences in Pay." *Journal of Economic Perspectives* 14(4): 75-99.

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

Borjas, George. J. 1980. "The relationship between wages and weekly hours of work: The role of division bias." *Journal of Human Resources*, 15(3): 409-423.

Bowles, Hannah R., Linda Babcock and Kathleen L. McGinn. 2005. "Constraints and Triggers: Situational Mechanics of Gender in Negotiations." J*ournal of Personality and Social Psychology* 89: 951-965.

Bowles, Hannah R., Linda Babcock and Lei Lai. 2007. "Social Incentives for Sex Differences in the Propensity to Initiate Bargaining: Sometimes It Does Hurt to Ask." *Organizational Behavior and Human Decision Processes* 103: 243-273.

Cameron, A. C., Gelbach, J. B., & Miller, D. L. 2011. "Robust inference with multiway clustering." *Journal of Business & Economic Statistics*, 29(2): 238-249.

Card, David, Francesco Devicienti and Agata Maida. 2014. "Rent-Sharing, Holdup, and Wages: Evidence from Matched Panel Data." *Review of Economic Studies* 81(1): 84-111.

Card, David, Joerg Heining and Patrick Kline. 2013. "Workplace Heterogeneity and the Rise of West German Wage Inequality." *Quarterly Journal of Economics* 128(3): 967-1015.

Cardoso, Ana Rute, Paulo Guimarães and Pedro Portugal. 2012. "Everything You Always Wanted to Know About Sex Discrimination." IZA Working Paper 7109.

Cardoso, Ana Rute and Pedro Portugal. 2005. "Contractual Wages and the Wage Cushion under Different Bargaining Settings." *Journal of Labor Economics* 23(4): 875-902.

Carlsson, Mikael, Julian Messina and Oskar Nordström Skans. 2011. "Wage Adjustment and Productivity Shocks." Sveriges Riksbank Working Paper #253.

Charles, Kerwin K. and John Guryan. 2008. "Prejudice and Wages: An Empirical Assessment of Becker's The Economics of Discrimination." *Journal of Political Economy* 116(5): 773-809.

Charles, Kerwin K. and John Guryan. 2011. "Studying Discrimination: Fundamental Challenges and Recent Progress." *Annual Review of Economics* 3(1), 479-511.

Del Bono, Emilia and Daniela Vuri. 2011. "Job mobility and the Gender Wage Gap in Italy," *Labour Economics* 18(1): 130-142.

Dickens, William and Lawrence F. Katz. 1986. "Inter-Industry Wage Differences and Industry Characteristics." In Kevin Lang and Jonathon Leonard, editors. *Unemployment and the Structure of Labor Markets.* London: Basil Blackwell.

Eeckhout, Jan and Philipp Kirchner. 2011. "Identifying Sorting – In Theory." *Review of Economic Studies* 78 (3): 872-906.

EPI (Economic Policy Institute). 2010. *The State of Working America* 2008-10. Washington DC: EPI.

Fortin, Nicole, Thomas Lemieux, and Sergio Firpo. 2011. “Decomposition Methods in Economics.” In Orley Ashenfelter and David Card (editors) *Handbook of Labor Economics*, volume 4a. Amsterdam: Elsevier. pp. 1-102.

Goldin, Claudia. 2014. “A Grand Gender Convergence: Its Last Chapter.” *American Economic Review* 104(4): 1091-1119.

Gronau, Reuben. 1974. “Wage Comparisons -A Selectivity Bias.” *Journal of Political Economy* 82(6): , (November/December 1974):1119-1143.

Groshen, Erica L. 1991. “The Structure of the Female/Male Wage Differential: Is It Who You Are, What You Do, orWhere You Work?” *Journal of Human Resources* 26(3): 457-472.

Guertzgen, Nicole. 2009. “Rent-Sharing and Collective Bargaining Coverage: Evidence from Linked Employer-Employee Data.” *Scandinavian Journal of Economics* 111 (2), pp. 323-349.

Guiso, Luigi, Luigi Pistaferri and Fabiano Schivardi. 2005. “Insurance within the Firm.” J*ournal of Political Economy* 113(5): 1054-1087.

Hall, Robert E. and Alan B. Krueger. 2012. “Evidence on the Incidence of Wage Posting, Wage Bargaining, and On-the-Job Search.” *American Economic Journal: Macroeconomics* 4(4): 56-67.

Hellerstein, Judith, David Neumark and Melissa McInerney. 2008. “Changes in Workplace Segregation in the United States between 1990 and 2000: Evidence from Matched Employer-Employee Data.” In Stefan Bender, Julia Lane, Kathryn Shaw, Fredrik Andersson, and Till von Wachter, editors, *The Analysis of Firms and Employees: Quantitative and Qualitative Approaches*. Chicago: University of Chicago Press, 2008.

Heckman. James J and Peter Siegelman. 1993. “The Urban Institute Audit Studies: Their Methods and Findings.” In Michael Fix and Raymond Struyk, editors. *Clear and Convincing Evidence: Measurement of Discrimination in America*. Washington DC: Urban Institute Press.

Hospido, Laura. 2009. “Gender Differences in Wage Growth and Job Mobility of Young Workers in Spain.” Investigaciones Economicas 33(1): 5-37.

ILO (International Labor Organization). 2011. Key Indicators of the Labour Market Database, 7th Edition. Available at www.ilo.org/empelm/pubs/WCMS_114060/lang–en/index.htm

INE (Statistics Portugal). 2012. *Estatísticas do Emprego*. 1.º Trimestre. Lisbon: Statistics Portugal.

Jann, Ben. 2008. “The Blinder-Oaxaca Decomposition for Linear Regression Models.” *Stata Journal* 8 (4): 453-479.

Kramarz, Francis and Oskar Nordstrom Skans. 2013. “When Strong Ties are Strong: Networks and Youth Labor Market Entry.” *Review of Economic Studies*. forthcoming.

Krueger, Alan B. and Lawrence H. Summers. 1988. “Efficiency Wages and the Inter-Industry Wage Structure.” *Econometrica* 56 (2): 259-293.

Lang, Kevin and Jee-Yeon K. Lehmann. 2012. “Racial Discrimination in the Labor Market: Theory and Empirics.” *Journal of Economic Literature* 50(4): 959-1006.

Lopes de Melo, Rafael. 2009. “Sorting in the Labor Market: Theory and Measurement.” Unpublished Working Paper (December).

Loprest, Pamela J. 1992. "Gender Differences in Wage Growth and Job Mobility." *American Economic Review* 82(2): 526-532.

Macis, Mario and Fabiano Schivardi. 2013. "Exports and Wages: Rent Sharing, Workforce Composition or Returns to Skills?" Johns Hopkins Department of Economics Unpublished Working Paper, August 2013.

Manning, Alan, 2011. "Imperfect Competition in the Labor Market." In Orley Ashenfelter and David Card (editors) *Handbook of Labor Economics*, volume 4b. Amsterdam: Elsevier. pp. 976-1041.

Manning, Alan and Joanna K. Swaffield. 2008 "The Gender Gap in Early-Career Wage Growth". *Economic Journal* 118(530): 983-1024.

Mare, David and Dean Hyslop. 2006. "Worker-Firm Heterogeneity and Matching: An Analysis Using Worker and Firm Fixed Effects Estimated from LEED." Statistics New Zealand Working Paper. Wellington, NZ: Statistics New Zealand.

Martins, Pedro 2009. "Rent-sharing Before and After the Wage Bill." *Applied Economics* 41 (17): 2133-2151.

Martins, Pedro S. and Pedro T. Pereira. 2004. "Does Education Reduce Wage Inequality? Quantile Regression Evidence from 16 Countries." Labour Economics 11 (2004): 355-371.

Mumford, Karen and Peter N. Smith. 2008. "What Determines the Part-time and Gender Earnings Gaps in Britain: Evidence from the Workplace." *Oxford Economic Papers* 61: i56-i75.

Mulligan, C. B., & Rubinstein, Y. 2008. Selection, investment, and women's relative wages over time. The Quarterly Journal of Economics, 1061-1110.

Neumark, David, Roy J. Bank and Kyle D. Van Nort. 1996. "Sex Discrimination in Restaurant Hiring: An Audit Study." *Quarterly Journal of Economics* 111(3): 915-941.

Nekby, Lena. 2003. "Gender Differences in Rent Sharing and Its Implications for the Gender Wage Gap: Evidence from Sweden." *Economics Letters* 81: 403-410.

Oaxaca, Ronald. 1973. "Male-Female Wage Differentials in Urban Labor Markets." *International Economic Review*, 14(3): 693–709.

Oaxaca, Ronald L. and Michael R. Ransom. 1999. "Identification in Detailed Wage Decompositions." *Review of Economics and Statistics* 81(1): 154-157.

OECD, 2012. "OECD Family Database." Available at www.oecd.org/social/family/database.

OECD, 2015a. "Gender Wage Gap" Available at http://www.oecd.org/gender/data/genderwagegap.htm.

OECD, 2015b. "Minimum Relative to Average Wages of Full-Time Workers." Available at http://stats.oecd.org/Index.aspx?DataSetCode=MIN2AVE

Papps, Kerry L. 2012. "Spillovers and Wage Determination Within Firms." Department of Economics University of Oxford Unpublished Working Paper.

Petersen, Trond and Laurie A. Morgan. 1995. "Separate and Unequal: Occupation-Establishment Sex Segregation and the Gender Wage Gap." *American Journal of Sociology* 101(2): 329-365.

Pissarides, Christopher and Jonathan Wadsworth. 1994. "On-the-job search: some empirical evidence from Britain." *European Economic Review*, 38(2), 385-401.

Ransom, Michael R. and Ronald L. Oaxaca. 2005. "Intrafirm Mobility and Sex Differences in Pay." *Industrial and Labor Relations Review* 58(2): 219-237.

Robinson, Joan. 1933. *The Economics of Imperfect Competition.* London: Macmillan.

Save-Soderbergh, Jenny. 2007. "Are Women Asking for Low Wages? Gender Differences in Wage Bargaining Strategies and Ensuing Bargaining Success." Swedish Institute for Social Research Working Paper 7/2007.

Skans, Oskar Nordstrom, Per-Anders Edin and Bertil Holmlund. 2008. "Wage Dispersion Between and Within Plants: Sweden 1985–2000." In Edward P. Lazear and Kathryn L. Shaw (editors). *The Structure of Wages: An International Comparison.* Chicago: University of Chicago Press.

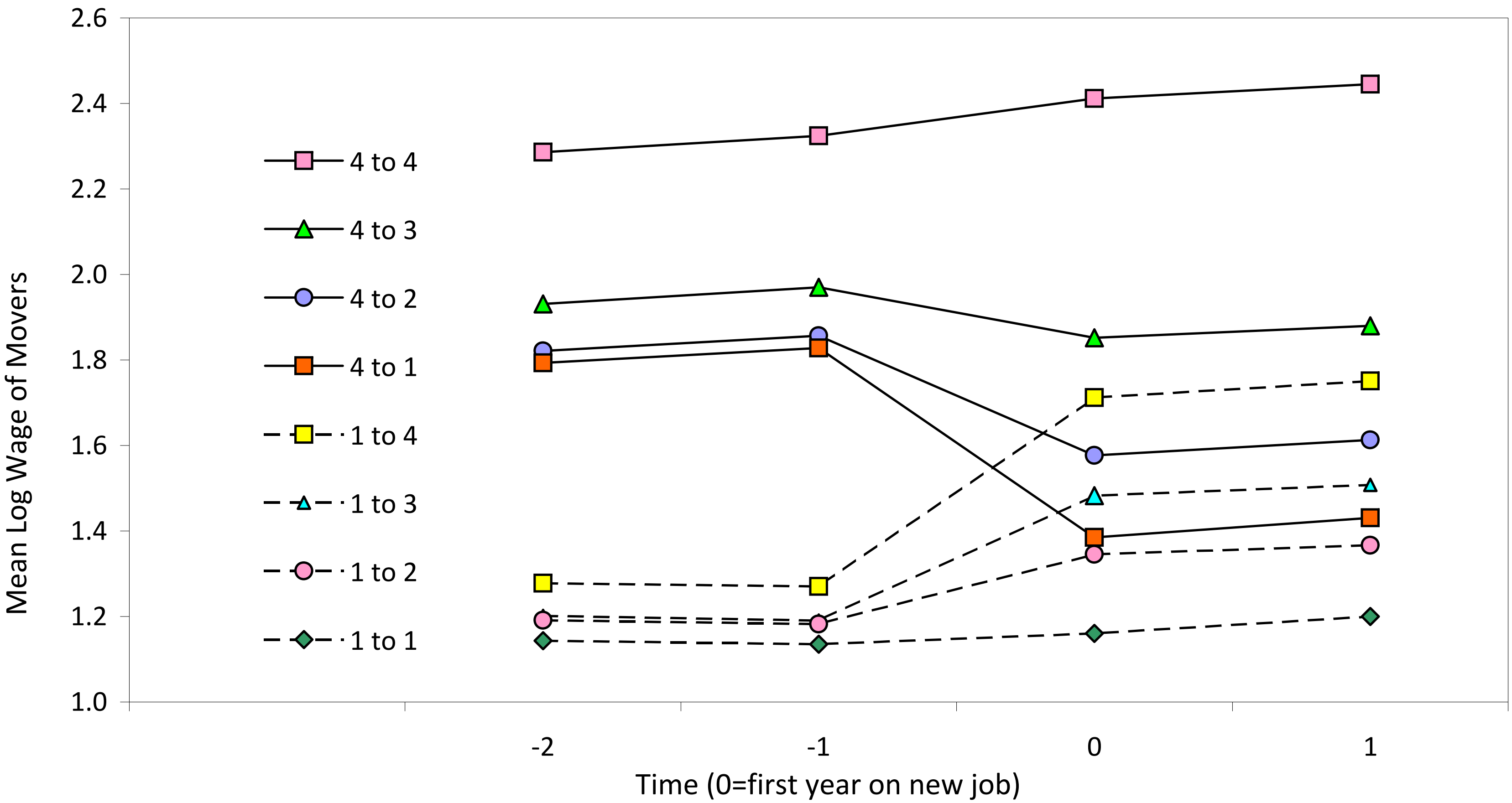
Strassner, Erich H. and Brian C. Moyer. 2002. "An Analysis of the Composition of Intermediate Inputs by Industry." U.S. Department of Commerce Bureau of Economic Analysis, Working Paper WP2002-05.

Stuhlmacher, Alice F and Amy E. Walters. 1999. "Gender differences in negotiation outcome: A meta-analysis". *Personnel Psychology*, 52: 653– 677.

Van Reenen, John. 1996. "The Creation and Capture of Rents: Wages and Innovation in a Panel of U.K. Companies." *Quarterly Journal of Economics* 111(1): 195-226.

