Financial risk-taking and the gender wage gap

Per-Anders Edin
Håkan Selin

IFAU | INSTITUTE FOR EVALUATION OF LABOUR MARKET AND EDUCATION POLICY

The Institute for Evaluation of Labour Market and Education Policy (IFAU) is a research institute under the Swedish Ministry of Employment, situated in Uppsala.

IFAU's objective is to promote, support and carry out scientific evaluations. The assignment includes: the effects of labour market and educational policies, studies of the functioning of the labour market and the labour market effects of social insurance policies. IFAU shall also disseminate its results so that they become accessible to different interested parties in Sweden and abroad.

Papers published in the Working Paper Series should, according to the IFAU policy, have been discussed at seminars held at IFAU and at least one other academic forum, and have been read by one external and one internal referee. They need not, however, have undergone the standard scrutiny for publication in a scientific journal. The purpose of the Working Paper Series is to provide a factual basis for public policy and the public policy discussion.

More information about IFAU and the institute's publications can be found on the website www.ifau.se

Financial risk-taking and
the gender wage gap[a]

Per-Anders Edin[b] and Håkan Selin[c]

September 7, 2020

Abstract

Could differences in risk attitudes explain parts of the gender wage gap? We present estimates on the association between labor market outcomes and financial risk-taking using individual level administrative data on individual wealth portfolios and wage rates. The individual's share of risky to total financial assets is significantly and positively associated with the wage rate. However, it turns out that our risk measure explains only a small part of the observed gender difference in wages.

Keywords: Wages; human capital; financial risk.

JEL Classification: J31; D8.

[b] Department of Economics at Uppsala University; Institute for Evaluation of Labour Market and Education Policy (IFAU); Uppsala Center for Labor Studies (UCLS), and STIAS. Email: Per-Anders.Edin@nek.uu.se

[c] IFAU and UCLS. Email: hakan.selin@ifau.uu.se

Contents

1 Introduction

Financial economists teach that risk attitudes determine investments in risky financial assets, and empirically we know that women on average chose safer financial portfolios (Croson and Gneezy, 2009). Labor economists teach that risk attitudes shape labor market outcomes via human capital investments, occupational choices, and reservation wages, and empirically we know that women on average earn lower wages than men. Moreover, survey and experimental evidence does not only show that women are more risk averse than men, but also that there is a substantial correlation in risk attitudes across different domains (Dohmen et al., 2011). These basic facts and intuitions give rise to a fundamental but largely unexplored research question: Could differences in risk attitudes explain parts of the gender wage gap?[1]

Historically, the literature on the gender wage gap focused on productivity differences and discrimination (Altonji and Blank, 1999). In recent years, a new literature has emerged, which discusses preference-based determinants of the gender gap, such as attitudes towards risk and competition, mainly using laboratory experiments. However, survey articles by Azmat and Petrongolo (2014) and Bertrand (2011) emphasize that non-lab evidence in this area is scarce. In particular, there are no previous studies relating gender differences in financial risk taking to gender differences in labor market outcomes.[2]

Using micro level administrative data on wealth portfolios and monthly wages, we examine if a high level of financial risk-taking is associated with a high wage, and we analyze the implications for the estimated gender gap. Crucially, the household's wealth portfolio can be decomposed into *individual* portfolios. This implies that we are able to relate individual portfolios to individual full-time equivalent wages. To the best of our knowledge, wage equations including an administrative measure of the risky asset share, which we define as

---

[1]Of course, it is possible to pose the reverse question, i.e. how gender differences in wages impact on financial decisions, see Thörnqvist and Vardardottir (2015). In this paper we do not attempt to isolate a causal parameter. Our aim is instead to explore if the gender wage gap shrinks when accounting for financial risk taking.

[2]Using self-reported survey measures, a few articles have examined the impact of risk aversion on the gender wage gap, see Le et al. (2011), Jung (2017) and Cho (2011). These studies find that risk attitudes explain only a modest fraction of the gender wage gap.

the ratio of stock holdings to total financial wealth, have never been presented. Broadly related studies either used survey measures of risk attitudes (Bonin et al., 2007) or survey measures of household wealth (Shaw, 1996). Our main specification, featuring the log of the wage rate on the left hand side, and the risky share on the right hand side, is derived from a model in which individuals make simultaneous investments in financial and human capital. In the context of such a life cycle model, we hypothesize that more risk averse individuals not only invest less in risky financial assets, but also less in risky human capital, and on average earn lower wages.

In line with our expectations, we find that women in our data hold significantly less stocks than men do. We also find that the risky share is indeed positively and significantly associated with the wage rate, and this finding is robust. But the implications for the estimated gender wage gap are modest. In our most naive model, where we regress the log wage on the raw risky share without adjusting for differences in wealth, we find that the risky share explains 7.5% of the observed gender difference in wages. However, we believe that financial risk taking explain considerably less of the observed gender gap. When adjusting the risky share for differences in wealth, the gender wage gap drops by only 3.1% when financial risk-taking is accounted for. Since financial risk taking is determined by more personal traits than risk aversion, we believe that the influence of risk aversion *per se* on the gender gap is even smaller.

Our data allow us to describe heterogeneity in the risky share measure along various dimensions. We pay special attention to associations at the occupational level in the highly gender-segregated Swedish labor market. Both the occupation-specific average of the log wage and its standard deviation are highly correlated to our risky share measure (after partialling out financial wealth). Moreover, we find an interesting asymmetry between male and female dominated occupations. While financial risk taking is low in many male dominated professions, there are literally no heavily female dominated occupations in which average financial risk taking is high.

The paper is structured in the following way. In Section 2 we provide an overview of our data sources, and Section 3 contains a descriptive analysis of the raw wealth data, with an emphasis on gender differences. We construct our

preferred measure of the risky share in Section 4, and thereafter we describe risk taking in different subgroups using this measure. In particular, we highlight the occupational dimension. In Section 5 we sketch a theoretical framework for stochastic returns to human capital investments, which rationalizes the empirical model. Regression results from wage equations including measures of financial risk taking are reported in Section 6. Finally, Section 7 concludes the paper.

2 Data

Owing to its high level of detail and general coverage, Swedish wealth data have been used extensively in economic research in recent years. In this section we demonstrate how we combine wealth data with other data sources, and we show how we select our estimation sample. The reader should keep in mind that we later on will focus on how the wage rate covaries with financial risk taking.

2.1 LINDA

Data come from LINDA (Longitudinal INdividual DAta for Sweden), which is a representative sample of 3.35% of the Swedish population (Edin and Fredriksson, 2000). It builds on information from administrative registers, and it consists of a large panel of individuals and their household members. An essential component of LINDA is the income register (Inkomst- och Förmögenhetsstatistiken), which originates from filed tax reports. Supplementary information on both labor and non-labor incomes are available from third-party reported income statements (kontrolluppgifter) from e.g. employers and commercial banks. In addition, LINDA contains detailed information on the level and field of education etc. Crucially, we access register information on the partner (and children in the household) if the sampled individual is either married or cohabiting with children in common. We primarily use data from year 2000, but we have longitudinal data on a subset of variables for a much longer time period (e.g. annual taxable earnings from 1968 and onwards). We cannot use data from recent

5

years, because administrative wealth data only exist for the years 1999-2007.[3]

2.2 The wage register

Data on wage rates in LINDA come from two sources, namely the wage structure statistics and a supplementary survey. First, the wage structure statistics (Lönestrukturstatistiken), which is administrated by the National Mediation Office, come from surveys to employers in both the public and private sectors. It includes data on monthly wages and salaries (in full time equivalents including taxable fringe benefits), occupation (3-digit code), and information on whether the individual worked part time or full time. All public sector workers are included in the survey, whereas 50 % of all private sector workers are covered.[4] Wage information is available for individuals aged 18-65.

Second, in order to remedy the incomplete coverage of private sector workers, Statistics Sweden conducts a supplementary survey to employers with employees who are in the LINDA sample, but not in the wage structure statistics. Information on the same set of variables is collected. Hence, the coverage is better in the LINDA sample than in the full population wage structure register. Still, public sector workers and employees in big private firms are overrepresented in the "wage sample", and we will pay careful attention to this issue in the analysis.

2.3 The wealth register

LINDA data have been linked to a unique Swedish wealth register, which was in place 1999-2007. Since the wealth register is a full population register, detailed wealth information is available for all individuals in the LINDA sample,

---

[3]Wealth register files are linked to the LINDA sample for the years 1999-2002. Due to some concerns regarding quality we do not use the 1999 file, but we will use the files for 2000 to 2002, with a strong emphasis on 2000.

[4]In the private sector, the wage structure statistics is based on a stratified sample with respect to industry and firm size. The probability to be in the sample increases with firm size. Firms with more than 500 employees are always in the survey and 76 % of firms with 200-499 employees are sampled. If an individual is employed in a small private firm, the probability to appear in the wage structure statistics is substantially smaller; less than 3 % of firms with less than 10 employees are sampled.

including household members. The data, which mainly are based on income statements from financial institutions, include rich information on various real and financial assets and liabilities used to assess the wealth tax.[5] Real assets, such as houses and apartments, are reported according to their estimated market values, whereas information on financial assets reflect their market value as of December 31.[6] When classifying the wealth data we use the classification suggested by Flood (2004, Appendix A).

Some limitations of the wealth data deserve to be mentioned. First, we do not observe unquoted shares nor assets in the funded component of the Swedish pension system as these were exempt from the wealth tax.[7] Second, there is probably underreporting of assets such as boats and objects of art for which third-party reporting was absent. Third, commercial banks were not required to report bank accounts balances to the Swedish Tax Agency unless the owner of the account received an interest of more than 100 SEK. Therefore, a substantial share (56 %) of individuals aged 31-60 lack bank holdings in our data. Still, as pointed out by e.g. Black et al. (2015), virtually everyone in the Swedish adult population owns a bank account. In the main analysis we impute bank holdings by setting imputed bank holdings equal to the maximum of observed bank holdings and SEK 10,000. We perform several robustness checks; we change the cut-off and we also implement the regression based imputation method of Calvet et al. (2007).

2.4   Sample selection and summary statistics

We limit our sample to men and women aged 31-60, who are sampled in LINDA. At age 31 most people have completed their educational degrees, which is cru-

---

[5]The wealth tax was abolished in 2007. For that reason no comprehensive wealth data are available after 2007.

[6]House property is valued based on tax assessed values and information on local house prices. Co-operative apartments "bostadsrätter" are valued based on the monthly fee to the housing co-operative and sales statistics.

[7]On data from the Swedish pension authorities, Säve-Söderbergh (2012) has analyzed investments in the mandatory individual accounts from the perspective of risk and gender. While gender differences are small between men and women who choose less risky portfolios in the mandatory retirement accounts, there are large gender differences among those who choose risky portfolios.

cial given our research question. The upper age limit is motivated by retirement behavior, which often has consequences for the wealth portfolio. In 2000 the average retirement age was around 62 for both men and women in Sweden (Sundén, 2006, Figure 2).[8]

In Section 5 we will specify an empirical wage equation, and the most central component of our study will be to regress wages on a risk measure. This analysis will be conducted on the subsample of individuals with monthly wages, which we will refer to as the "wage sample". As this sample is selected in a non-random way we, of course, compare the wage sample with the randomly selected total LINDA sample of individuals aged 31-60. Table A1 of Appendix A.1 reports the means of observable characteristics in the total sample and in the wage sample for men and women, respectively. There is a larger share of women in the wage sample as compared to the entire sample. This is due to the fact that women are overrepresented among public sector employees, which are all sampled in the wage register. In percentage terms the raw earnings gap is larger in the complete sample than in the wage sample. More women in the entire sample report very low earnings. In the wage sample the mean female wage is 79 % of the mean male wage. A very large fraction of men work full time, whereas part time work is more prevalent among women. The gender differences in real assets, financial assets, debt, and net wealth are substantial in both samples, but larger in the complete sample than in the wage sample. Somewhat more individuals in the wage sample hold stocks, but the ratios of stocks to financial assets are fairly similar in the two samples.

3   The wealth portfolio

In this section we describe features of wealth portfolios with a special emphasis on gender differences. In Figure 1 we show box plots and report descriptive statistics for (a) real assets, (b) financial assets, (c) debt and (d) net wealth for individuals aged 31-60 in the LINDA sample. From (a) we see that more men

---

[8]Graphs showing some key wealth variables for different age and gender groups are provided in Appendix A.3. As the graphs are snapshots of the wealth distribution in 2000 the wealth-age correlation captures both cohort effects and life cycle (age) effects.

than women own real assets (e.g. houses and co-operative apartments) and that mean real assets (conditional on owning real assets) are larger for men than for women. From (b) we infer that mean financial assets is substantially larger for men than women. When interpreting Figure 1b one should remember that we measure bank holdings imperfectly and that we imputed bank account balances for a majority of the sampled individuals. Therefore, everyone has positive financial assets and median values are quite similar for men and women. By contrast, in the raw data 31% lack financial wealth entirely. There are also gender differences in the propensity to borrow and the level of debt. Figure 1c reveals that men borrow more money than women. Figure 1d, finally, shows the distribution of net wealth, which is obtained by subtracting debt from the sum of real and financial assets. A large fraction of the sample, 31%, has negative net wealth (in raw data). On average, men own larger net wealth than women, but the dispersion is larger for men as well.

In Figure 2 we zoom in at the composition of financial asset portfolios by gender. For each gender separately, the figure shows the aggregate value of an asset divided by aggregate financial wealth. The most striking discrepancy between men and women is that individual stocks make up a substantial larger fraction of men's total financial assets (34.4%) than for women (23.9%). On the other hand, in relative terms, women invest more in mutual funds, 23.2% for men vs. 29.5% for women. Hence, when summing individual stocks and mutual funds the gender difference is less pronounced.[9]

---

[9]The null hypothesis that average stocks and mutual funds (divided by gender-specific average financial wealth) is the same in the male and female samples is not rejected in a two sample t-test; the p-value is 0.11. This statistic reflects shares in the aggregate financial portfolios of both genders. If one instead examines differences in the average risky shares, there are statistically significant differences across genders not only when individual stocks are in the numerator, but also when both stocks and mutual funds are in the numerator.
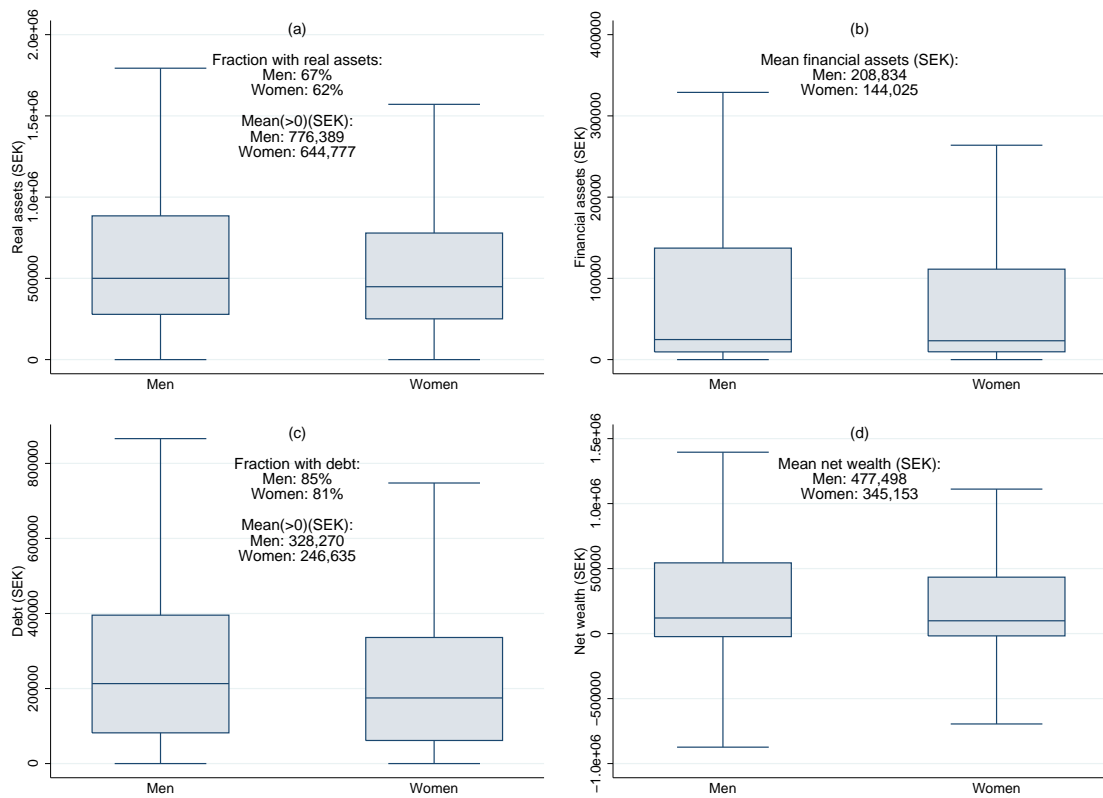
Figure 1: Real assets, financial assets, debt and net wealth by gender. Note that everyone in the sample has positive financial assets due to the imputation procedure described in Section 2.3. Gray-shaded "boxes" represent interquartile ranges (IQR), and the median is indicated with a line inside the box. "Whiskers" span data within 1.5 of the IQR of the nearer quartile. The sample includes all individuals aged 31-60 who were sampled in LINDA in year 2000.