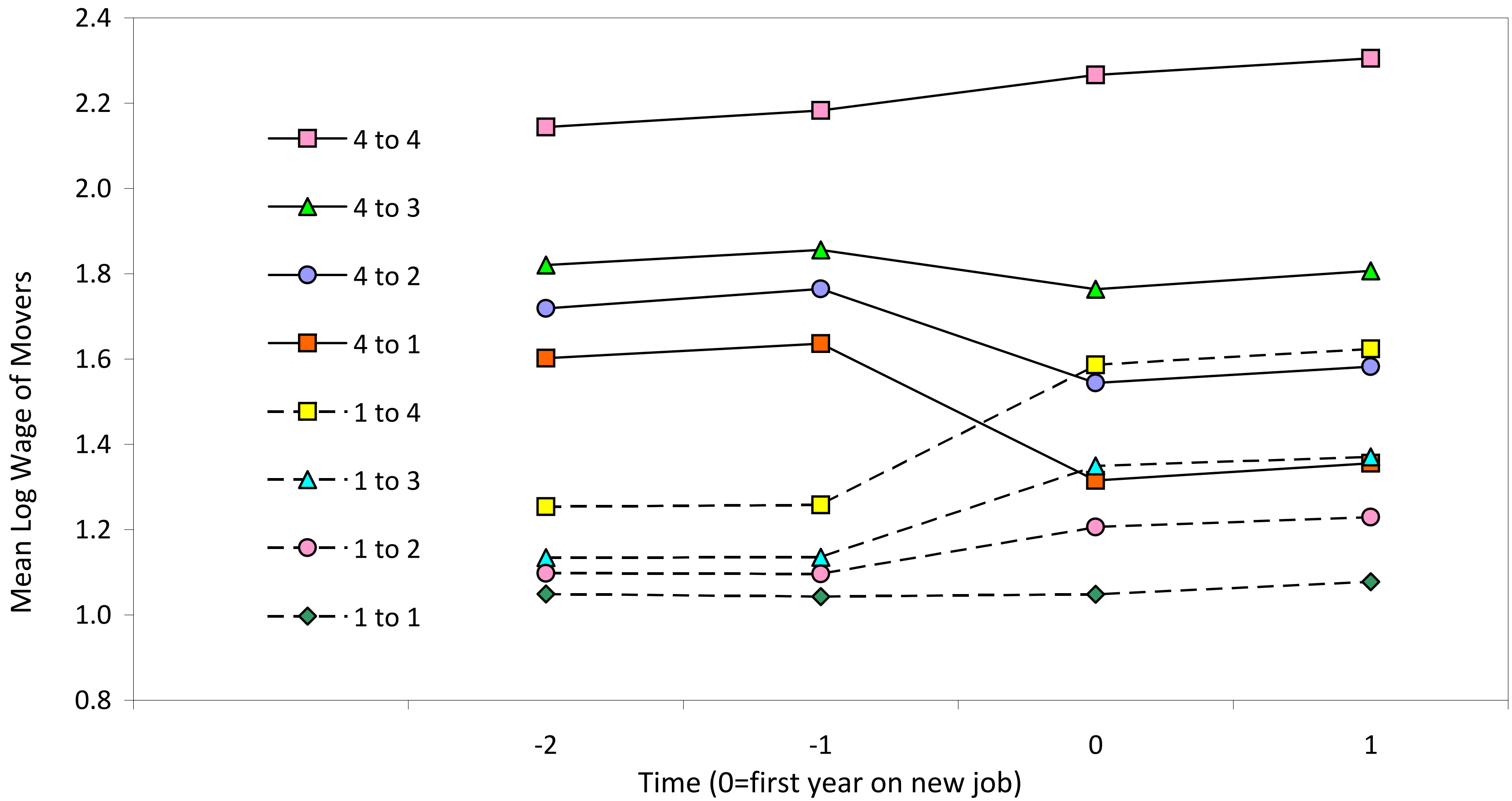
Wood, Robert G., Mary E. Corcoran and Paul N. Courant. 1993. "Pay Differences among the Highly Paid: The Male-Female Earnings Gap in Lawyers' Salaries." *Journal of Labor Economics* 11(2): 417-441.

Figure 1: Mean Log Wages of Male Job Changers, Classified by Quartile of Mean Co-Worker Wage at Origin and Destination Firm



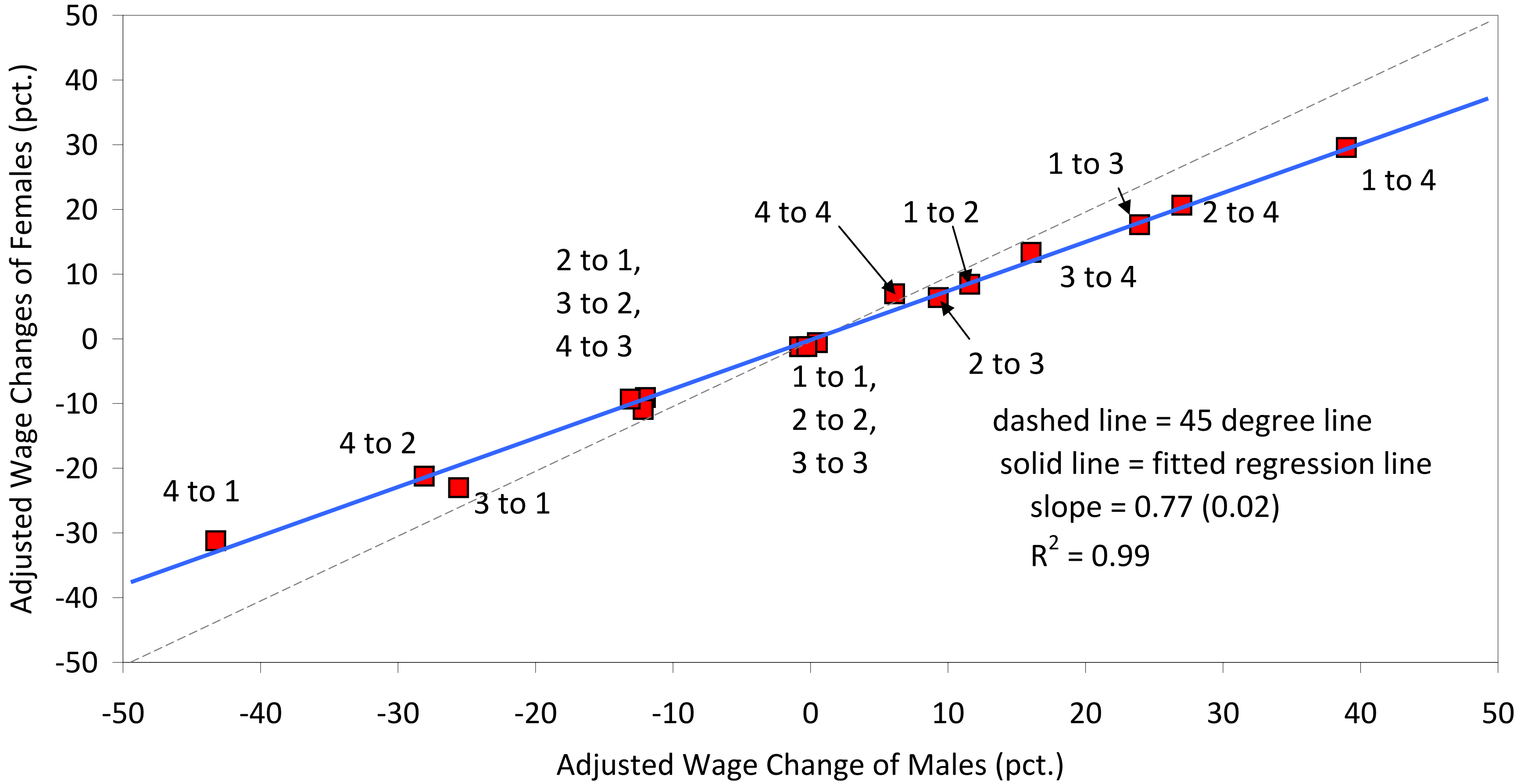
Notes: figure shows mean wages of male workers at mixed-gender firms who changed jobs in 2004-2007 and held the preceding job for 2 or more years, and the new job for 2 or more years. Each job is classified into quartiles based on mean log wage of co-workers (quartiles are based on coworker wages in last year on old job and first year on new job).

Figure 2: Mean Wages of Female Job Changers, Classified by Quartile of Mean Co-Worker Wage at Origin and Destination Firm



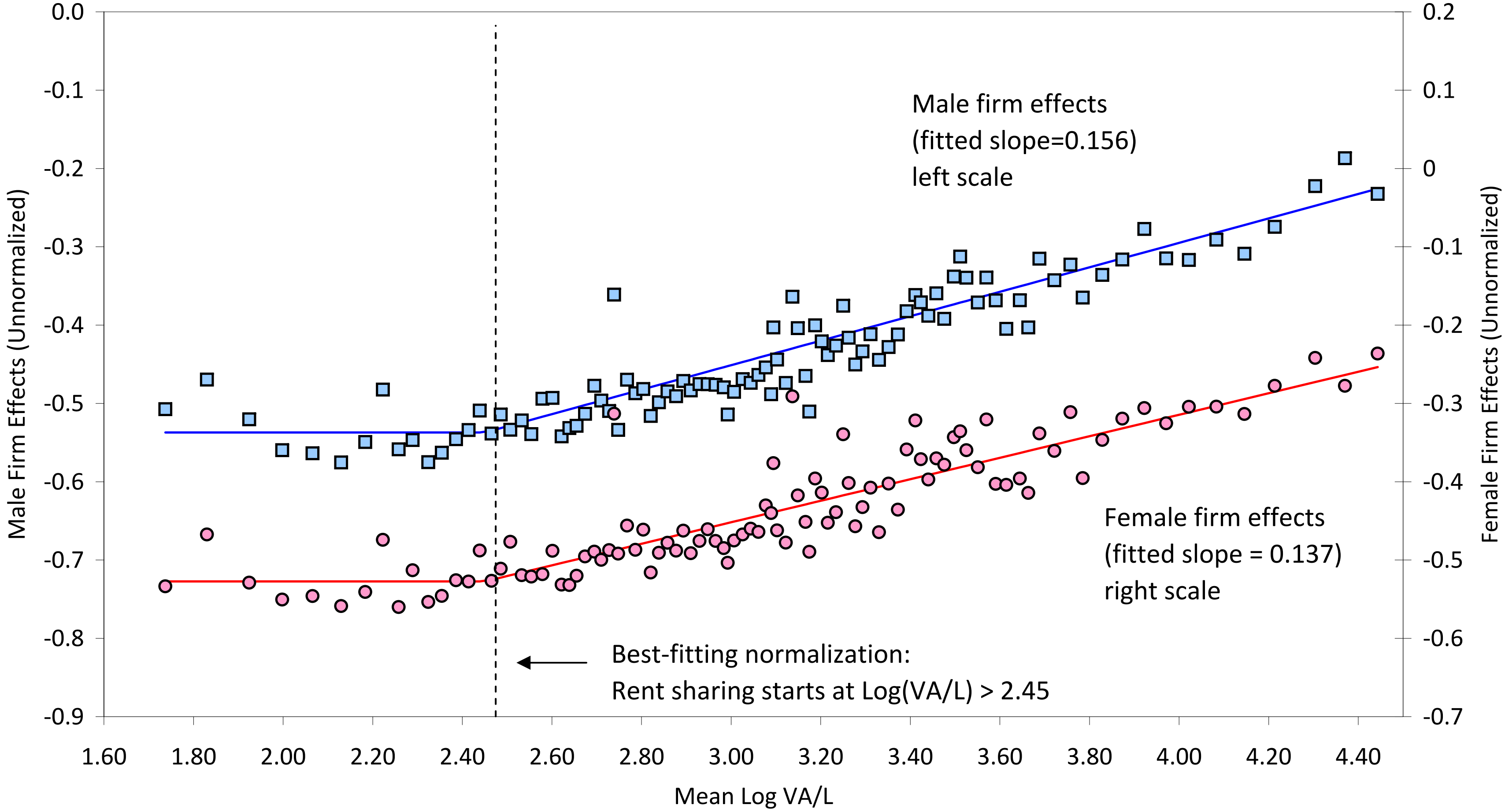
Notes: figure shows mean wages of female workers at mixed gender firms who changed jobs in 2004-2007 and held the preceding job for 2 or more years, and the new job for 2 or more years. Each job is classified into quartiles based on mean log wage of co-workers (quartiles are based on coworker wages in last year on old job and first year on new job).

Figure 3: Comparison of Adjusted Wage Changes of Male/Female Job Movers by Quartile of Coworker Wages of Origin and Destination Jobs



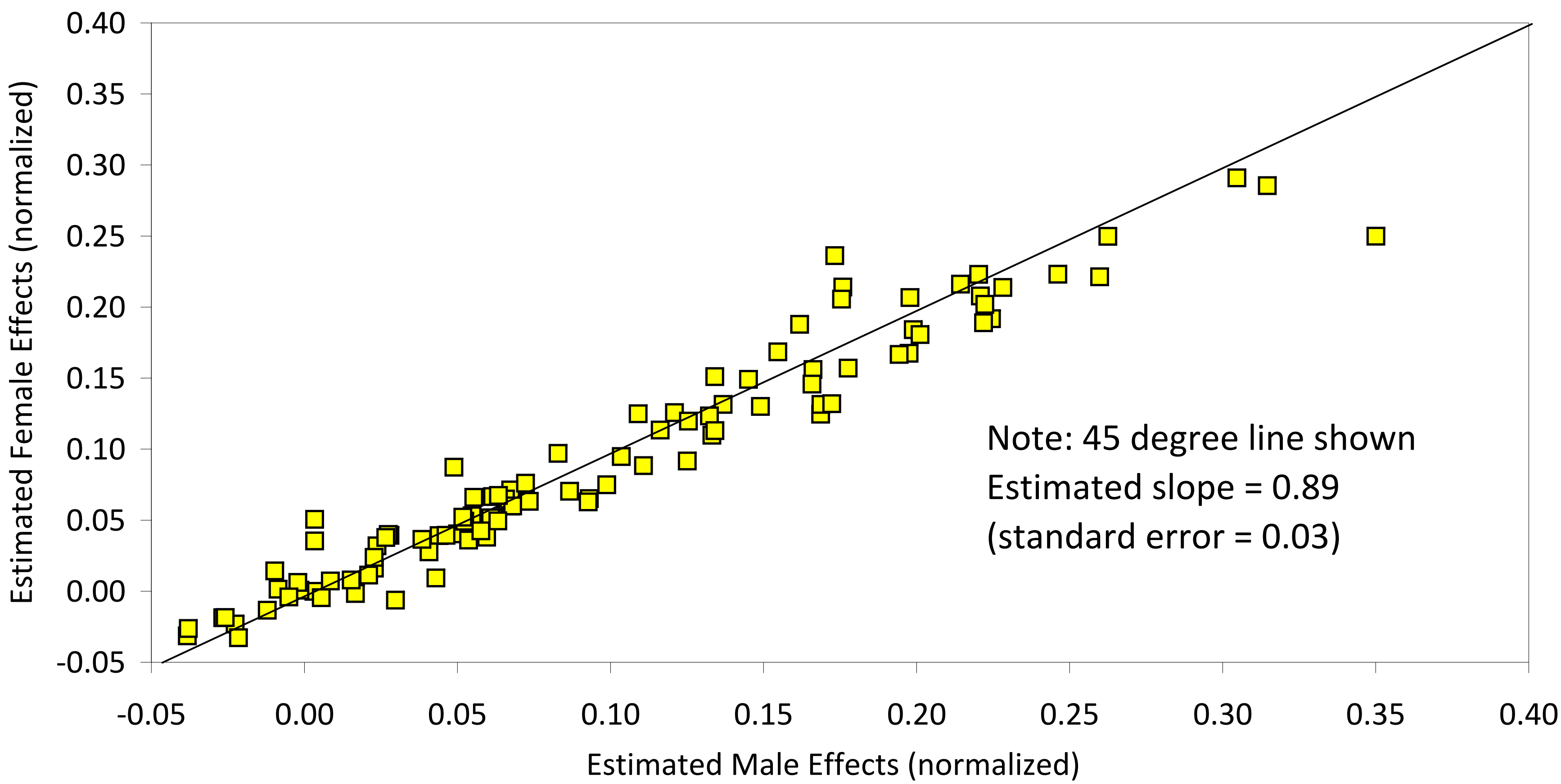
Notes: points represent regression adjusted mean log wage changes of male and female job movers in different origin/destination quartiles of mean coworker wages. For example "4 to 1" point shows mean wage changes for men and women who move from 4th quartile of coworker wages to 1st quartile. Fitted line is estimated by OLS to 16 points in the Figure.