Figure 2: Components of financial assets as shares (in %) of total financial assets, by gender.

4 Financial risk-taking in the population: a description

In this section we first discuss the risky asset share, and we propose a measure that is independent of financial wealth. We then examine subgroup differences in this risky share measure. Finally, we visualize the correlation between the risky share measure and some quantities of interest at the occupational level.

4.1 The risky share and its relation to financial wealth

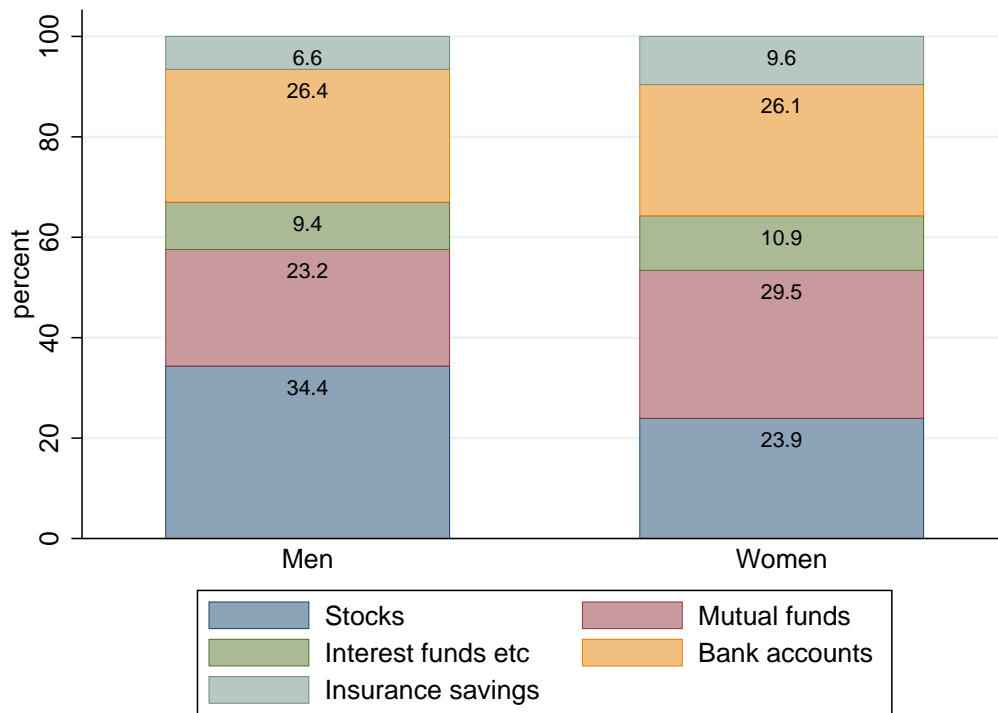In the standard continuous-time model featuring risky and safe financial assets and CRRA preferences, the risky portfolio share is independent of both time and total wealth (Merton, 1971). How does the risky portfolio share correlate with wealth and age in our data? To answer that question we first need to define the risky share. In the baseline, we define the risky share as the ratio of individual stocks to total financial wealth, while excluding mutual funds from the numerator. While investors holding a lot of mutual funds tend to mainly be subject to market risk, investors of individual stocks in addition tend to expose themselves to the idiosyncratic risk pertaining to specific individual stocks.[10] Therefore, we consider stock investments as our preferred measure of financial risk-taking.

In Figure 3 we plot the risky share as a function of log financial wealth, separately for each gender. We graph the relationship of interest by first creating 20 equally sized bins with respect to financial wealth, while pooling men and women. The dots represent gender-specific averages of the risk measure in each bin.[11] Figure 3 suggests that the risky asset share in general increases in wealth.

---

[10]Calvet et al. (2007) analyzed disaggregated Swedish financial data and ranked household portfolios based on idiosyncratic risk. They concluded that "households with low idiosyncratic risk often hold concentrated portfolios of mutual funds, whereas households with high idiosyncratic risk hold concentrated portfolios of individual stocks."(p.724)

[11]Since we define ventile groups based on the pooled sample, the gender composition varies across these groups. The top wealth groups are male dominated, but men are in fact also slightly over represented at the bottom of the distribution. Among the large number of individuals with an imputed financial wealth of SEK 10,000 (the bottom six ventile groups), the gender composition is quite even.

There is a cluster of dots at the very bottom at the wealth distribution. It is generated by the large number of observations for which bank holdings have been imputed to be SEK 10,000. Accordingly, at low levels of wealth the slope is very steep due to measurement problems, at upper-middle parts of the distribution it flattens out, and then the slope increases again at high levels of wealth. The patterns are fairly similar for men and women, but there are striking differences in levels, especially at the middle of the distribution. We deliberately omit confidence intervals; the gender difference in the risky share is indeed higly significant in all groups (the p-value is always below 0.001).[12]

Figure 3 suggests that the risky share is positively correlated with wealth. However, we do not want our regression results and descriptive analysis to be influenced by this correlation. Therefore, we construct a measure of the risky share, which does not depend on financial assets. More specifically, we partial out the influence of financial wealth by regressing the risky share on 100 percentile dummies, using observations with imputed financial assets of SEK 10,000 as a reference category.[13] We do this on the total sample of 62,717 men and 60,026 women aged 31-60. The obtained residuals reflect variation *within* a given a wealth percentile group.

4.2 The risky share in different subgroups

How does financial risk taking differ across demographic groups once conditioning on financial wealth? In Table 1 we split the LINDA sample (individuals aged 31-60) into two subsamples along five dimensions: age, having a partner, having children, holding a university degree, and being employed in the public sector. In each subsample we compare average risky share residuals for men and women. By construction, the residuals sum to zero for the total sample.

---

[12]If we instead define the risky share with stocks *plus* mutual funds in the numerator, the gender difference becomes much smaller. Still, the basic pattern is similar: At low levels of wealth the slope is very steep due to measurement problems, at upper-middle parts of the distribution it flattens out, and then the slope increases at high wealth levels. The within-group gender difference is typically significant at the 5 percent level at the middle of the distribution, but insignifant at the top.

[13]In the reference group, which consists of 41,009 individuals, the risky share is zero by construction. Therefore, there will be a mass point in the distribution of residuals at zero.
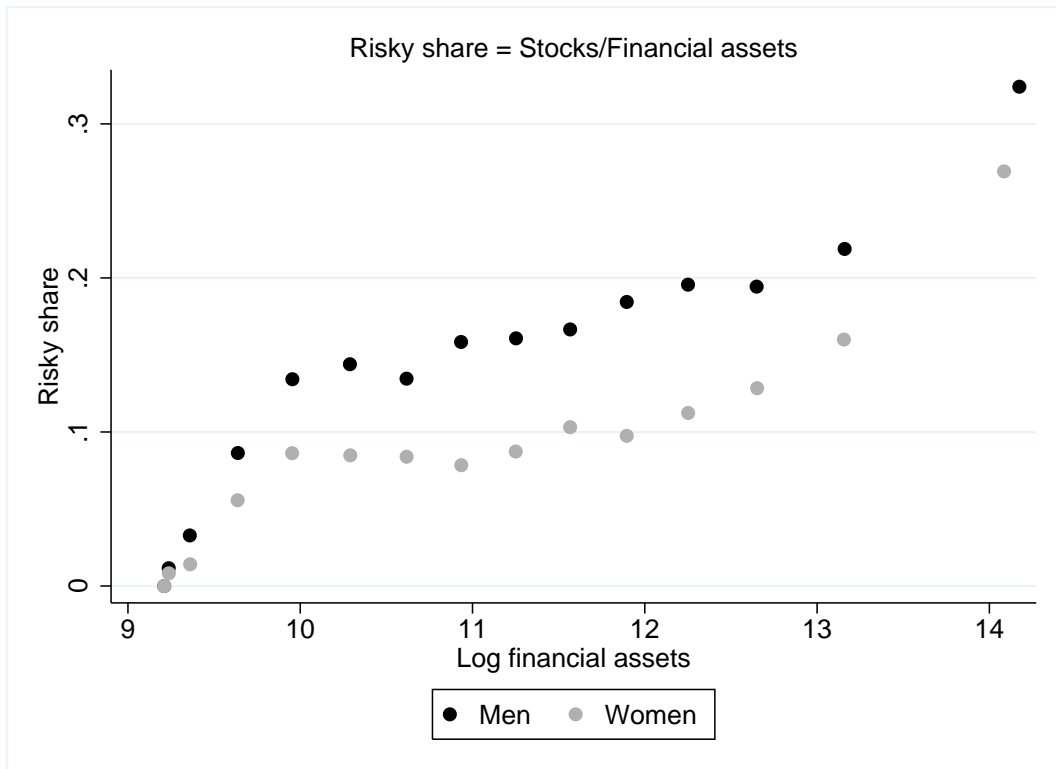
Figure 3: Stocks/financial assets as a function of log financial assets. The total sample (men and women pooled) is divided into 20 equally sized bins with respect to financial assets. The dots represent averages in each bin. 62,717 men and 60,026 women.