Figure 4: Firm Fixed Effects vs. Log Value Added/Worker



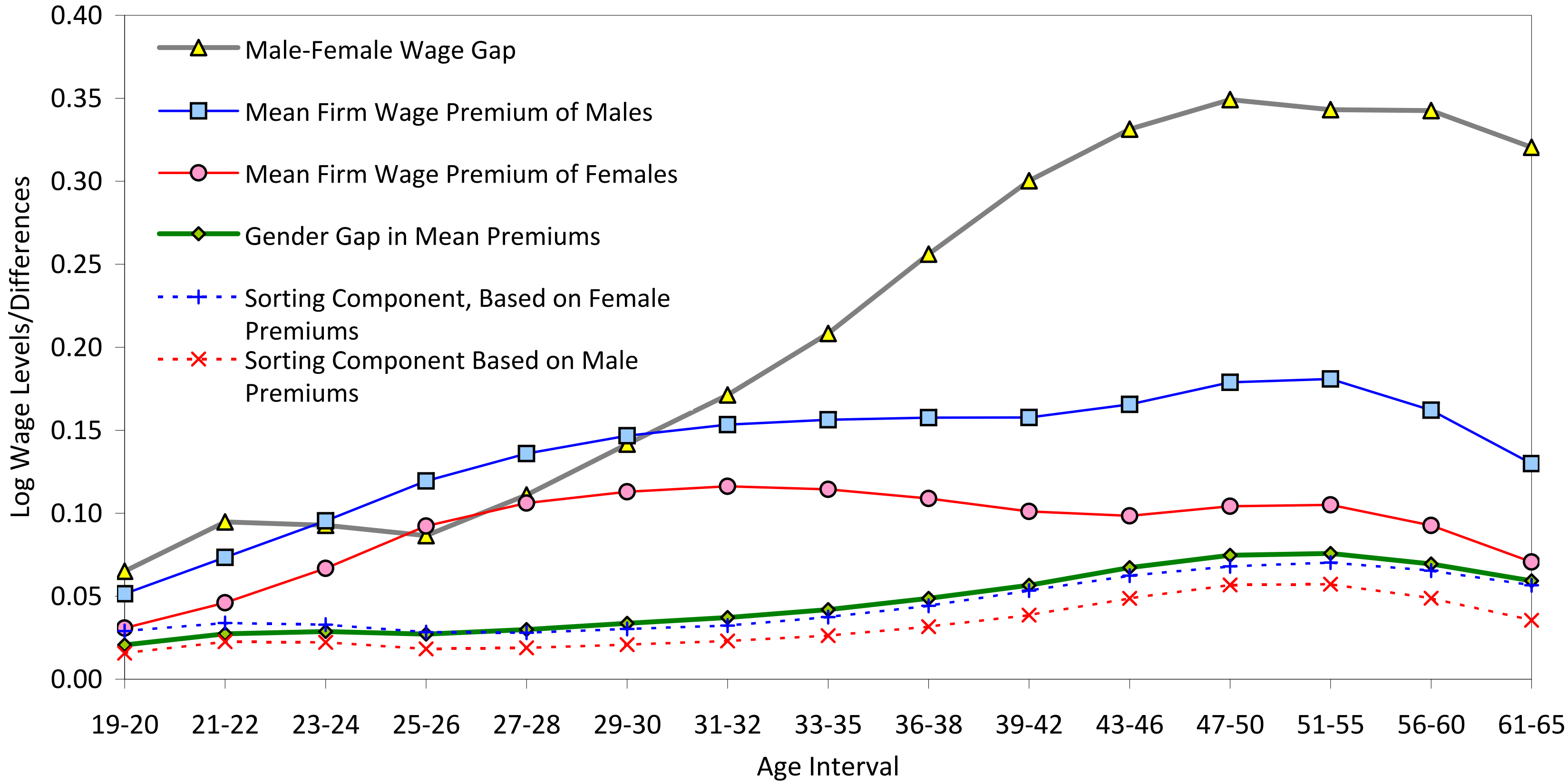
Note: points shown represent mean estimated firm-specific wage premiums from AKM models for men and women, averaged across firms with value added data available in 100 percentile bins of mean log value added per worker. See text for explanation of arbitrary normalization of the firm effects. which is imposed for estimation of the AKM models.

Figure 5: Estimated Firm Effects for Female and Male Workers:
Firm Groups Based on Mean Log VA/L



Note: figure shows bin scatter plot of estimated firm-specific wage premiums for female workers against estimated firm-specific wage premiums for male workers. Firm-level data is grouped into 100 percentile bins based on mean log value added per worker at the firm. Estimated slope is estimated across percentile bins by OLS.

Figure 6: Evolution of Wage Gap and Components Over the Lifecycle



Notes: figure shows unadjusted male-female wage gap, means of firm-specific wage premiums earned by males and females, and difference in mean premiums, which is the total contribution of firm-specific wage components to the gender wage gap. Dashed lines show the effect of differential sorting of males and females to specific firms, evaluated using male and female firm-specific wage premiums.

Table 1: Descriptive Statistics for Various Samples of Employees in QP, 2002-2009

	Overall Population of QP Employees:		Connected Sets of Workers/Firms: All		Connected Sets of Workers/Firms: Dual-Connected		Overall Population with Av. Value Added Data	
	Males	Females	Males	Females	Males	Females	Males	Females
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Age:								
Mean Age	38.1	36.9	38.0	36.5	38.0	36.4	37.9	36.3
Fraction ≤ 30 years old	0.30	0.33	0.30	0.34	0.30	0.34	0.31	0.35
Fraction ≥ 50 years old	0.19	0.14	0.18	0.13	0.19	0.13	0.18	0.13
Education:								
Mean Years Schooling	8.0	8.8	8.0	8.9	8.6	9.1	7.8	8.6
Fraction with High School	0.18	0.23	0.18	0.23	0.21	0.24	0.17	0.24
Fraction with Degree	0.09	0.13	0.09	0.14	0.11	0.15	0.07	0.11
Mean Log Real Hourly Wage	1.59	1.41	1.62	1.43	1.71	1.48	1.54	1.36
(standard dev.)	(0.55)	(0.50)	(0.55)	(0.51)	(0.58)	(0.53)	(0.48)	(0.43)
Mean Monthly Hours	162.6	158.0	162.5	157.9	162.8	157.1	164.1	159.9
(standard dev.)	(24.7)	(30.1)	(24.8)	(29.9)	(24.0)	(30.5)	(23.6)	(29.4)
Fraction in Lisbon	0.35	0.35	0.36	0.37	0.42	0.40	0.32	0.34
Fraction in Oporto	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.14
Mean Firm Size (No. emp's)	730	858	804	978	1,091	1,230	500	886
Fraction Females at Firm	0.24	0.70	0.24	0.70	0.30	0.64	0.24	0.67
Mean Log VA/Worker							3.05	2.88
Number person-year obs.	9,070,492	7,226,310	8,225,752	6,334,039	6,012,521	5,012,736	5,786,148	4,204,851
Number of persons	2,119,687	1,747,492	1,889,366	1,505,517	1,450,288	1,247,503	1,441,626	1,082,058
Number of firms	349,692	336,239	216,459	185,086	84,720	84,720	153,998	141,889

Notes: Overall sample in columns 1-2 includes paid workers age 19-65 with potential experience ≥1. Individuals with inconsistent employment histories are excluded. Wages are measured in real (2009=100) Euros per hour. Value added (VA) is measured in thousands of real Euros per year. All statistics are calculated across person-year observations. See text for definitions of connected and dual connected sets.

Table 2: Summary of Estimated Two-way Fixed Effects Models for Male and Female Workers

	All Males (1)	All Females (2)
Standard deviation of log wages	0.554	0.513
Number of person-year observations	8,225,752	6,334,039
Summary of Parameter Estimates:		
Number person effects	1,889,366	1,505,517
Number firm effects	216,459	185,086
Std. dev. of person effects (across person-yr obs.)	0.420	0.400
Std. dev. of firm effects (across person-yr obs.)	0.247	0.213
Std. dev. of Xb (across person-yr obs.)	0.069	0.059
Correlation of person/firm effects	0.167	0.152
RMSE of model	0.143	0.125
Adjusted R-squared of model	0.934	0.940
Comparison job-match effects model:		
Number of job-match effects	2,689,648	2,087,590
RMSE of match-effects model	0.128	0.113
Adjusted R-squared of match-effects model	0.946	0.951
Std. deviation of job match effect	0.062	0.054
Inequality decomposition of two-way fixed effects model:		
Share of variance of log wages due to:		
person effects	57.6	61.0
firm effects	19.9	17.2
covariance of person and firm effects	11.4	9.9
Xb and associated covariances	6.2	7.5
residual	4.9	4.4

Notes: See text. Models includes dummies for individual workers and individual firms, year dummies interacted with education dummies, and quadratic and cubic terms in age interacted with education dummies (total of 44 parameters). Comparison job-match effects models include dummies for each worker-firm job match as well as other covariates in basic model. Samples include only observations in largest connected set.

Table 3: Contribution of Firm-Level Pay Components to Gender Wage Gap at Dual Connected Firms

	Wage Gap (1)	Means of Firm Premiums: Male Prem. Among Men (2)	Means of Firm Premiums: Female Prem. Among Women (3)	Total Contribution of Firm Components (4)	Decompositions: Sorting: Using M Effects (5)	Decompositions: Sorting: Using F Effects (6)	Decompositions: Bargaining: Using M Distribution (7)	Decompositions: Bargaining: Using F Distribution (8)
All	0.234	0.148	0.099	0.049	0.035	0.047	0.003	0.015
				(21.2)	(14.9)	(19.9)	(1.2)	(6.3)
By Age Group:								
Up to age 30	0.099	0.114	0.087	0.028	0.019	0.029	-0.001	0.009
				(28.2)	(18.9)	(29.3)	1.2	(9.3)
Ages 31-40	0.228	0.156	0.111	0.045	0.029	0.040	0.004	0.016
				(19.7)	(12.6)	(17.8)	(1.9)	(7.0)
Over Age 40	0.336	0.169	0.099	0.069	0.050	0.064	0.005	0.019
				(20.6)	(15.0)	(19.1)	(1.5)	(5.6)
By Education Group:								
< High School	0.286	0.115	0.055	0.059	0.045	0.061	-0.002	0.015
				(20.8)	(15.6)	(21.4)	0.6	(5.2)
High School	0.262	0.198	0.137	0.061	0.051	0.051	0.010	0.010
				(23.3)	(19.6)	(19.5)	(3.8)	(3.7)
University	0.291	0.259	0.213	0.047	0.025	0.029	0.018	0.022
				(16.1)	(8.7)	(9.9)	(6.2)	(7.4)

Notes: Sample includes male and female workers in "dual connected" set -- see Table 1, columns 5-6. Estimated firm effects are from models described in columns 1 and 2 of Table 2. Entries in parentheses represent the percent of the overall male female wage gap (in column 1) that is explained by source described in column heading.

Table 4: Contribution of Firm-Level Pay Components to Gender Wage Gap, by Modal Occupation

		Means of Firm Premiums:		Total Contribution	Decompositions			
					Sorting		Bargaining	
	Wage Gap	Male Prem. Among Men	Female Prem. Among Women	of Firm Components	Using M Effects	Using F Effects	Using M Distribution	Using F Distribution
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
All	0.234	0.148	0.099	0.049	0.035	0.047	0.003	0.015
				(21.2)	(14.9)	(19.9)	(1.2)	(6.3)
By Modal Occupation								
Managers	0.251	0.217	0.205	0.012	-0.001	0.001	0.012	0.013
(69% male)				(4.9)	(-0.3)	(0.2)	(4.6)	(5.2)
Professionals	0.143	0.264	0.232	0.032	0.005	0.015	0.017	0.027
(47% male)				(22.0)	(3.5)	(10.4)	(11.6)	(18.5)
Technicians	0.135	0.249	0.189	0.060	0.038	0.040	0.020	0.022
(63% male)				(44.3)	(28.0)	(29.4)	(14.8)	(16.3)
Clerks	0.159	0.239	0.177	0.063	0.049	0.052	0.011	0.014
(40% male)				(39.5)	(30.8)	(32.6)	(6.9)	(8.7)
Services	0.162	0.065	0.052	0.013	0.013	0.023	-0.010	-0.001
(33% male)				(7.7)	(8.3)	(14.0)	(-6.3)	(-0.6)
Craft	0.405	0.094	0.016	0.078	0.048	0.084	-0.005	0.030
(70% male)				(19.3)	(11.8)	(20.7)	(-1.3)	(7.5)
Operatives etc	0.290	0.145	0.096	0.050	0.036	0.040	0.010	0.014
(72% male)				(17.1)	(12.3)	(13.8)	(3.3)	(4.8)
Elementary	0.190	0.100	0.052	0.048	0.030	0.057	-0.009	0.018
(49% male)				(25.3)	(16.0)	(29.9)	(-4.6)	(9.3)

Notes: see notes to Table 3. Workers are classified into their most common occupation during years they are observed. Farm and fishing workers are included with operatives.

Table 5: Contribution of Firm-Level Pay Components to Gender Wage Gap: All Workers versus Workers in "Female" and "Male" Occupations

	Means of Firm Premiums:		Total Contribution of Firm Components	Decompositions			
				Sorting		Bargaining	
Wage Gap	Male Prem. Among Men	Female Prem. Among Women		Using M Effects	Using F Effects	Using M Distribution	Using F Distribution
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
A. All Workers at Dual Connected Firms							
(6,012,521 males and 5,012736 females at 84,720 firms)							
0.234	0.148	0.099	0.049	0.035	0.047	0.003	0.015
			(21.2)	(14.9)	(19.9)	(1.2)	(6.3)
B. Workers with "Female" Occupations at Firms that have Males and Females in "Female" Occupations							
(1,572,387 males and 3,403,802 females at 43,239 firms							
0.240	0.127	0.097	0.031	0.026	0.043	-0.012	0.005
			(12.8)	(10.8)	(17.8)	(-5.1)	(1.9)
C. Workers with "Male" Occupations at Firms that have Males and Females in "Male" Occupations							
(2,935,719 males and 801,113 females at 21,969 firms)							
0.137	0.177	0.133	0.044	0.015	0.027	0.016	0.028
			(31.9)	(11.1)	(20.0)	(11.9)	(20.8)

Notes: see notes to Table 3. Sample in panel A includes male and female workers who are employed at dual connected firms. Sample in panel B includes male and female workers who mainly work in "female" occupations, and are employed at "dual-connected female-occupation" firms. Sample in panel C includes male and female workers who mainly work in "male" occupations, and are employed at "dual-connected male-occupation" firms. See text for explanation of dual connected sets. Workers are classified into gender-occupation groups based on the female share of the occupation(s) they hold in all years. Decompositions in panel A are based on estimated two-way fixed effects models fit to all men and all women; those in panel B are based on estimated models fit to men and women in mainly female occupations; those in panel C are based on estimated models fit to men and women in mainly male occupations.

Table 6: Estimated Relationship Between Estimated Firm Effects and Measures of Surplus per Worker

	Number Firms (1)	Regressions of Firm Effects on Measure of Surplus: Males (2)	Females (3)	Ratio : Col (3) / Col (4) (4)
Surplus Measure:				
1. Excess Mean Log Value Added per Worker	47,477	0.156	0.137	0.879
		(0.006)	(0.006)	(0.031)
2. Mean Log Sales per Worker	75,163	0.072	0.064	0.897
		(0.005)	(0.004)	(0.036)
3. Excess Mean Log Sales per Worker	75,163	0.092	0.081	0.883
		(0.006)	(0.006)	(0.038)

Notes: Columns 2-3 report coefficients of surplus measure indicated in row heading in regression models in which the dependent variables are the estimated firm effects for males or females. All specifications include a constant, and are estimated at the firm level, weighting by the total number of male and female workers at the firm. Ratios in column 4 are estimated by IV, treating average change in female wages as dependent variable, average change in male wages as endogenous explanatory variable, and change in surplus measure as the instrument. Robust standard errors in parentheses.