Accordingly, the sign of a group average shows whether or not the group is above or below the total sample mean.

The most striking feature of Table 1 comes as no surprise in light of our descriptive analysis of Section 3: Men take more financial risk than women. This holds true in all subsamples, and in the aggregate the difference is around 3.8 percentage points. The gender difference is, however, particularly salient among those with a university degree (around 5.6 percentage points).

From Table 1, column 1, we learn that the difference in risk-taking between young (aged 31-45) and old (46-60) is small. It is also quite modest for individuals in couples and "singles"[14], and the same holds for individuals with and without children in the household. The two bottom categories, education and public sector employment, have a stronger connection to human capital investments and educational choices, and there we see larger overall differences in risk-taking. A common perception is that public sector employees are better protected than private sector workers, but lower paid. From Table 1, column 1, we infer that public sector employees have a significantly lower risky share than others. This association is not contradicting an hypothesis that more risk averse individuals sort into public sector employment. In the next subsection we will describe the occupational dimension in greater detail.

4.3   Occupations

Our wage data contain 3-digit occuapation codes (according to the ISCO-88 classficiation). Hence, we are able to examine how our variables of interest behave at the occupation level. This is interesting, because measures of within-occupation wage dispersion have earlier been used by researchers as proxies for earnings risk in occupations.[15] However, to the best of our knowledge, no one has described how such measures correlate with financial risk taking at the occupational level.

In Figure 4 we collapse the data into occupation-specific averages. In (a) we

---

[14]In the Swedish registry data a couple is observed if the two partners are married or cohabiting with common children. Hence, "singles" do not only include genuine singles, but also partners who are cohabiting without common children.

[15]See Hartog (2011) for an overview of the literature on risk-augmented Mincer equations.

Table 1: Average risky share (residuals)

	All	Men	Women
	(1)	(2)	(3)
Young (aged 31-45)	0.003	0.020	-0.014
	(0.0007)	(0.0011)	(0.0008)
Old (aged 46-60)	-0.003	0.017	-0.025
	(0.0008)	(0.0012)	(0.0010)
In a couple	0.005	0.027	-0.017
	(0.0007)	(0.0011)	(0.0008)
"Singles"	-0.008	0.006	-0.023
	(0.0008)	(0.0012)	(0.0010)
With children	0.006	0.026	-0.012
	(0.0007)	(0.0011)	(0.0008)
Without children	-0.008	0.011	-0.031
	(0.0008)	(0.0011)	(0.0010)
University degree	0.015	0.045	-0.011
	(0.0011)	(0.0018)	(0.0014)
No university degree	-0.007	0.008	-0.023
	(0.0006)	(0.0009)	(0.0007)
Public sector	-0.017	0.013	-0.028
	(0.0009)	(0.0020)	(0.0009)
Not in public sector	0.007	0.020	-0.013
	(0.0006)	(0.0009)	(0.0009)
All	0.000	0.019	-0.019
	(0.0005)	(0.0008)	(0.0006)

Note: Standard error of the mean is in paranthesis. Each pair, e.g. young and old, is defined based on the entire LINDA sample of individuals aged 31-60, which contains 122,743 observations (62,717 men and 60,026 women). "Couples" refer to formally married individuals or to cohabiting individuals with common children. "Singles" refers to genuine singles and cohabiting without common children. "Children" means "children in the household".

graph the average log monthly wage as a function of the average risky share residual. There is a strong, approximately linear, relationship between the average log wage and the risky share residual. Employees in high-wage occupatations tend to hold risky portfolios, and vice versa. In (b) we instead plot the intra-occupation standard deviation in log wages as a function of the average risky share. The standard deviation in log wages (or in log wage residuals) across different groups has earlier been used as measure of cross sectional labor market risk, see Christiansen et al. (2007). We see that there is a substantial correlation also in this dimension. One should keep in mind, however, that Figure 4(a) poorly corresponds to the life cycle model we will outline in Section 5, in which occupational and educational choices are themselves outcomes.

In Figure 5 we plot the share of females in each occupation against the average risky share residual. Interestingly, there is huge dispersion in the gender shares. In some occupations the female share is close to zero, while it is almost 100 % in others. This reflects the strong gender segregation in the Swedish labor market. In comparison with Figure 4 it is much more difficult to discern a distinct linear function in Figure 5. Occuations with a large share of male employees exibit both low and high average risky shares. Still, it is remarkable that the upper right corner of the figure is empty: There are *no* heavily female dominated occcupations where employees on average hold very risky financial portfolios!

In Table 2 we sort occupations by the average risky share residual, and we report summary statistics for the "top 10" and "bottom 10". In the very top we find "Directors and chief executives", who also earn the highest wages, but few executives are women (only 7%). In the top 10 group two occupations have an even gender distribution (Business professionals and Health professionals), whereas the others are strongly dominated by men. By contrast, in the bottom 10 group there are no occcupations with a relatively even gender distribution. Seven occupations are female and three are male dominated. As can be seen from Figure 4(a), these occupations are typically to be found at the bottom of the wage distribution (e.g. cleaners).
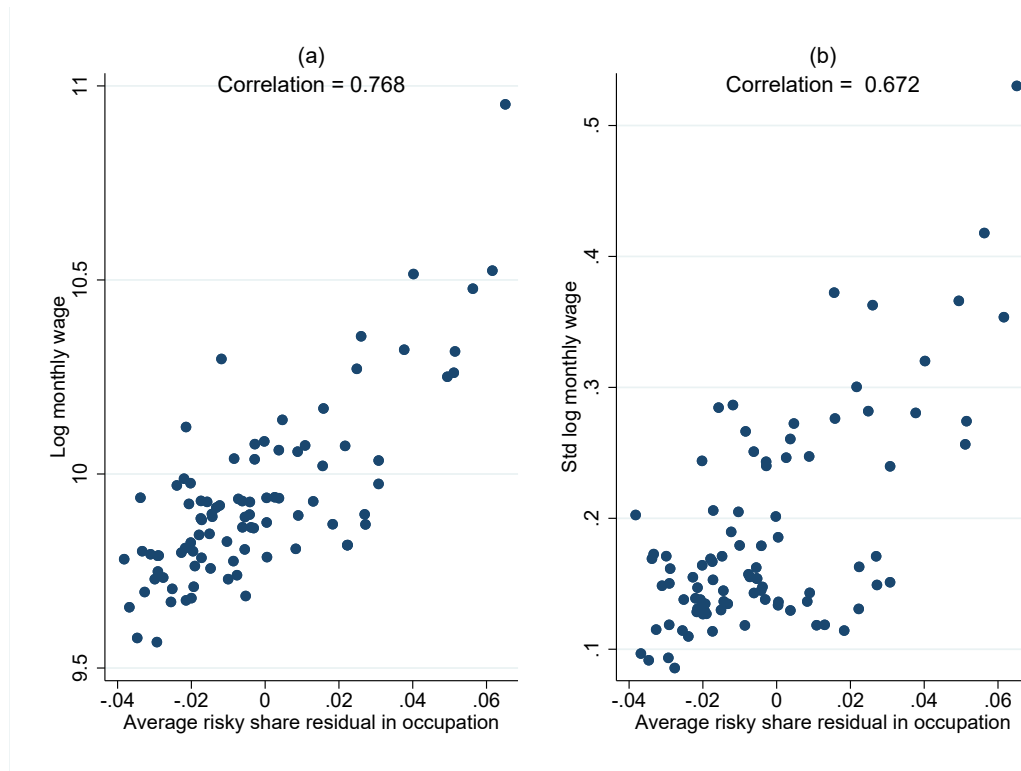
17

Figure 4: Mean and standard deviation in the log wage as functions of the average risky share residual at the occupation level. Occupations with less than 50 observations are excluded.
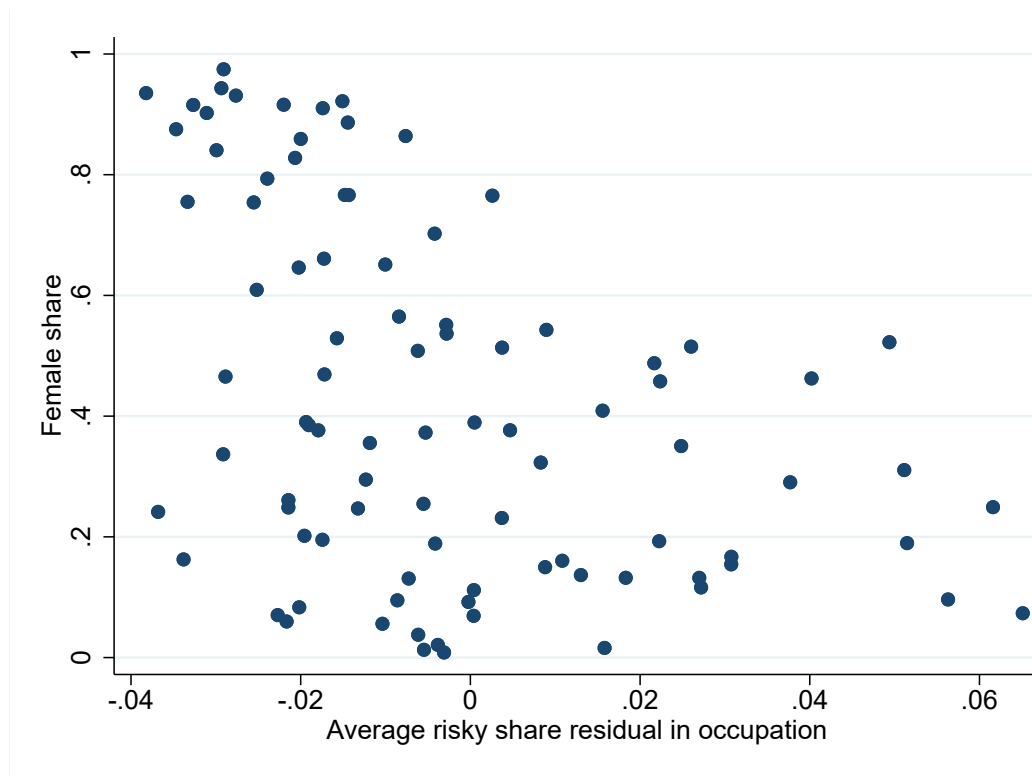
Figure 5: Female share as function of the average risky share residual at the occupation level. Occupations with less than 50 observations are excluded.

Table 2: Occupations sorted by the average risky share residual

	Risky share (res)		Log wage		Female	Obs	Occupation (ISCO88 code)
	Mean	Std	Mean	Std	share		
Top 10							
1.	0.065	0.273	10.953	0.530	0.07	177	Directors and chief executives (121)
2.	0.062	0.255	10.524	0.354	0.25	955	Other department managers (123)
3.	0.056	0.263	10.478	0.418	0.10	135	Ship and aircraft controllers and technicians (314)
4.	0.051	0.252	10.316	0.274	0.19	812	Architects. engineers and related professionals (214)
5.	0.051	0.264	10.261	0.257	0.31	132	Physicists. chemists and related professionals (211)
6.	0.049	0.252	10.251	0.366	0.52	1290	Business professionals (241)
7.	0.040	0.240	10.515	0.320	0.46	839	Health professionals (except nursing) (222)
8.	0.038	0.231	10.320	0.281	0.29	940	Computing professionals (213)
9.	0.031	0.221	10.035	0.240	0.17	2180	Physical and engineering science technicians (311)
10.	0.031	0.188	9.974	0.151	0.15	123	Chemical-processing-plant operators (815)
Bottom 10							
1.	-0.038	0.155	9.781	0.203	0.94	247	Cashiers. tellers and related clerks (421)
2.	-0.037	0.138	9.657	0.097	0.24	174	Market gardeners and crop growers (611)
3.	-0.035	0.098	9.577	0.092	0.88	1035	Shoe cleaning and other street services elementary occupations (912)
4.	-0.034	0.138	9.939	0.169	0.16	166	Printing-. binding- and paper-products machine operators (825)
5.	-0.033	0.176	9.801	0.173	0.76	543	Customs. tax and related government associate professionals (344)
6.	-0.033	0.124	9.696	0.115	0.92	8773	Personal care and related workers (513)
7.	-0.031	0.161	9.793	0.149	0.90	696	Numerical clerks (412)
8.	-0.030	0.158	9.729	0.171	0.84	1273	Other office clerks (419)
9.	-0.029	0.134	9.567	0.093	0.94	653	Domestic and related helpers. cleaners and launderers (913)
10.	-0.029	0.116	9.789	0.119	0.34	190	Rubber- and plastic-products machine operators (823)

Note: Occupations with less than 50 observations are excluded.

5 Motivation of empirical specification

In this section we motivate the empirical specification. In particular, we derive our main empirical equation from a model of investments in human capital and financial capital.

5.1 Risk and human capital investments

Most of the literature on human capital assumes that investments are made under certainty. But in real life the returns to investments in human capital are far from certain. In practice, it is impossible for a student to perfectly forecast the marginal return to an additional year of schooling. Along the same lines, it is impossible for workers to know the returns to investments in new skills with certainty. Levhari and Weiss (1974) distinguished between two sources of uncertainty associated with human capital investments:

- *Uncertain inputs*: The individual has imperfect knowledge of (i) own ability and/or (ii) the quality of schooling or on the job training.

- *Uncertain outputs*: The individual has imperfect knowledge of future demand and supply conditions. Hence, the market value of an investment is uncertain.

According to the mean-variance model, a workhorse model for analyzing financial investments, investors trade-off the expected return of an investment against its variance. The more risky an investment is, the higher is the required expected rate of return *ceteris paribus*. A similar logic applies to risky human capital investments.[16] This is, however, with some qualifications, because human capital assets differ from financial assets in some important respects. In particular, human capital assets cannot be bought and sold on the market.

5.2 Basic model structure

Following Williams (1979) and Shaw (1996) we consider a life-cycle model in which individuals simultaneously make investments in human capital and fi-

---

[16]See the discussion in Christiansen et al. (2007).

nancial capital under uncertainty. Heterogeneity in risk aversion plays a key role, and it is used to identify the empirical relationship of interest. Individuals invest in three distinct assets: one risky financial asset with expected return μ_f and variance σ_f^2, one safe financial asset with risk-free return $r \leq \mu_f$ and, finally, one human capital asset.[17] In discrete time, individual i's labor income at age t is

$$y_{it} = (1 - s_{it} - l_{it})k_{it}, \tag{1}$$

where s is time devoted to acquisition of human capital, l is leisure and k is the stock of human capital (the wage rate). New human capital in the beginning of period $t + \Delta t$ is $\theta(t, t + \Delta)s_{it}k_{it}$, where

$$\theta(t, t + \Delta) \sim \log N(\mu_\theta, \sigma_\theta) \tag{2}$$

is the stochastic return to human capital investments.

Shaw (1996) used the models of Williams (1979) and Merton (1971) to derive surprisingly informative expressions for the optimal human capital investment, s_{it}, and the optimal risky share, i.e. risky financial assets as a share of total financial assets, which we denote by $RISK$. In continuous time, when shocks to the returns to risky financial capital and human capital do not covary ($\sigma_{\theta,f} = 0$) we have

$$s_{it} = \frac{\mu_\theta - \eta_{it}}{\sigma_\theta^2 \rho_{it}} \tag{3}$$

$$RISK_{it} = \frac{\mu_f - r}{\sigma_f^2 \rho_{it}} \tag{4}$$

---

[17]An alternative way to model human capital investments is to consider a discrete choice set of alternatives characterized by different earnings variances, see e.g. Wiswall and Zafar (2015) and Nielsen and Vissing-Jorgensen (2006). In our model, the individual chooses *how much* human capital to invest given a constant variance in the returns.

where ρ refers to the Pratt-Arrow index of relative risk aversion.[18] η is the marginal rate of substitution between financial wealth and human capital, which is further defined in Appendix A.2. The latter is typically increasing over the life-cycle as the gain in expected future earnings falls closer to retirement.