Table 7: Effects of Changes in Measured Surplus per Worker on Wages of Stayers

		Estimated Rent Sharing Coefficients:		
	Number Firms	Male Stayers	Female Stayers	Ratio
	(1)	(2)	(3)	(4)
Surplus Measure and Sample:				
1. Excess Mean Log Value Added per Worker (Winsorized at +/- 0.50). Sample = Stayers at Firms with VA data, 2006-9	33,104	0.049 (0.007)	0.045 (0.008)	0.911 (0.086)
2. Excess Mean Log Value Added per Worker (Not Winsorized). Sample = Stayers at Firms with VA data, 2006-9	33,104	0.035 (0.006)	0.031 (0.006)	0.894 (0.091)
3. Excess Mean Log Value Sales per Worker (Winsorized at +/- 0.50). Sample = Stayers at Firms with Sales data, 2005-8	44,266	0.021 (0.006)	0.018 (0.005)	0.876 (0.182)

Notes: Dependent variables are average change in wages of male or female workers at a firm (regression-adjusted for quadratic in age). Table entries are coefficients of the measured change in surplus per worker, as defined in row heading. Ratios in column 4 are estimated by IV, treating average change in female wages as dependent variable, average change in male wages as endogenous explanatory variable, and change in surplus measure as the instrument. Standard errors, clustered by firm, in parentheses.

Appendix A: Data Issues

Data Appendix (including description of procedure for matching QP and SABI

Table A1: Descriptive Statistics for Overall QP and Analysis Sample

Table A2: Matching Rates of Observations in QP to Firms in SABI

Table A3: Comparison of Models Relating Estimated Firm Effects for Male and Female Workers to Mean Log Value Added per Worker, Estimated on Subsamples with More Restrictive Matching Criteria Between QP and SABI

Appendix B: Supplementary Material

Figure B1: Trends in Real Hourly Wages of Men and Women

Figure B2: Job Survival Rates for New Jobs Starting 2002-2008

Figure B3: Regression-Adjusted Changes in Wages for Male Movers
Across Coworker Wage Quartiles

Figure B4: Regression-Adjusted Changes in Wages for Female Movers
Across Coworker Wage Quartiles

Figure B5: Mean Residuals for Males by Decile of Worker and Firm Effects

Figure B6: Mean Residuals for Females by Decile of Worker and Firm Effects

Figure B7: Goodness of Fit and Estimated Rent Sharing Coefficients for Alternative Normalization Thresholds

Table B1: Distributions of Number of Jobs Held in Sample Period, by Gender, and Mean Log Wage by Number of Jobs Held

Table B2: Wages of Job Changes for Movers with 2+ Years of Data Before/After Job Change

Table B3: Comparison of Firms in Dual Connected Set with Low Average Value Added per Worker with Other Firms

Table B4: Decompositions of Gender Wage Gap by Industry

Appendix C: Estimated Models for Subset of Workers with 12+ Years of Education, Age 25+

Appendix D: Models Based on Sales per Worker Instead of VA/L

Appendix A: Data Sources, Definitions, and Matching Procedure

a. Quadros de Pessoal

The Quadros de Pessoal (QP) is an annual census of employees at private sector firms in Portugal. Firms with at least one paid employee are required to submit information on their full workforce as of the survey reference week (in October). Individuals working as independent contractors are excluded from coverage. (Information on the QP survey universe and other details are reported in IZA, 2010). For multi-plant firms we aggregate employees at different establishments to the firm level to reflect the fact that our financial data are firm-wide, using the unique firm identifier that is available in the data.

The QP asks firms to report each employee's gender, education, occupation, and date of hire, as well as information on pay and hours. Pay rates in Portugal are normally expressed in Euros per month, net of all payroll taxes. We construct a "regular monthly salary" by summing an individual's monthly base salary and any reported regular salary supplements. The latter include payments for meals, tenure-related premiums, and other payments that are received regularly. Employers also report "normal" hours, which are monthly contractual hours, as set by the prevailing collective bargain or firm regulations, and do not include overtime. We define the hourly wage as (monthly base salary + regular supplements)/ normal hours of work. This corresponds to a "straight time" hourly wage exclusive of overtime, as is used for example in the Bureau of Labor Statistics' Occupational Employment Statistics (OES) Survey (see US Department of Labor, 2005).

The QP dataset for 2002-2009 includes over 20 million observations on 4.5 million workers. To construct our analysis sample we drop the entire history for a person if: (1) the hiring date for any job is missing or inconsistent across observations (0.6% of observations dropped); (2) the individual is observed in two consecutive years at different firms, but the hiring date for the second job is the same as the hiring date for the first job (6.9% of observations dropped); the hourly wage in any year is extremely high or extremely low (0.3% dropped); the change in the log hourly wage from one year to the next is less than -1 or greater than 1 (1.6% dropped). After these deletions we retain only person-year observations in which the worker is between the ages of 19 to 65 (1.6% of observations dropped), with at least two years of potential labor market experience (i.e., age-education-6>1) (0.7% dropped) and is employed as a wage-earner rather than an owner, or unpaid family worker (dropping 9.3% of observations).

Appendix Table A1 shows the characteristics of the male and female observations in the entire QP, and our analysis sample as summarized in Table 1 of the paper. The two samples are quite similar in terms of age, education, location, mean hourly wage, and mean monthly hours of work.

b. SABI

Bureau van Dijk's SABI (Sistema de Analisis de Balances Ibericos) data base has annual data for non-financial firms including: a firm tax identifier; income statement information (including sales revenue and other items); total employment; the firm's name, address, industry, shareholder capital; and date of formation. The data are available from 2000 onward, but coverage expanded substantially in 2005, and information on employment is missing for many firms prior to 2006.

Bureau van Dijk constructs value added (VA) for firms in SABI as follows:

VA = After-tax Profit + Employee Expenses (including pension costs) + Depreciation + Interest Paid + Taxes Paid.

This definition corresponds to the economic concept of value added, which is the sum of payments to labor and capital.

c. Matching QP and SABI

The following variables are common to QP and SABI and can be used to match observations for a given firm in a given year in the two data sets: (1) location -- zip code and county (concelho) in SABI, parish (freguesia) and county in QP; (2) 5 digit industry;

(3) year of firm creation; (4) shareholder capital; (5) annual sales. We do not use employment in our matching procedure, but we use it as a check variable.

In QP, total sales of the firm are reported for the previous calendar year. We therefore use sales in year t-1 from SABI to match observations between the two data sets in each year. In SABI, both sales and shareholder capital are reported in thousands of euro, whereas in QP they are reported in euros. We therefore round both variables in QP to the thousands. Sales and shareholder capital are treated as missing if the reported values are zero. The zip codes reported in SABI were converted to parishes, with the exception of a few codes that cross parish boundaries and a few that appear to be non-existent codes.

We use a multi-step matching procedure which uses exact matching at each stage, and sequentially relaxes the number of variables that have to match exactly. Firms that are matched at one step are removed from both data sets, leaving unmatched observations for the next step. The steps are as follows:

1. Exact match based on 5 variables: location, industry, year of firm creation, sales and shareholder capital. We first attempt an exact match using sales and shareholder capital for 2009 (the other variables are time-invariant), then work backwards to 2005. We initially use parish and 5-digit industry to look for exact matches. We then repeat the process using county and 3-digit industry.

2. Exact match based on 4 variables: location, industry, and any two of: year of firm creation, annual sales, or shareholder capital. As in step 1, we initially use parish and 5-digit industry to look for exact matches, then use county and 3-digit industry.

3. Exact match based on 3 variables: location, industry, and any one of: year of firm creation, annual sales, or shareholder capital. As in step 1, we initially use parish and 5-digit industry to look for exact matches, then use county and 3-digit industry.

In this step, once a potential match was found, we compared data from QP and SABI to check the plausibility of the match. Specifically, we checked annual observations on sales and shareholder capital for all years from 2005 to 2009 in which non-missing data were available in both data sets. A match was validated only if the deviation between SABI and QP did not exceed 1% in any year for either sales or shareholder capital, or, in cases with a larger deviation in any one year, if the values in all other years were exactly the same in both data sets.

4. Exact match based on 2 variables: location and any one of industry, year of firm creation, annual sales, or shareholder capital. As in step 3, potential matches were evaluated and only retained if the same criterion was met.

We successfully matched a total of 301,417 firms between QP and SABI - representing about 80% of the firms that ever appear in SABI, and 53% of firms that appear at least once (with a worker in our analysis sample) in QP from 2002 to 2009. Of the matches, 52% were matched on all five variables, 31% were matched on four variables,12% were matched on three variables, and the remaining 5% were matched on two varables. The match rate by firm size (based on average number of employees in QP) are as follows: 1-10 workers - 50.7%; 11-50 workers - 68.61%, 51-100 workers - 67.0%, 101-500 workers - 69.2%, over 500 workers - 61.0%. Appendix Table A2 shows the match rates by major industry and by gender, calculated across person-year observation in our main QP analysis sample.

As a check on the differential quality of the matches, we used different subsets of firms to estimate the relationship between the estimated gender-specific wage premiums at each firm, and mean log value added per worker. The results are reported in Appendix Table A3. The top row of the table reproduces the specifications from Table 6 of the paper, which are based on data from 47,477 firms. Row 2 shows the same specifications dropping firms matched in steps 3 and 4 of our matching procedure; row 3 shows the results after dropping firms matched in steps 2-4; and row 4 shows the estimated models using only the observations that were matched exactly on 5 variables in the first step of our process. The estimated effects of mean log value added per worker in the models for the estimated male and female wage premiums are slightly higher (+9%) when the lower-quality matches are dropped, but the ratio of the estimated effects for females and males is quite stable. Based on these results we do not believe that the inclusion of the lower quality matches has any substantive effect on our conclusions.

References

IZA (Institute for the Study of Labor). "Study Documentation for Quadros de Pessoal." Accessed at http://idsc.iza.org/metadata/PDF/401.pdf (April 19, 2015).

U.S. Department of Labor Bureau of Labor Statistics. "Appendix B. Survey Methods and Reliability Statement for the May 2005 Occupational Employment Statistics Survey." Accessed at http://www.bls.gov/oes/2005/may/appendixb.pdf (April 19, 2015).

Appendix Table A1: Descriptive Statistics for Overall QP and Analysis Sample

	Overall Population of Employees in QP		Analysis Sample	
	Males	Females	Males	Females
	(1)	(2)	(3)	(4)
Age:				
Mean Age	38.9	37.0	38.1	36.9
Fraction ≤ 30 years old	0.28	0.32	0.30	0.33
Fraction ≥ 50 years old	0.21	0.16	0.19	0.14
Education:				
Mean Years Schooling	8.0	8.8	8.0	8.8
Fraction with High School	0.18	0.23	0.18	0.23
Fraction with University Degree	0.10	0.14	0.09	0.13
Mean Log Real Hourly Wage	1.61	1.42	1.59	1.41
(standard dev.)	(0.58)	(0.52)	(0.55)	(0.50)
Mean Monthly Hours	161.9	156.7	162.6	158.0
(standard dev.)	(25.9)	(31.8)	(24.7)	(30.1)
Fraction in Lisbon	0.35	0.36	0.35	0.35
Fraction in Oporto	0.13	0.13	0.13	0.13
Mean Firm Size (Number employees)	668	839	730	858
Fraction Female Workers at Firm	0.25	0.66	0.24	0.70
Number person-year obs.	11,651,615	9,011,089	9,070,492	7,226,310
Number of persons	2,550,576	2,040,863	2,119,687	1,747,492
Number of firms	431,991	391,982	349,692	336,239

Notes: Overall sample in columns 1-2 includes all observations available in QP with consistent data for age, gender and education. Analysis sample in columns 3-4 excludes individuals with inconsistent employment histories. Person-year observations are also conditioned on being a paid worker in the year, age 19-65, with potential experience ≥2. Wages are measured in real (2009=100) Euros per hour. Lisbon refers to Greater Lisbon and Setubal, Oporto refers to Greater Oporto (NUTS-3 classifications).

Appendix Table A2: Matching Rates of Observations in QP to Firms in SABI

	Percent of All Observations in Industry	Percent of Observations Matched	Percent of Observations from Females	Percent Matched by Gender	
				Male	Female
	(1)	(2)	(3)	(4)	(5)
All Industries	100.0	70.9	44.3	73.3	67.8
Agriculture	1.7	52.0	40.6	50.9	53.7
Fishing	0.1	84.2	27.1	82.2	89.5
Mining	0.4	80.8	9.7	80.5	83.7
Food Products	3.5	75.2	49.1	74.9	75.6
Textiles	8.0	81.3	71.7	80.9	81.5
Wood Products	2.8	78.3	27.0	76.7	82.6
Paper	1.5	79.3	34.5	77.4	82.8
Chemicals	1.8	82.2	34.2	80.8	84.9
Other Mineral Products	2.0	81.4	29.3	81.0	82.4
Metal Fabrication	7.0	80.4	25.8	80.0	81.6
Utilities	0.8	86.5	17.7	87.3	82.7
Construction	12.4	69.5	8.3	69.0	74.1
Trade	19.5	79.4	46.6	79.8	78.9
Hotels	6.6	75.7	62.4	78.2	74.3
Transportation	5.9	71.9	22.9	72.9	68.5
Finance	2.8	27.5	44.4	27.6	27.3
Business Services	10.6	82.6	49.1	83.5	81.6
Education	2.1	42.8	76.2	44.1	42.5
Health	5.9	35.9	88.3	44.1	34.8
Recreation Services	1.0	64.2	44.2	66.4	61.5
Other	3.6	30.6	67.1	32.4	29.7