Equations (3) and (4) both illustrate the mean-variance trade-offs facing the individuals. Equation (3) shows that the optimal human capital investment depends positively on the "net return", which is given by $\mu_\theta - \eta_{it}$, and negatively on the product of the variance of the return, σ_θ^2, and relative risk aversion, ρ. In the same spirit, (4) reveals that the risky share depends positively on the risk premium, $\mu_f - r$, and negatively on the product of σ_f^2 and ρ. When shocks to the returns to risky financial capital and human capital covary (i.e. $\sigma_{\theta,f} \neq 0$) individuals may use the financial portfolio to hedge for shocks to human capital returns. This would give rise to an additional term on the right hand side of (4).

From (3) and (4) we make the central observation that relative risk aversion both determines human capital investments and the risky share. Intuitively, a more risk averse individual will invest less in the risky financial asset *and* spend less time on human capital acquisition with uncertain future prospects. Combining (3) and (4) gives:

$$s_{it} = b_{it} RISK_{it}, \tag{5}$$

with $b_{it} = \frac{(\mu_\theta - \eta_{it})\sigma_f^2}{(\mu_f - r)\sigma_\theta^2}$. Hence, at a given point in continuous time, the individual's human capital investment, expressed in time use, is proportional to the risky share.

---

[18]Following Shaw (1996) we make the approximation that the Pratt-Arrow index of relative risk aversion with respect to gambles with human capital equals risk aversion with respect to gambles with financial capital, see the discussion in Appendix A.2.

5.3 The empirical equation

The growth in human capital (given uncertainty in θ) satisfies the following stochastic differential equation:

$$dk_{it} = [\omega_i + \mu_\theta b_{it} RISK_{it}] k_{it} dt + \sigma_\theta b_{it} RISK_{it} k_{it} dZ, \tag{6}$$

where ω represents deterministic wage growth and dZ is the increment of a standardized Wiener process with mean zero. We approximate the solution to (6) as

$$k_{it} \approx k_{0i} e^{t(\omega_i + \mu_\theta \bar{b}_i RISK_i)}, \tag{7}$$

i.e. we treat (6) as an ordinary deterministic differential equation while assuming that $\bar{b}_i$ and $RISK_i$ are constant at the individual level. Given our focus on the gender wage gap we assume that $t\omega_i = \beta_0 + \beta_1 GENDER_i$ and $t\mu_\theta \bar{b}_i = \beta_2$. Moreover, we let $k_{0i} = e^{\gamma X_i}$, where X_i is a vector of variables that are predetermined in the "initial period". Finally, let the observed wage rate be $w_i = k_{it} e^{\epsilon_i}$, where ϵ is an error term. By taking logs we arrive at the following empirical log wage equation:

$$\log w_i = \beta_0 + \beta_1 GENDER_i + \beta_2 RISK_i + \gamma X_i + \epsilon_i, \tag{8}$$

which we estimate by OLS on cross sectional individual level data. In the main specification, we will define $RISK_i$ as the ratio of stocks to the individual's total financial assets. Crucially, $\log w_i$ and $RISK_i$ are simultaneously determined in the model, and $RISK_i$ is highly likely to be endogenous also in a statistical sense. Therefore, we emphasize that we do not interpret β_2 as a causal effect; we will use the term "assocation" rather than "effect". In our baseline specification, we will, however, partial out the influence of wealth from $RISK_i$ using the procedure described in Section 4.1.

Which variables should represent the individual's skill endowment in the "initial period", k_{0i}, and appear in the X_i vector? In our life cycle model, the wage rate is a function of human capital investments since birth. Ideally, we

therefore want to control for factors that are pre-determined at birth. In some specifications, we control for county and country of birth and age (cohort). We will also present specifications, where we include controls for e.g. education and occupation. We expect these factors to mediate the influence of risk preferences on the wage.

Even though we motivate our empirical specification by modeling risk attitudes, we fully acknowledge that finanical risk taking is determined by more personal traits than risk aversion. Differences in financial risk taking may originate from ambiguity aversion (Borghans et al., 2009), financial literacy (Lusardi and Mitchell, 2014; Almenberg and Dreber, 2015), or self-confidence (Barber and Odean, 2001), and we cannot discriminate between these partly related mechanisms. Suppose, however, that these other variables covary with risk aversion and the wage rate with the same sign, an assumption we find plausible. Then the assocation between the wage and risk aversion is smaller than implied by a literal interpretation of our model framework. For that reason, we will not focus on quantifying the parameters of the theoretical model. We will instead highlight to what extent gender differences in financial risk taking can explain gender differences in wages. Given these other factors, we believe that a correct estimate of the impact of the risky share on the gender wage gap provides an upper bound of the impact of risk aversion *per se* on the gender gap. Of course, attenuation bias from measurement errors is the main challenge to this interpretation, and we will discuss measurement issues in Section 6.2 below.

5.4   Comparison with related models and approaches

In our model, we ask how wages (and wealth portfolios) are affected by varying the degree of risk aversion. The moments of the earnings processes can be thought of as being endogenously determined. By contrast, an important literature in financial economics examines how the financial portfolio is affected by exogenous changes in the variance of non-financial income (earnings) and in the covariance between earnings shocks and stock returns, see e.g Guiso et al. (1996), Vissing-Jorgensen (2002), and Betermier et al. (2012). A central hypothesis is that risk-averse agents reduce the risky share if they are exposed to higher

earnings risk. The estimation equations typically feature the risky share (or the probability to hold stocks) on the left hand side, and moments of the earnings and stock return processes on the right hand side. Unlike the model presented above, these models do not contain endogenous human capital investment decisions. The hedging literature addresses a different research question, even though fairly similar variables are involved in the analysis.

Moreover, individual heterogeneity in risk attitudes conceptually distinguishes our model from the structural approaches of Nielsen and Vissing-Jorgensen (2006) and Wiswall and Zafar (2015). Both examine the determinants behind education decisions and recover homogeneous relative risk aversion parameters from discrete choice models.[19] None of these papers, however, use data on financial portfolios. The same holds true for the literature on risk augmented Mincer equations, which started with King (1974) and recently was surveyed by Hartog (2011). The idea is that someone who chooses a specific education does not receive a deterministic wage rate, but instead makes draws from a wage distribution conditional on the career choice. Therefore, researchers add moments (variance and skewness) of education-occupation-specific residuals to the Mincer equation, and they often find evidence of a positive risk premium.[20] Finally, it should also be mentioned that risk preferences are determining reservation wages in standard job search models. E.g., Krueger and Mueller (2016) find a positive effect of risk loving on the reservation wage ratio. Mechanisms operating through occupational choices and risk-taking in wage bargaining are indeed closely related to the human capital mechanisms we are modeling explicitly.

[19]Nielsen and Vissing-Jorgensen (2006) use estimated income processes for 50 educational groups, whereas Wiswall and Zafar (2015) use experimentally generated data on beliefs about future earnings in a number of U.S. college major choices.

[20]Relatedly, Dillon (2018) recently estimated a life cycle model of career choice, while considering occupation-specific wage and employment risk. She finds that earnings risk differs substantially across occupations; individuals perceive these differences and demand compensation for it.

6 Regression results

In this section we report results from estimating the wage equation (8). We first present the most basic specifications. Later on we will consider a richer set of outcomes, robustness checks, and results by household type. Throughout this section, our preferred measure of financial risk taking is the risky share, i.e. the ratio of stocks to financial wealth. In the baseline specification, we use the residualized measure of the risky share, which is independent of the financial wealth level. The focus on the risky share measure follows from our model framework. A binary indicator for holding stocks is a common alternative measure in the finance literature. Our main conclusions in this study are essentially unaltered if we replace the residualized risky share with a residualized measure of an indicator for holding stocks.

6.1 Wages

Our most central dependent variable is the log wage rate. Note that we are not controlling for financial wealth in the regressions, i.e. we only partial out wealth from the risky share, but not from the log wage. Financial wealth is not predetermined at birth, and we want to allow for sorting in as many dimensions as possible. In Section 6.2 we show that results are similar when controlling for financial wealth. In Table 3 we estimate (8), while alternating the set of control variables included in the X vector. Throughout, we report the coefficients for the gender dummy and the risky share residual. Moreover, we report *the risk impact on the gender gap (%)*, which we define as $\frac{\hat{\beta}_1 - \hat{\beta}_1^{restricted}}{\hat{\beta}_1^{restricted}} \times 100$, where $\hat{\beta}_1^{restricted}$ is the coefficient for the gender dummy from a "standard" regression without the risky share regressor.

In the specification without controls, which is reported in column 1 of Table 3, the coefficient for the risky share amounts to 0.14, and it is precisely estimated. Hence, a one percentage point increase in the risky share is associated with an increase of 0.14 percent in the wage rate. The inclusion of the risky share has a clear impact on the estimated gender wage gap. However, the impact is smallish; it lowers the gender wage gap by 3.1 %. In column 2 we include control

variables that plausibly are predetermined at birth: a quadratic in age, county of birth dummies and country of birth dummies.[21] The key estimates are virtually unaltered when including the predetermined controls. We consider the results of column 2 as being our baseline results, and they suggest that differences in financial risk taking between men and women cannot explain much of the observed gender wage gap.[22]

Is it possible to make the association between risk taking and the wage rate insignificant by adding more control variables that are known to reduce the gender wage gap? In column 3 we continue by adding a large set of dummies for the level and field of education and its interactions to the set of controls. Educational choices are *not* predetermined at birth, and in our life cycle model we expect risk preferences to determine human capital investments, which in turn determines the wage rate. When controlling for education the estimated gender gap shrinks significantly. However, the coefficient for the risky share decreases even more, and the impact on the estimated gender wage gap is halved when going from column 2 to 3. In column 4, we add an upper layer of mediators when also controlling for occupation and industry. Still, there is a significant gender gap of 7.2 log points, which cannot be explained by observable characteristics. One should also note that the ambitious set of controls cannot entirely pick up the association between the risky share and the wage: Even though the risky share coefficent is now by an order of magnitude smaller than in the first two columns, it is still significantly distinct from zero at a level of 1 %.

In Table 4 we instead estimate equation (8) using the *raw* risky share as the main regressor of interest. The regression output is displayed in Table 4. Interestingly, both the risky share coefficients and the impact on the gender gap are roughly doubled relative to the baseline specification of Table 3. In the specification without controls (column 1) the impact on the gender gap is now 7.5 %. We consider 7.5 % as an upper bound of the impact of financial risk taking on

---

[21]Since we use a cross section from a specific year (2000) age is equivalent to birth cohort.

[22]By running quantile regressions we also examined how the impact differs at different quantiles of the wage distribution, see Appendix A.4. The results indicate that the impact is larger at the top of the distribution. But the differences are not dramatic; at the 90th percentile the risk variable explains 5% of the gender wage gap.

Table 3: Log wage on risky share residuals (stocks/financial assets)

	(1)	(2)	(3)	(4)
Female	-0.192**	-0.193**	-0.140**	-0.0721**
	(0.00249)	(0.00245)	(0.00268)	(0.00274)
Risky share (res)	0.144**	0.145**	0.0635**	0.0177**
	(0.00785)	(0.00777)	(0.00618)	(0.00518)
Impact on gender gap (%)	-3.124	-3.080	-1.605	-0.821
Predetermined	No	Yes	Yes	Yes
Schooling	No	No	Yes	Yes
Occupation	No	No	No	Yes
Adjusted R-squared	0.118	0.153	0.490	0.725
Observations	56,113	56,113	56,113	56,113

Note: Robust standard errors in parenthesis. * indicates significance at 5% level and ** at 1% level.

the gender wage gap.[23]

Using the raw risky share is the simplest possible specification, and it does have an intuitive appeal. However, the downside is obvious: There is a trivial feedback from earnings on financial wealth (those who earn more can save more), and since stock ownership is positively correlated with wealth, the risky share coefficients of Table 4 partly reflects the correlation between the log wage and financial wealth. Henceforth, we focus on specifications with the risky share residual as the regressor of interest. In what follows, we also restrict our attention to specifications where we only control for predetermined observable characteristics. This choice is guided by our life-cycle model, in which the wage rate is a function of human capital investments during the course of the entire life cycle. We emphasize that results with and without predetermined controls in general do not differ much.

Is the association between the risky share and the wage similar for men and

---

[23]A caveat is potential measurement errors in the risky share, which may attenuate the estimates towards zero. We further discuss this issue in Section 6.2.

Table 4: Log wage on the raw risky share (stocks/financial assets)

	(1)	(2)	(3)	(4)
Female	-0.184**	-0.185**	-0.139**	-0.0715**
	(0.00243)	(0.00240)	(0.00266)	(0.00273)
Risky share	0.299**	0.285**	0.120**	0.0427**
	(0.00736)	(0.00727)	(0.00591)	(0.00491)
Impact on gender gap (%)	-7.532	-7.205	-2.987	-1.640
Predetermined	No	Yes	Yes	Yes
Schooling	No	No	Yes	Yes
Occupation	No	No	No	Yes
Adjusted R-squared	0.149	0.181	0.494	0.725
Observations	56,113	56,113	56,113	56,113

Note: Robust standard errors in parenthesis. * indicates significance at 5% level and ** at 1% level.

women? *A priori* we did not know what to expect. In Table 5 we report regressions where the gender dummy has been interacted with the risky share. Column 1 is similar to the baseline specification in column 2 of Table 3, but with an interaction term added. We make two observations from column 1. First, the impact on the gender gap is unaffected when the interaction term is included.[24] Second, there are indeed significant differences between men and women. In both gender groups the association is positive, but the association is significantly stronger for men, and the significant interaction term does not seem to be fully driven by non-linearities in the relationship between log wages and the risky share.[25] One may only speculate about the underlying mechanisms. One story could be that labor market sorting based on risk attitudes is stronger for

---

[24]We replace $\hat{\beta_1}$ in $\frac{\hat{\beta_1} - \hat{\beta_1}^{restricted}}{\hat{\beta_1}^{restricted}} \times 100$ with $\hat{\Delta}(\hat{\beta}, \hat{\gamma}, \overline{RISK}, \overline{X}) = \log w(GENDER_i = female) - \log w(GENDER_i = male)$, which is the gender effect on the log wage evaluated at the means of the other independent variables.

[25]When adding a fourth degree polynomial in the risky share residual, the point estimate for the interaction term is still -0.036, with a p-value of 0.017.

30

men than women. Alternatively, if women, conditional on choosing a risky but high-paid job, are more prone than men to rebalance their portfolios to safer assets (hedging), we would perhaps also observe this pattern. But we emphasize that these remarks are speculative.

In column 2 of Table 5 we run a similar regression on the same sample, but we replace the log wage rate with log earnings as dependent variable. The latter quantity is the product of work hours and the wage rate. In general, the same preference parameters determining risk aversion affect optimal labor supply choices in life-cycle models with both consumption and labor supply choices. Column 2 shows that the risky share coefficient for males is somewhat larger for earnings and wage rates: the point estimate is 0.19 for earnings compared with 0.17 for wages. Among women we see the opposite: 0.07(=0.187-0.114) for earnings vs. 0.11(=0.172-0.0635) for wage rates. For men there is a strong mapping between the wage rate and earnings; 94 % in the wage sample has full time contracts, the same figure for women is 54 % only. As a consequence, the gender gap in the wage sample is twice as large for earnings. Expressed in terms of log points the impact on the gender gap is similar for wages and earnings (0.61 vs. 0.58).

6.2 Specification issues

A potential problem associated with the analysis is that the "wage sample" is non-randomly selected. In particular, women are "oversampled" since they are more likely to be public sector employees. Moreover, as we highlighted in Section 4.2, individuals may sort into the group of public sector employees based on risk attitudes. Fortunately, we have data on labor earnings for all individuals in the sample. This enables us to compare the association between the risky share and earnings for individuals with and without data on wage rates. We report this specification test in the Appendix, Table A2.[26] For both men and

---

[26]When comparing the wage and no wage samples we exclude the bottom earnings decile (defined based on the total sample). The reason is that the no wage sample – in contrast to the wage sample – contains a large mass of individuals (mostly women) with very low earnings. When including people with very low earnings the wage sample and no wage sample are less comparable.

Table 5: Interacted specifications

	(1) Log wage	(2) Log earnings
Female	-0.194**	-0.404**
	(0.00246)	(0.00462)
Risky share (res)	0.172**	0.187**
	(0.0115)	(0.0167)
Female × Risky share	-0.0635**	-0.114**
	(0.0150)	(0.0262)
Impact on gender gap (%)	-3.066	-1.413
Adjusted R-squared	0.154	0.149
Observations	56,113	56,050

Note: Robust standard errors in parenthesis. * indicates significance at 5% level and ** at 1% level. All specifications include a set of predetermined control variables: a quadratic in age, county of birth, and country of birth dummies. Risky share residuals were obtained by regressing the raw risky share on 100 percentile dummies in financial wealth in the full LINDA sample of individuals aged 31-60.

women, the risky share coefficient is slightly lower in the no wage sample, but the differences between the two samples are far from being statistically significant.[27] We consider this result to be reassuring, and it indicates that our baseline results for wage rates also hold in representative samples.