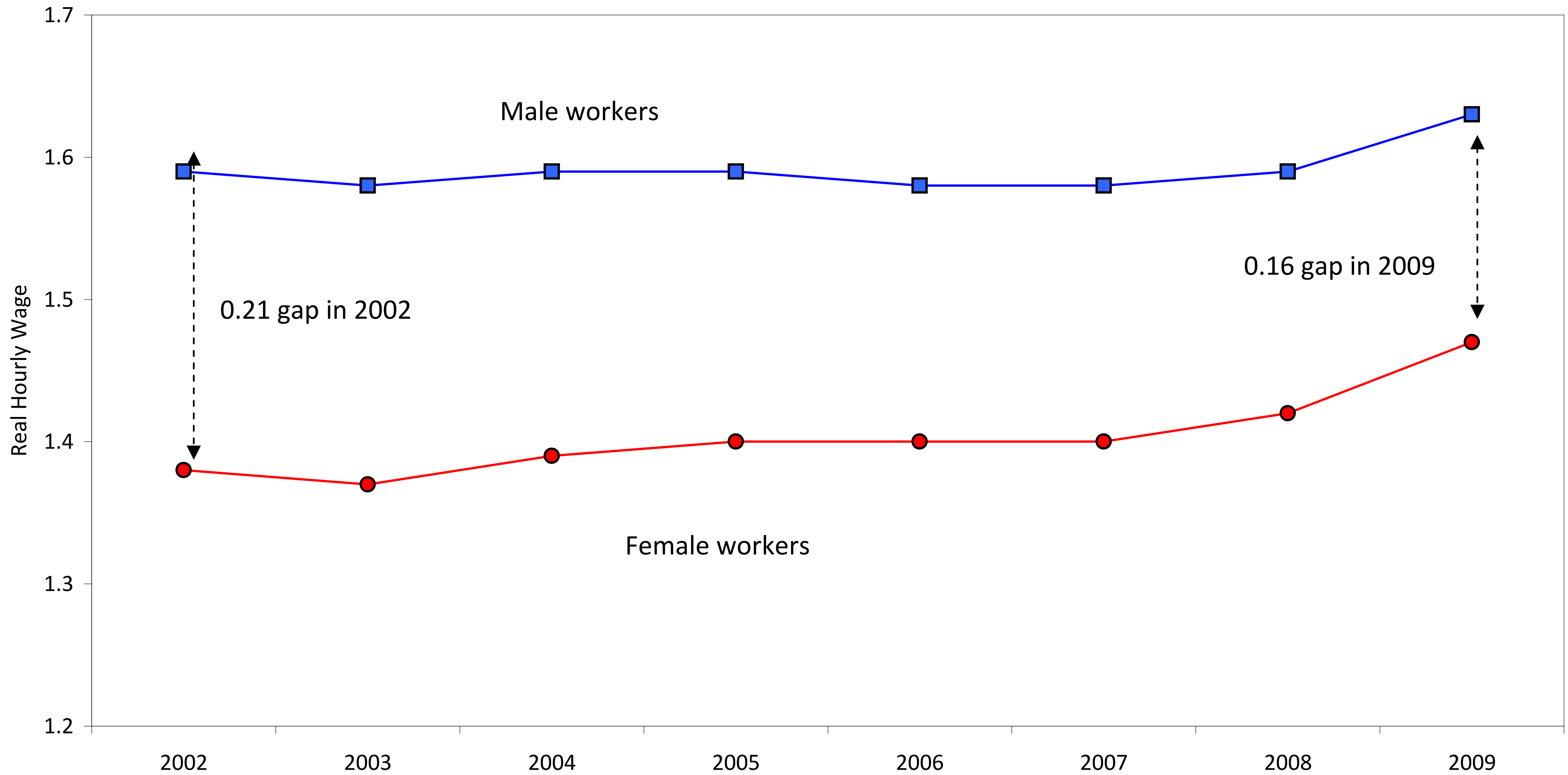
Note: All statistics are calculated across person-year observations in QP analysis sample for 2002-2009. "Matched" means that employer of person in given year can be matched to firm in SABI. Sample contains 9,070,492 person-year observations for males and 7,226,310 for females.

Appendix Table A3: Comparison of Models Relating Estimated Firm Effects for Male and Female Workers to Mean Log Value Added per Worker, Estimated on Subsamples with More Restrictive Matching Criteria Between QP and SABI

	Number Firms	Regressions of Firm-Specific Wage Premiums on log(VA/L)		Ratio: Females to Males
		Males	Females	
	(1)	(2)	(3)	(4)
1. All Available Matched Firms	47,477	0.156	0.137	0.879
		(0.006)	(0.007)	(0.031)
2. Firms Matched with 3 or More Exact Matching Variables	46,679	0.159	0.139	0.878
		(0.006)	(0.006)	(0.031)
3. Firms Matched with 4 or More Exact Matching Variables	44,552	0.162	0.143	0.878
		(0.006)	(0.006)	(0.031)
5. Firms Matched with 5 Exact Matching Variables	30,023	0.170	0.148	0.868
		(0.006)	(0.007)	(0.039)

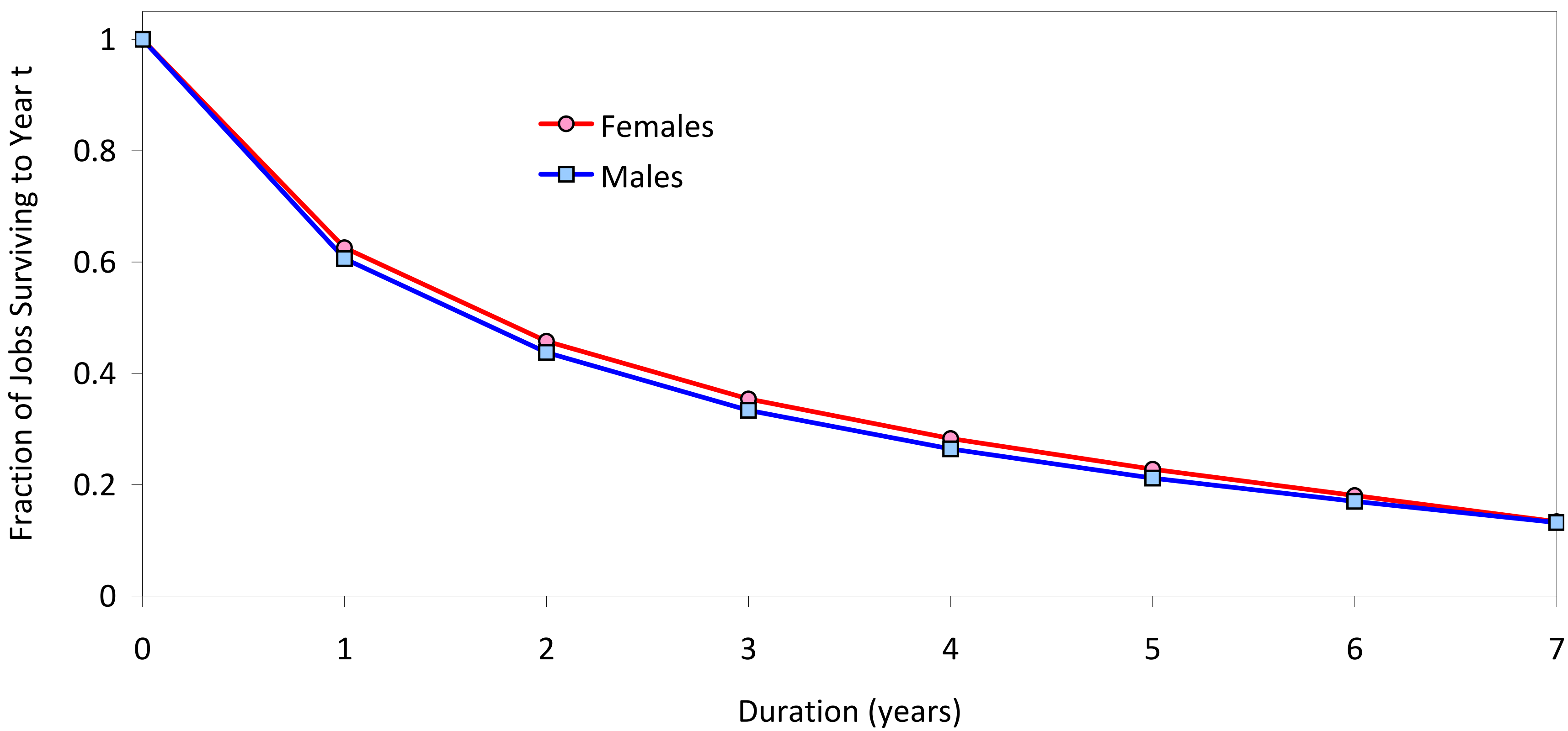
Notes: Columns 2-3 report coefficients of mean log value-added per worker in excess of 2.4 in regression models in which the dependent variables are the estimated firm-specific wage premiums for the gender group identified in the column headings. All specifications include a constant. Models are estimated at the firm level, weighted by the total number of male and female workers at the firm. Ratio estimates in columns 4 are obtained by IV method. Standard errors in parentheses.

Appendix Figure B1: Trends in Real Hourly Wage of Men and Women



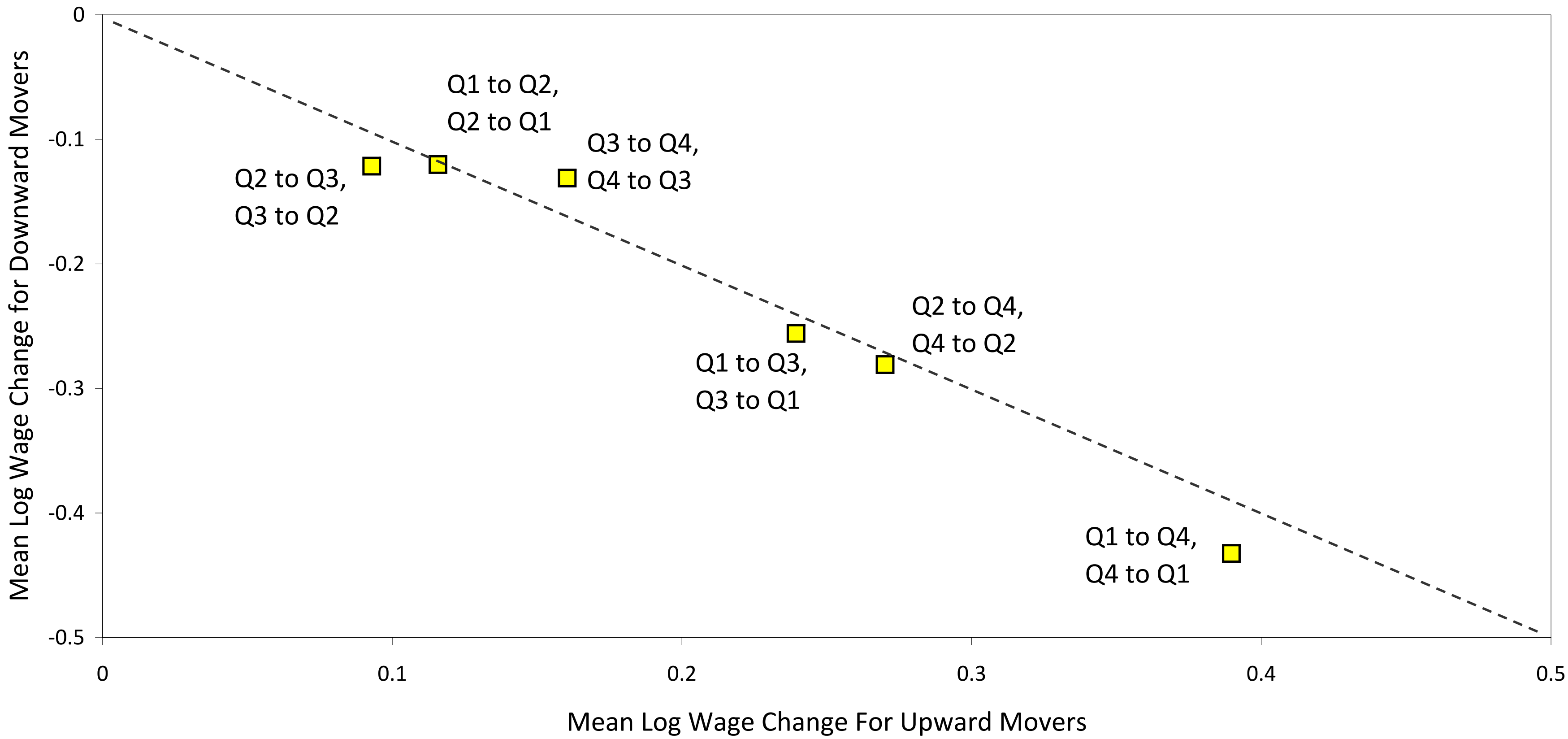
Note: figure plots mean log wages of male and female workers in main analysis sample by year.

Appendix Figure B2: Job Survival Rates for New Jobs Starting 2002-2008



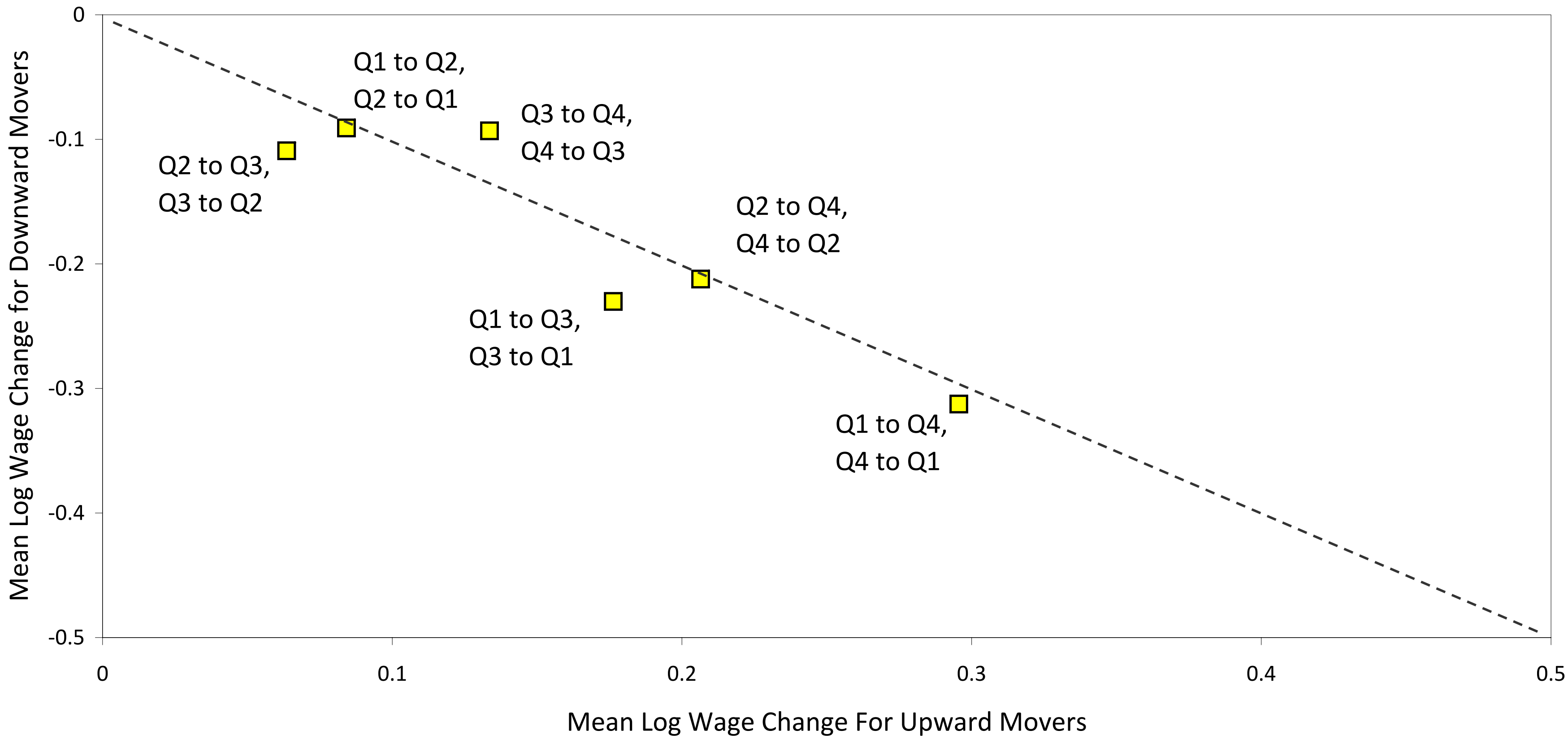
Note: figure plots Kaplan-Meier estimates of job survival rates for male and female employees starting new jobs in 2002-2008 period.

Appendix Figure B3: Regression-Adjusted Changes in Wages for Male Movers
Across Coworker Wage Quartiles



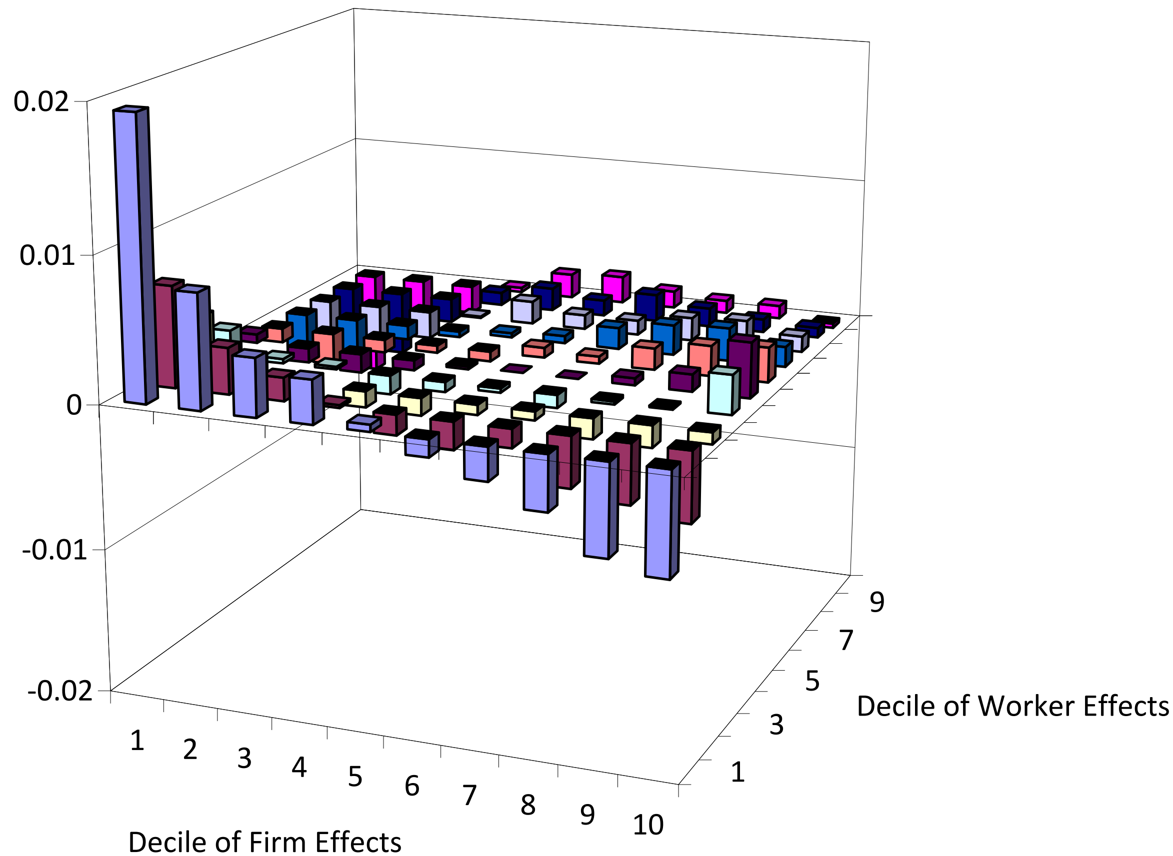
Note: Figure plots regression adjusted mean wage changes over 4 year interval for job changers who move across coworker wage quartile groups indicated. Dashed line represents symmetric changes for upward and downward movers.

Appendix Figure B4: Regression-Adjusted Changes in Wages for Female Movers
Across Coworker Wage Quartiles



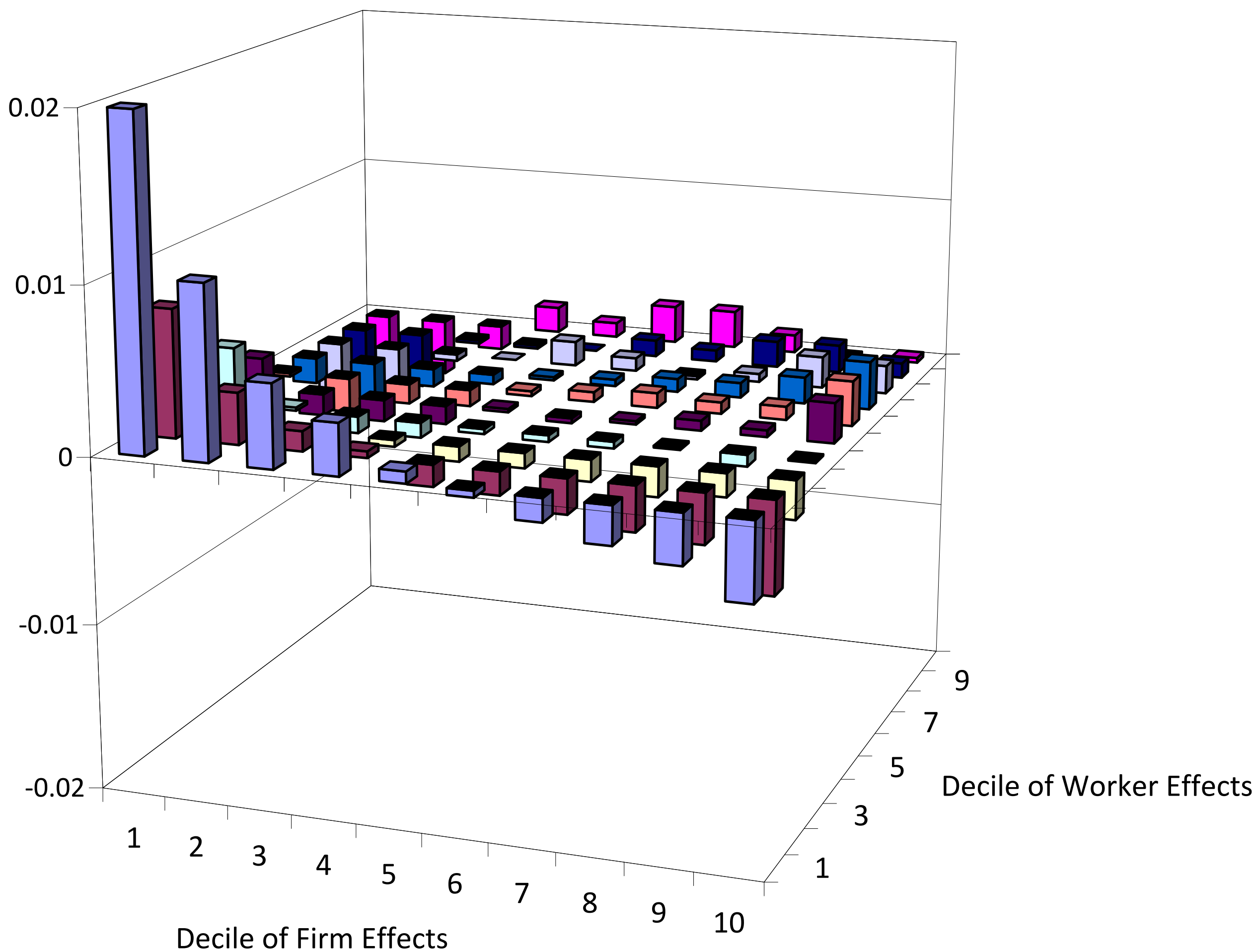
Note: Figure plots regression adjusted mean wage changes over 4 year interval for job changers who move across coworker wage quartile groups indicated. Dashed line represents symmetric changes for upward and downward movers.

Appendix Figure B5: Mean Residuals for Males by Decile of Worker and Firm Effects



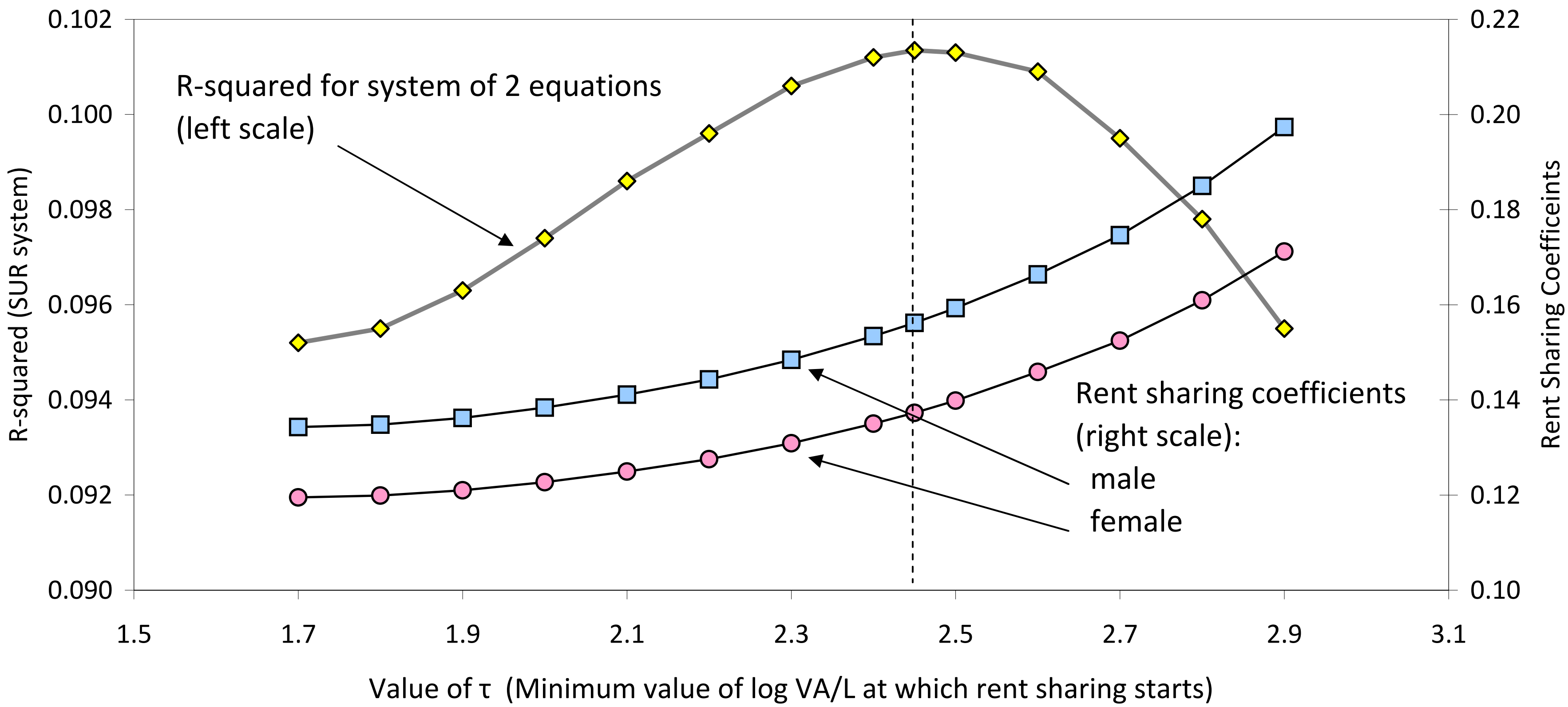
Note: figure plots mean residuals from wage model for male workers (column 1 of Table 2), for 100 bins, classified by decile of estimated firm effect and decile of estimated worker effect.

Appendix Figure B6: Mean Residuals for Females by Decile of Worker and Firm Effects



Note: figure plots mean residuals from wage model for female workers (column 2 of Table 2), for 100 bins, classified by decile of estimated firm effect and decile of estimated worker effect.

Appendix Figure B7: Goodness of Fit and Estimated Rent Sharing Coefficients for Alternative Choices of Minimum Rent Sharing Threshold τ



Note: figure shows R-squared for system of 2 equations and coefficient estimates of male and female rent sharing models for alternative values of parameter τ. Best fitting value is τ=2.45, shown by vertical line.

Appendix Table B1: Distributions of Number of Jobs Held in Sample Period, by Gender, and Mean Log Wage by Number of Jobs Held

	Distribution of Number of Jobs Held 2002-2009 (Person-year weighed)		Distribution of Number of Jobs Held 2002-2009 (Person-weighed)		Mean Log Wage of Persons, By Number of Jobs Held 2002-2009		
	Males	Females	Males	Females	Males	Females	Male-Female Gap
# Jobs	(1)	(2)	(3)	(4)	(5)	(6)	(7)
1	67.81	70.37	72.50	74.29	1.56	1.38	0.17
2	20.93	20.42	18.71	18.51	1.45	1.31	0.15
3	7.91	6.84	6.39	5.53	1.43	1.29	0.14
4	2.52	1.87	1.85	1.35	1.41	1.28	0.13
5	0.68	0.41	0.46	0.27	1.39	1.27	0.12
6	0.13	0.08	0.08	0.05	1.39	1.26	0.13
7	0.02	0.01	0.01	0.01	1.37	1.22	0.14
8	0.00	0.00	0.00	0.00	1.39	1.48	-0.09
# Obs.	9,070,492	7,226,310	2,119,687	1,747,492	2,119,687	1,747,492	--

Notes: tabulations based on overall population of male and female employees in QP data set -- see columns 1 and 2 of Table 1. There are 15 males and 7 females with 8 jobs in the sample, accounting for 120 person-year observations for men and 56 person-year observations for women.

Appendix Table B2: Wages of Job Changes for Movers with 2+ Years of Data Before/After Job Change

Origin/ destination quartile	Number Changes (1)	Pct. Of Changes (2)	Mean Log Real Wages of Movers: 2 years before (3)	1 year before (4)	1 year after (5)	2 years after (6)	3 Year Change (%): Raw (7)	Adjusted* (8)	(Std Err) (9)
Males									
1 to 1	13,787	43.2	1.14	1.14	1.16	1.20	5.6	0.5	(0.5)
1 to 2	9,139	28.7	1.19	1.18	1.35	1.37	17.6	11.6	(0.6)
1 to 3	6,283	19.7	1.20	1.19	1.48	1.51	30.6	23.9	(0.7)
1 to 4	2,682	8.4	1.28	1.27	1.71	1.75	47.3	39.0	(1.2)
2 to 1	7,293	21.2	1.34	1.35	1.22	1.27	-6.5	-12.0	(0.6)
2 to 2	12,326	35.8	1.37	1.38	1.40	1.42	5.0	-0.8	(0.6)
2 to 3	10,356	30.0	1.41	1.42	1.54	1.57	15.9	9.3	(0.5)
2 to 4	4,496	13.0	1.49	1.49	1.81	1.84	35.3	27.0	(0.9)
3 to 1	4,356	11.9	1.49	1.52	1.24	1.30	-19.4	-25.6	(0.7)
3 to 2	8,835	24.2	1.54	1.55	1.45	1.48	-5.8	-12.2	(0.6)
3 to 3	15,107	41.3	1.61	1.63	1.65	1.67	6.4	-0.3	(0.5)
3 to 4	8,246	22.6	1.73	1.75	1.94	1.97	24.7	16.0	(0.7)
4 to 1	1,634	5.4	1.79	1.83	1.39	1.43	-36.2	-43.3	(1.6)
4 to 2	3,245	10.7	1.82	1.86	1.58	1.61	-20.9	-28.1	(1.2)
4 to 3	6,589	21.7	1.93	1.97	1.85	1.88	-5.2	-13.1	(0.9)
4 to 4	18,830	62.1	2.29	2.32	2.41	2.45	15.9	6.1	(0.9)
Females									
1 to 1	24,130	60.9	1.05	1.04	1.05	1.08	2.9	-0.6	(0.4)
1 to 2	9,094	23.0	1.10	1.10	1.21	1.23	13.2	8.4	(0.5)
1 to 3	4,490	11.3	1.13	1.14	1.35	1.37	23.6	17.6	(0.6)
1 to 4	1,888	4.8	1.25	1.26	1.59	1.62	37.0	29.6	(1.2)
2 to 1	6,705	29.8	1.20	1.22	1.12	1.16	-4.5	-9.1	(0.5)
2 to 2	7,711	34.3	1.26	1.28	1.28	1.31	4.2	-1.2	(0.5)
2 to 3	5,495	24.5	1.33	1.35	1.44	1.46	12.6	6.4	(0.8)
2 to 4	2,562	11.4	1.44	1.45	1.69	1.73	29.0	20.7	(0.9)
3 to 1	3,283	16.7	1.38	1.40	1.15	1.20	-17.4	-23.0	(1.3)
3 to 2	4,762	24.2	1.42	1.45	1.34	1.37	-4.5	-10.9	(1.1)
3 to 3	7,245	36.8	1.51	1.53	1.54	1.56	5.3	-1.2	(0.7)
3 to 4	4,381	22.3	1.64	1.66	1.81	1.86	22.0	13.4	(0.9)
4 to 1	1,014	6.2	1.60	1.64	1.32	1.36	-24.6	-31.3	(2.8)
4 to 2	1,516	9.2	1.72	1.76	1.54	1.58	-13.7	-21.2	(1.3)
4 to 3	2,844	17.3	1.82	1.86	1.76	1.81	-1.3	-9.3	(0.9)
4 to 4	11,064	67.3	2.14	2.18	2.27	2.31	16.1	7.0	(0.8)

Notes: entries are mean log real daily wages for job changers to/from mixed-gender firms with at least 2 years of wages at the old job and the new job. Origin/destination quartiles are based on mean wages of coworkers in year before (origin) or year after (destination) job move.

*Four year wage change is regression-adjusted, using coefficient estimates from model fit to job stayers. Model includes dummies for age and education, and quadratic in age fully interacted with education. Standard errors of adjusted changes (in column 9) are two-way clustered across workers and firms and account for sampling errors in regression adjustment model using method of Kline (2014).

Appendix Table B3: Comparison of Firms in Dual Connected Set with Low Average Value Added per Worker with Other Firms

	Firms with Low VA/L	Firms with Higher VA/L	Firms with Missing VA/L
Mean Firm Characteristics (Each firm weighted equally)			
Mean Size (# employees per year)	13.4	22.7	18.4
Median Size (# employees per year)	4.9	7.4	6.0
Mean Education of Workers	7.7	9.0	8.7
Mean Percent Female	52.1	41.9	49.3
Mean Log Wage of Workers	1.19	1.48	1.39
Mean Log Sales/Worker	3.35	4.24	3.81
Industry Distribution:			
Pct. in Hotels and Restaurants	28.3	10.0	16.1
Pct. in Textiles and Apparel	11.2	3.9	5.2
Pct. in Trade	18.7	30.4	23.4

Note: sample includes firms in dual connected set only. There are 84,720 firms in total in this set: 11,731 with low mean log value added per worker (VA/L), 35,746 with higher VA/L. and 37,243 without data on VA/L. Low VA/L is defined as having VA/L < 2.45.

Appendix Table B4: Contribution of Firm Level Pay Components to Gender Wage Gap at Firms in Different Industries

			Male-Female Wage Gap		Mean Firm Premiums			Decompositions:			
								Sorting		Bargaining	
Industry	Fraction of All Workers in Industry	Female Share	Unadjusted	Adjusted*	Male Premium Among M's	Female Premium Among F's	Total Contribution of Firm Components	Using M Effects	Using F Effects	Using M Effects	Using F Effects
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
All	100.0	45.5	0.234	0.243	0.148	0.099	0.049	0.035	0.047	0.003	0.015
Agriculture	1.2	48.7	0.218	0.212	0.051	0.022	0.029	0.030	0.019	0.010	-0.002
Food Products	4.3	48.9	0.315	0.274	0.119	0.041	0.078	0.038	0.041	0.037	0.041
Textiles	8.5	66.7	0.263	0.204	0.048	0.013	0.035	0.002	0.032	0.003	0.033
Wood Products	2.5	36.4	0.202	0.205	0.052	0.053	-0.001	-0.021	0.006	-0.007	0.020
Paper and Publ.	1.7	34.9	0.220	0.299	0.211	0.151	0.061	0.038	0.026	0.034	0.023
Chemicals	2.3	34.8	0.254	0.294	0.284	0.181	0.104	0.061	0.042	0.061	0.043
Non-Met. Minerals	2.1	34.8	0.323	0.305	0.201	0.123	0.078	0.049	0.084	-0.006	0.029
Metals	8.1	30.0	0.253	0.245	0.153	0.165	-0.012	-0.012	-0.007	-0.005	0.001
Utilities	1.1	16.8	0.101	0.251	0.406	0.293	0.112	0.099	0.030	0.082	0.014
Construction	8.2	10.7	-0.042	0.234	0.071	0.097	-0.026	-0.015	-0.002	-0.024	-0.011
Trade	17.9	49.2	0.211	0.211	0.106	0.083	0.023	0.015	0.015	0.008	0.008
Hotels and Rest.	6.7	59.1	0.164	0.152	0.003	0.019	-0.015	0.001	0.011	-0.027	-0.017
Transport	6.9	25.4	-0.083	0.138	0.288	0.312	-0.024	-0.050	-0.033	0.008	0.026
Finance	4.0	43.4	0.207	0.181	0.416	0.390	0.025	0.008	0.006	0.019	0.017
Business Services	12.4	46.2	0.218	0.202	0.123	0.066	0.057	0.046	0.067	-0.010	0.011
Education	2.1	72.2	0.234	0.128	0.307	0.241	0.065	0.005	0.016	0.050	0.061
Health	5.3	85.3	0.243	0.169	0.087	0.052	0.035	0.030	0.043	-0.007	0.006
Recreation	1.1	43.2	0.217	0.244	0.268	0.193	0.075	0.056	0.036	0.039	0.019
Other	3.3	61.7	0.132	0.170	0.179	0.152	0.027	0.008	0.017	0.010	0.019

Notes: See note to Table 3. Adjusted wage gap is gap in adjusted wages, obtained by fitting OLS model to log wages of males (with controls for years of education, cubic in potential experience and year dummies) and using this model to obtain residual wages for both men and women. Fishing and mining industries, which together account for 0.4 percent of employment, are excluded from table.

Appendix Table B5: Contribution of Firm-Level Pay Components to Gender Wage Gap -- Alternative Normalization

		Means of Firm Premiums:			Decompositions			
					Sorting		Bargaining	
	Wage Gap	Male Prem. Among Men	Female Prem. Among Women	Total Contribution of Firm Components	Using M Effects	Using F Effects	Using M Distribution	Using F Distribution
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
All	0.234	0.146	0.076	0.071	0.035	0.047	0.024	0.035
				(30.2)	(15.0)	(19.9)	(10.2)	(15.2)
By Age Group:								
Up to age 30	0.099	0.112	0.063	0.049	0.019	0.029	0.020	0.030
				(49.3)	(18.9)	(29.3)	(20.0)	(30.4)
Ages 31-40	0.228	0.154	0.088	0.066	0.029	0.040	0.025	0.037
				(28.9)	(12.6)	(17.8)	(11.1)	(16.3)
Over Age 40	0.336	0.166	0.076	0.090	0.050	0.064	0.026	0.040
				(26.8)	(15.0)	(19.1)	(7.7)	(11.8)
By Education Group:								
< High School	0.286	0.112	0.032	0.080	0.045	0.061	0.019	0.036
				(28.1)	(15.6)	(21.4)	(6.7)	(12.5)
High School	0.262	0.196	0.114	0.082	0.051	0.051	0.031	0.031
				(31.3)	(19.6)	(19.5)	(11.8)	(11.7)
University	0.291	0.257	0.189	0.068	0.025	0.029	0.039	0.043
				(23.3)	(8.7)	(9.9)	(13.4)	(14.6)

Notes: see Table 4. Numbers in parentheses represent the percent of the overall male female wage gap (in column 3) that is explained by source described in column heading. In this table only, firm-specific wage premiums are normalized by assuming that the mean surplus for workers in both genders working in the hotel and restaurant industry is 0.

Appendix Table B6: Relationship Between Estimated Firm Effects and Mean Total Hours of Workers of Same Gender

	Models for Males				Models for Females			
	No Industry Controls		Industry Controls		No Industry Controls		Industry Controls	
	OLS	IV	OLS	IV	OLS	IV	OLS	IV
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log Mean Hours of Workers at Firm (Same Gender)	-0.16 (0.03)	-0.12 (0.05)	-0.06 (0.03)	0.02 (0.05)	-0.05 (0.03)	-0.13 (0.05)	0.03 (0.02)	0.03 (0.04)
First Stage Coeff.	--	0.54 (0.00)	--	0.45 (0.01)	--	0.65 (0.00)	--	0.60 (0.00)

Notes: Dependent variable in columns 1-4 is estimated firm-specific wage premium for male employees at a firm. Dependent variable in columns 5-8 is estimated firm-specific wage premium for female employees. Entries represent coefficients of log mean hours (including regular hours and overtime) of the gender group at the firm. Models in columns 3-4 and 7-8 include dummies for 20 major industries. All specifications include a constant. Models in even-numbered columns are estimated by IV, using the log mean hours of workers at the same firm in the other gender group as an instrument. Estimated first stage coefficients are reported in second row of the table. All models are fit to micro data for workers in the dual-connected set (n=11,025,257), with standard errors (in parenthese) clustered by firm (n=84,720 firms).

Appendix Table B7: Estimated Relationship Between Estimated Firm Effects and Measures of Surplus per Worker, Controlling for Industry, Location, and Firm Size

	Number Firms	Regressions of Firm Effects on Measure of Surplus		
		Males	Females	Ratio : Col (3) / Col (4)
	(1)	(2)	(3)	(4)
Surplus Measure:				
1. Excess Mean Log Value Added per Worker	47,477	0.140	0.117	0.840
		(0.006)	(0.007)	(0.036)
2. Mean Log Sales per Worker	75,163	0.070	0.061	0.865
		(0.003)	(0.003)	(0.041)
3. Excess Mean Log Sales per Worker	75,163	0.086	0.072	0.837
		(0.004)	(0.005)	(0.043)

Notes: Columns 2-3 report coefficients of surplus measure indicated in row heading in regression models in which the dependent variables are the estimated firm effects for males or females. All specifications include 20 dummies for major industry, dummies for location in Lisbon and Porto, and linear and quadratic controls for mean employment at the firm. Models are estimated at the firm level, weighting by the total number of male and female workers at the firm. Ratios in column 4 are estimated by IV, treating average change in female wages as dependent variable, average change in male wages as endogenous explanatory variable, and change in surplus measure as the instrument. Robust standard errors in parentheses.

Appendix Table B8: Decomposition of Male-Female Wage Gap, Based on Observable Measure of Surplus

	Males (1)	Females (2)	Male-Female Difference (3)
1. Total Firm-specific Component of Wages (from Table 3)	0.148	0.099	0.049
			(21.2)
2. Rent sharing coefficient (from row 1 of Table 6)	0.156	0.137	0.019
3. Mean net surplus (based on excess mean log value added per worker)	0.743	0.566	0.178
4. Firm-specific Component of Wages Attributable to Measured Productivity (= row 2 × row 3)	0.116	0.078	0.038
			(16.5)
5. Share of Total Firm-specific Component Attributable to Measured Productivity (= row 4 ÷ row 1)	0.784	0.785	0.776
Counterfactuals:			
a. Assign females the male firm effects	0.116	0.088	0.028
(sorting effect, using male coefficients)			(11.9)
b. Calculate mean female firm effect using male firm distribution	0.116	0.102	0.014
(bargaining effect, using male distribution)			(6.0)
c. Assign males the female firm effects	0.102	0.078	0.024
(sorting effect, using female coefficients)			(10.4)
d. Calculate mean male firm effect using female firm distribution	0.088	0.078	0.011
(bargaining effect, using female distribution)			(4.6)

Note: decomposition based on regression models presented columns 2 and 3 (row 1) of Table 6. See text. Entries in parentheses in column 3 represent the share of the overall male-female wage gap (0.234) that is explained by the rent sharing component under alternative counterfactuals.

Appendix Table B9: Descriptive Statistics for Job Stayers at Firms with Value Added Data Available for 2006-2009

	All Firms with Male and Female Stayers	
	Males (1)	Females (2)
Mean Age	38.25	37.21
Mean Education	7.99	8.49
Mean Firm Size (workers in QP)	631	1024
Mean Fraction of Females at Firm	0.29	0.59
Mean Log Real Hourly Wage 2006	1.62	1.40
(standard deviation)	(0.48)	(0.44)
Mean Log Real Hourly Wage 2009	1.70	1.48
(standard deviation)	(0.48)	(0.44)
Mean Log Value Added per Worker 2006	0.76	0.59
(standard deviation)	(0.50)	0.49
Mean Log Value Added per Worker 2009	0.76	0.58
(standard deviation)	(0.51)	(0.50)
Number of Workers	283,346	200,907
Number of Firms	33,104	

Note: Sample in columns 1-2 contains workers at dual connected firms with financial data for 2006-2009 who were employed at the firm continuously from 2006 to 2009. Age, education, firm size and fraction female refer to 2006.

Appendix Table B10: Effects of Changes in Excess Log(VA/L) on Wages of Stayers, with Selection Correction

	Estimated Rent Sharing Coefficients:		
	Males (1)	Females (2)	Ratio (3)
Baseline Models (from Row 1 of Table 7):			
Coefficient on Change in Log Value	0.049	0.045	0.911
Added per Worker	(0.007)	(0.008)	(0.086)
Models With Selection Correction:			
Coefficient on Change in Log Value	0.049	0.045	0.911
Added per Worker	(0.007)	(0.008)	(0.086)
Coefficient on Inverse Mills Ratio	-0.002	-0.013	-0.007
(based on fraction of stayers at firm)	(0.006)	(0.005)	(0.003)

Notes: See notes to Table 7. Sample includes job stayers over the period from 2006 to 2009 at firms with data on value added per worker. Dependent variable is average change in wages of male or female workers from 2006 to 2009 at a firm (regression-adjusted for quadratic in age). Inverse Mills ratio term in selection corrected models is constructed using the fraction of all workers at the firm in 2006 who are still employed at the firm in 2009. Standard errors, clustered by firm, in parentheses.

Appendix C: Estimated Models Using Only Workers with High School or More Education Age 25 or Older

In this appendix we re-estimate a series of models, using only workers with high school or more education who are over the age of 25. As shown in the following table, this sample restriction leads to a substantial reduction in the fraction of workers who are ever observed at any time in the 2002-2009 period earning an hourly wage within 5% of the minimum wage:

	Percent Ever Near Min. Wage	
	Males	Females
Dual-connected Sample	7.3	17.7
Higher Education/Age Sample (dual connected)	1.7	3.1

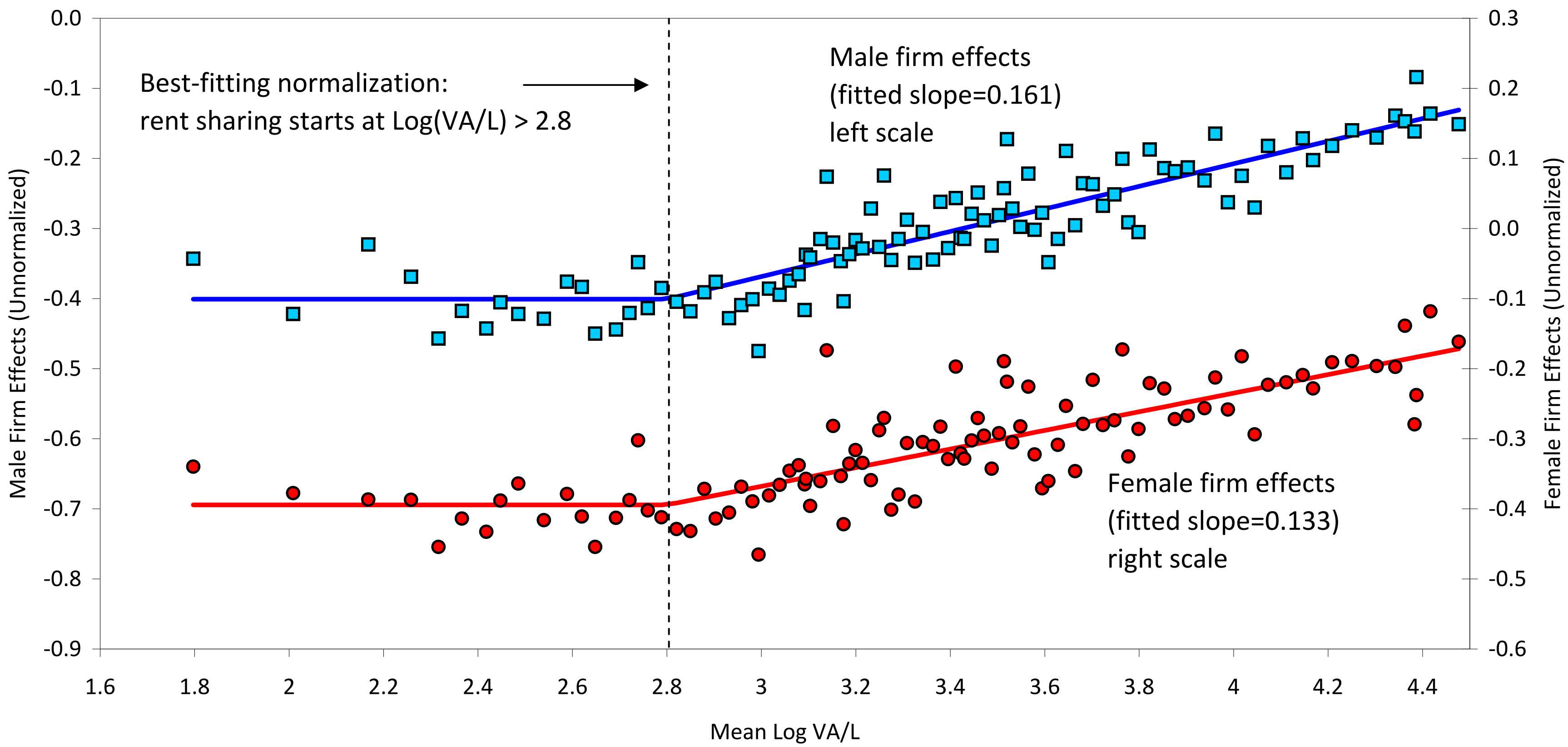
Appendix Table C1 presents estimated two-way fixed effects models for the higher education/age sample. For ease of comparison, the first two columns show the models for our main sample (reproduced from Table 2). Columns 3 and 4 show parallel models estimated on the samples of men and women with higher education and age who are in the corresponding largest connects subsamples.

Appendix Figure C1 graphs the (unnormalized) estimated firm effects for age men and women from the higher education and age sample against mean log value added per worker. This figure shares many features with the corresponding figure for the overall sample (Figure 4): in particular, there is a clear visual "kink" in the relationship between the estimated firm effects and mean log value added per worker. Using the same procedure as we describe in the text for our main sample, we identified the kink point as occurring when mean log value added per worker is above 2.80 (versus the kink at 2.45 indentified in our main sample). We then define "excess value added per worker" as min[0 , VA/L - 2.80] where "VA/L" is mean log value added per worker.

Appendix Table C2 presents decompositions of the importance of firm-specific wage premiums of the gender wage gap, using the same format as Table 3 in the paper, but restricting attention to men and women with higher education and age in the corresponding dual-connected set, and normalizing the firm effects by assuming that firms with VA/L < 2.80 pay a zero wage premium, on average.

Appendix Table C3 presents models that relate the estimated male and female firm effects to VA/L, estimated at the firm level. For reference the first row reproduces the results from our main sample, taken from row 1 of Table 6. The second row presents a parallel set of models using the firm effects estimated for men and women in the higher education and age sample.

Appendix Figure C1: Firm Fixed Effects vs. Mean Log Value Added/Worker
for Workers with High School or More, Age 25+



Note: points shown represent mean estimated firm-specific wage premiums from AKM models for men and women with high school or more education age 25 or older, averaged across firms with value added data available in 100 percentile bins of mean log value added per worker. Firm effects are arbitrarily normalized for estimation of the AKM models.

Appendix Table C1: Comparison of Estimated AKM Models for Males and Females

	All Workers in Connected Sets		Wkrs. High School or More Education, Age ≥25	
	All Males	All Females	Males	Females
	(1)	(2)	(3)	(4)
Standard deviation of log wages	0.554	0.513	0.659	0.583
Number of person-year observations	8,225,752	6,334,039	1,760,208	1,843,182
Summary of Parameter Estimates:				
Number person effects	1,889,366	1,505,517	405,819	451,496
Number firm effects	216,459	185,086	57,054	70,090
Std. dev. of person effects (across person-yr obs.)	0.420	0.400	0.564	0.492
Std. dev. of firm effects (across person-yr obs.)	0.247	0.213	0.279	0.268
Std. dev. of Xb (across person-yr obs.)	0.069	0.059	0.083	0.066
Correlation of person/firm effects	0.167	0.152	0.022	-0.009
RMSE of model	0.143	0.125	0.151	0.140
Adjusted R-squared of model	0.934	0.940	0.947	0.942
Comparison job-match effects model:				
Number of job-match effects	2,689,648	2,087,590	550,229	607,974
RMSE of match-effects model	0.128	0.113	0.135	0.126
Adjusted R-squared of match-effects model	0.946	0.951	0.958	0.953
Std. deviation of job match effect	0.062	0.054	0.068	0.062
Inequality decomposition of two-way fixed effects model:				
Share of variance of log wages due to:				
person effects	57.6	61.0	73.3	71.0
firm effects	19.9	17.2	18.0	21.0
covariance of person and firm effects	11.4	9.9	1.6	-0.7
Xb and associated covariances	6.2	7.5	3.3	4.5
residual	4.9	4.4	3.9	4.2

Notes: See text. Models includes dummies for individual workers and individual firms, year dummies interacted with education dummies, and quadratic and cubic terms in age interacted with education dummies (total of 44 parameters). Comparison job-match effects models include dummies for each worker-firm job match as well as other covariates in basic model. Samples include only

Appendix Table C2: Contribution of Firm-Level Pay Components to Gender Wage Gap --- Workers with High School or More Education, Age 25 or Older

		Means of Firm Premiums:			Decompositions			
					Sorting		Bargaining	
	Wage Gap (1)	Male Prem. Among Men (2)	Female Prem. Among Women (3)	Total Contribution of Firm Components (4)	Using M Effects (5)	Using F Effects (6)	Using M Distribution (7)	Using F Distribution (8)
All	0.272	0.170	0.131	0.038	0.015	0.020	0.018	0.023
				(14.0)	(5.4)	(7.4)	(6.6)	(8.6)
By Age Group:								
Up to age 30	0.115	0.127	0.097	0.030	0.008	0.016	0.014	0.023
				(26.4)	(6.6)	(14.0)	(12.4)	(19.8)
Ages 31-40	0.226	0.174	0.140	0.034	0.011	0.015	0.019	0.023
				(14.9)	(4.6)	(6.6)	(8.2)	(10.2)
Over Age 40	0.313	0.212	0.180	0.032	0.006	0.010	0.021	0.026
				(10.1)	(1.9)	(3.3)	(6.8)	(8.2)
By Education Group:								
High School	0.274	0.152	0.100	0.052	0.031	0.035	0.018	0.022
				(19.1)	(11.2)	(12.7)	(6.5)	(7.9)
University	0.287	0.195	0.170	0.025	0.000	0.007	0.018	0.026
				(8.9)	(0.0)	(2.5)	(6.4)	(9.0)

Notes: see Table 3. Numbers in parentheses represent the percent of the overall male female wage gap (in column 3) that is explained by source described in column heading.

Appendix Table C3: Comparison of Models for Firm Effects Using Base Sample and Workers with High School or More Education, Age 25 or Older

	Number Firms (1)	Regressions of Firm Effects on log(VA/L): Males (2)	Regressions of Firm Effects on log(VA/L): Females (3)	Ratio: Females to Males (4)
1. Base Sample -- Dual Connected Firms	47,477	0.156 (0.006)	0.137 (0.007)	0.879 (0.031)
2. Restricted Sample -- Only Dual Connected Firms, Based on Workers with High School or Higher Education, Age 25 or Older	15,499	0.161 (0.004)	0.133 (0.004)	0.825 (0.023)

Notes: Columns 2-3 report coefficients of excess mean log value-added per worker in regression models in which the dependent variables are the estimated firm effects for the gender group identified in the column headings. Models are estimated at the firm level, weighted by the total number of male and female workers at the firm. Ratio estimates in columns 4 are obtained by IV method -- see text. Standard errors in parentheses.

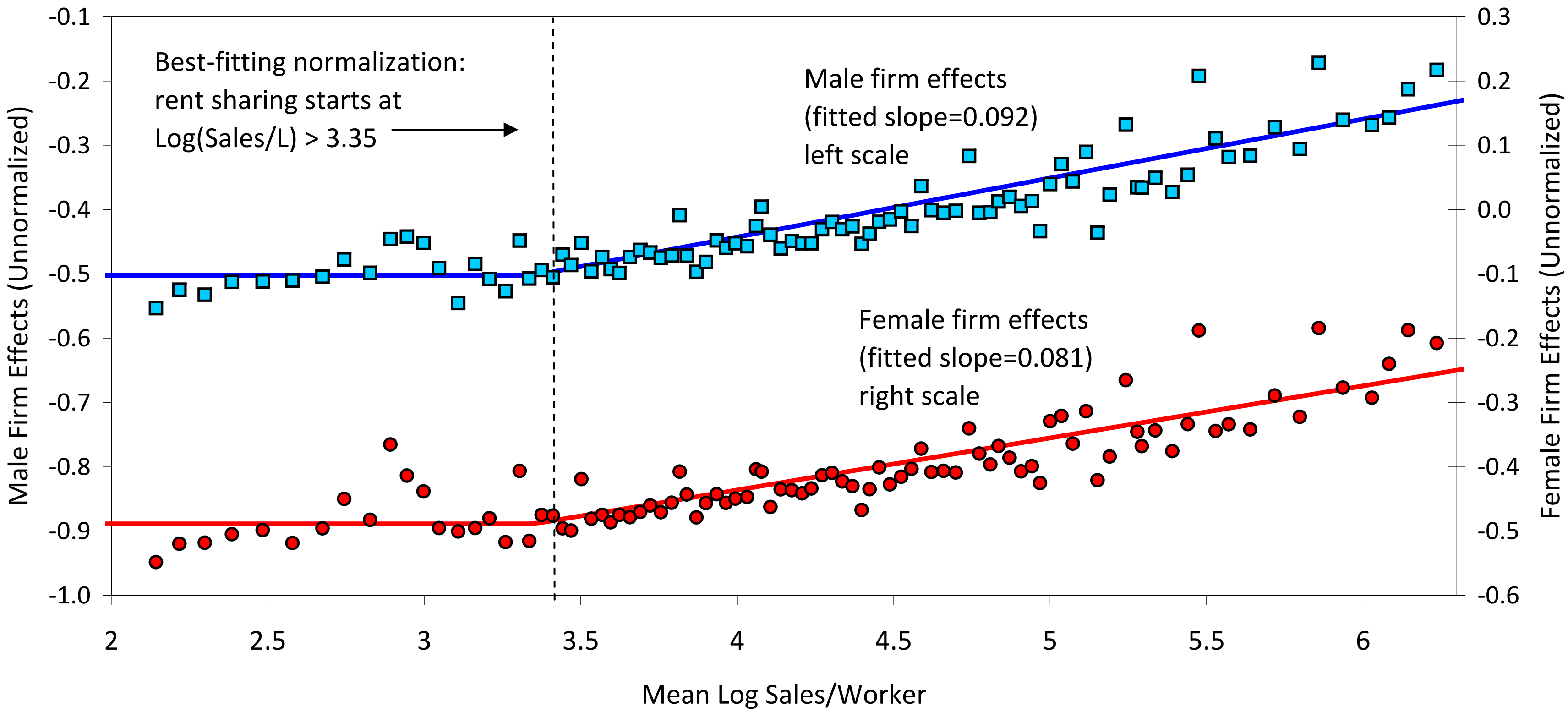
Appendix D: Estimated Models Using Sales per Worker Instead of Value Added per Worker

In this appendix we re-estimate a series of models using information on log sales per worker as an alternative to log value added per worker.

Appendix Figure D1 graphs the (unnormalized) estimated firm effects for men and women against mean log sales per worker. This figure shares many features with the corresponding figure using mean log value added per worker as a measure of the surplus available for bargaining (Figure 4). In particular, there is a clear visual "kink" in the relationship between the estimated firm effects and mean log sales per worker. Using the same procedure as we describe in the text for our main sample, we identified the kink point as occurring when mean log sales per worker is above 3.35 (versus the kink at 2.45 in the relationship with mean log value added per worker indentified in our main sample). We then define "excess sales per worker" as min[0 , S/L - 3.35] where "S/L" is mean log sales per worker.

Appendix Table D1 presents decompositions of the importance of firm-specific wage premiums of the gender wage gap, using the same format as Table 3 in the paper, but normalizing the firm effects by assuming that firms with S/L < 3.35 pay a zero wage premium, on average.

Appendix Figure D1: Firm Fixed Effects vs. Mean Log Sales/Worker



Note: points shown represent mean estimated firm-specific wage premiums from AKM models for men and women, averaged across firms with value added data available in 100 percentile bins of mean log sales per worker. Firm effects are arbitrarily normalized for estimation of the AKM models. Data for one percentile group with apparent outlier in sales data not shown.

Appendix Table D1: Contribution of Firm-Level Pay Components to Gender Wage Gap -- Sales/Worker Normalization

	Wage Gap	Means of Firm Premiums: Male Prem. Among Men	Means of Firm Premiums: Female Prem. Among Women	Total Contribution of Firm Components	Decompositions: Sorting: Using M Effects	Decompositions: Sorting: Using F Effects	Decompositions: Bargaining: Using M Distribution	Decompositions: Bargaining: Using F Distribution
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
All	0.234	0.110	0.059	0.052	0.035	0.047	0.005	0.017
				(22.2)	(14.9)	(19.9)	(2.3)	(7.3)
By Age Group:								
Up to age 30	0.099	0.077	0.046	0.030	0.019	0.029	0.001	0.012
				(30.6)	(18.9)	(29.3)	(1.3)	(11.7)
Ages 31-40	0.228	0.118	0.071	0.047	0.029	0.041	0.007	0.019
				(20.8)	(12.7)	(17.8)	(3.0)	(8.1)
Over Age 40	0.336	0.131	0.059	0.072	0.050	0.064	0.007	0.021
				(21.3)	(15.0)	(19.1)	(2.2)	(6.3)
By Education Group:								
< High School	0.286	0.077	0.015	0.062	0.045	0.061	0.001	0.017
				(21.7)	(15.6)	(21.4)	(0.3)	(6.1)
High School	0.262	0.160	0.097	0.064	0.051	0.051	0.013	0.012
				(24.2)	(19.6)	(19.5)	(4.8)	(4.7)
University	0.291	0.221	0.172	0.049	0.025	0.029	0.020	0.024
				(17.0)	(8.7)	(9.9)	(7.0)	(8.3)

Notes: see Table 3. Numbers in parentheses represent the percent of the overall male female wage gap (in column 3) that is explained by source described in column heading. In this table only, firm-specific wage premiums are normalized by assuming that firms with mean log sales per worker less than 3.35 pay have on average no surplus.