Another concern is attenuation bias due to measurement error in the risky share. One possible way to account for measurement error is to average the risky share over several years. In column 2 of Table 6 we report a regression, where we replaced the contemporaneous risky share with a 3-average over 2000-02. The point estimate for the risky share increases (in a small but significant way), providing some indications of attenuation bias in the baseline estimates. The perhaps most acute measurement issue is, however, the imputation of bank holdings, which are missing for a substantial proportion of the sample. We have done a host of robustness tests regarding the imputation procedure, with no substantial changes to the results.[28] The most radical robustness check in this respect is reported in column 3 of Table 6, where we simply exclude 61 % of the sample, i.e. those with imputed bank holdings, and we rerun the baseline regression. In this sample, people are wealthier and earn higher wages; the gender gap is therefore larger (Albrecht et al., 2015). The risky share is associated with a significantly larger wage than in the baseline. The impact on the gender gap is now larger in percentage terms, 4.3 % compared to 3.1 % (despite the fact that the initial gap was larger). Still, the results of column 3 leave the qualitative conclusions unaffected.

We have chosen to focus on the ratio of stocks to financial wealth. An investor is exposed to idiosyncratic risk when investing in individual stocks. What happens if one also includes mutual funds in the numerator of the risky share? The answer is provided in column 4 of Table 6. The association between the risky share and the wage rate now becomes significantly lower. In Section

---

[27]When testing for inequality of β_{2a} and β_{2b}, where a and b refer to different regressions, we first estimate a and b as a system of seemingly unrelated regressions (SUR) using the suest command in Stata. We then perform a t test, where $\beta_{2a} = \beta_{2b}$ is the null hypothesis. We use this procedure when a and b are run on the same, overlapping and non-overlapping samples.

[28]In the baseline specification, we set bank holdings to SEK 10,000 for everyone with reported bank holdings below SEK 10,000. The results are insensitive to changing the cut-off to SEK 5,000 or SEK 20,000. Moreover, following Calvet et al. (2007) we imputeded bank holdings using a regression based approach.

3 we recognized that gender differences are much more pronounced for stocks than for the sum of stocks and mutual funds. Hence, it comes as no surprise that the risky share explains close to zero of the gender wage gap when the numerator of the risky share is the sum of stocks and mutual funds.

A final specification issue relates to the risky share residuals. As explained in Section 4.1 above, these represent deviations from group means in wealth percentile groups. Even though the mean of residuals by construction is zero in all groups, higher moments of the distributions vary across wealth groups; the variance increases and skewness decreases when moving upwards in the wealth distribution, and the mean raw risky share increases. This heteroskedastictiy may cause problems when not controlling for wealth dummies, because people with high wealth, with more extreme negative values of risky share residuals, will also tend to earn more.[29] This problem is absent when controlling for wealth dummies directly. Such a specification, which uses the raw risky share as the independent variable of interest, is presented in column 5 of Table 6. The results are very similar.

6.3   The household dimension

In the first two columns of Table 7 we estimate our main model for "singles" and individuals in couples separately using our individual risky share measure. In Swedish administrative data "singles" both refer to genuine singles and cohabiting without common children, while "couples" refer to formally married individuals or to cohabiting individuals with common children. The most striking difference between column 1 and 2 is that the estimatad gender pay gap is almost twice as large among individuals in couples, which most likely relates to the role of children for long run human capital accumulation, see e.g. Angelov et al. (2016) and Lundborg et al. (2017).

---

[29]To illustrate, in the top wealth groups the mean raw risky share is substantially larger than at the bottom. This implies that someone with a zero (or just low) risky share has a larger negative value of the risky share residual at the top than at the bottom of the distribution. Since people with large wealth holdings also tend to earn high wages we observe a negative association between the risky share residual and the wage at very low (negative) values of the risky share residuals.

Table 6: Log wage regressions: various specifications

	(1) Baseline	(2) 3-year average share	(3) Bank holdings > SEK 10,000	(4) Mutual funds included	(5) Wealth controls
Female	-0.193**	-0.193**	-0.227**	-0.199**	-0.185**
	(0.00245)	(0.00245)	(0.00432)	(0.00245)	(0.00232)
Risky share	0.145**	0.158**	0.200**	0.0488**	0.148**
	(0.00777)	(0.00836)	(0.0126)	(0.00504)	(0.00757)
Impact on g.g. (%)	-3.080	-3.130	-4.320	-0.292	-3.300
Adjusted R-squared	0.153	0.154	0.177	0.147	0.233
Observations	56,113	55,928	21,951	56,113	56,113

Note: Robust standard errors in parenthesis. * indicates significance at 5% level and ** at 1% level. All specifications include a set of predetermined control variables: a quadratic in age, county of birth, and country of birth dummies. In the baseline specification (column 1), risky share residuals were obtained by regressing the raw risky share on 100 percentile dummies in financial wealth in the full LINDA sample of individuals aged 31-60. When constructing the risky share measure in column 2 we first take the average stock holdings, financial wealth and bank holdings for each individual for the years 2000-02. Average bank holdings are imputed by taking the maximum of the real 3-year average and SEK 10,000. Then we define the raw risky share as the ratio of average stock holdings 2000-02 to average imputed financial wealth 2000-02. We obtain risky share residuals by regressing the raw average risky share on percentile dummies in imputed average financial wealth. In column 4, we use a similar procedure as in column 1 and 3, while including mutual fund holdings in the numerator of the risky share. In column 5 we use the raw risky share as regressor, and we control for 100 percentile dummies.

Up to now our analysis has been conducted entirely at the individual level, and we have neglected the partner's assets when defining the risky share. In fact, variation in individual wealth portfolios is crucial to our research design, because household variation cannot explain gender differences among individuals in couples. Still, using individual level data is non-standard, since most wealth data sources provide information on household level wealth variables. If spouses specialize into different types of asset holdings, while balancing the total household portfolio, our individual measure does not necessarily reflect risk preferences. Researchers using Swedish wealth data often use aggregate household wealth portfolios, see e.g. Thörnqvist and Vardardottir (2015) who examined the effect of the wife's bargaining power on the household wealth portfolio. We constructed risky share measures at the household level by replacing the ratio of the individual's stocks to the individual's financial assets with the ratio of the two partners' stock holdings to the two partner's financial wealth.[30] In column 3 of Table 7 we re-estimate our model on partners, while replacing the individual risky share with the household's risky share. Comparing columns 2 and 3, it is evident that the risky share estimates are remarkably similar.[31]

7 Conclusion

Since long economists have recognized that risk attitudes may shape labor market outcomes. In this paper we found that financial risk-taking is positively associated with the wage rate, often viewed as a summary measure of accumulated human capital. The association is robust to various changes in the specification and in different subsamples. The result is a non-trivial one, and it indicates labor market sorting based on risk attitudes. Consistent with our priors, we documented significant gender differences in financial risk taking using a representative sample of Swedes aged 31-60, and we examined if gender

---

[30]We construct the risky share residuals by regressing 100 percentile dummies in own financial wealth and 100 percentile dummies in the partner's financial wealth.

[31]There is a small effect on the gender gap in column 3. This is an artifact of our sampling design, where we sample *individuals*, not couples, aged 31-60 with observed wage rates.

Table 7: The household dimension

	(1)	(2)	(3)
	"Singles"	Individuals in couples	
Female	-0.120***	-0.231***	-0.237***
	(0.00388)	(0.00310)	(0.00307)
Individual risky share (res)	0.115***	0.148***	
	(0.0131)	(0.00952)	
Household risky share (res)			0.144***
			(0.00893)
Impact on gender gap (%)	-3.083	-2.936	-0.333
Adjusted R-squared	0.092	0.191	0.191
Observations	18,799	37,314	37,314

Note: Robust standard errors in parenthesis. * indicates significance at 5% level and ** at 1% level. All specifications include a set of predetermined control variables: a quadratic in age, county of birth, and country of birth dummies. Risky share residuals were obtained by regressing the raw risky share on 100 percentile dummies in financial wealth in the full LINDA sample of individuals aged 31-60. LINDA contains wealth data for the family members of the sampled individual. For individuals in couples the raw risky household share is defined as the ratio of the household's stock holdings to the housould's (imputed) financial wealth. Household level variables are obtained by summing the two spouses' wealth variables. We generate residuals by regressing the raw household risky share on 100 percentile group dummies in own financial wealth and 100 percentile dummies in the partner's financial wealth.

differences in portfolio composition possibly could explain parts of the gender wage gap. The risky share impacted the estimated gender gap in wages, but the impact seems relatively small.

In our model we asked how heterogeneity in risk attitudes impacts labor outcomes. In our empirical discussion we deliberately avoided interpreting the estimates in terms of "effects". From a model perspective, we regressed one endogenous quantity – the realized wage – on another endogenous quantity – the realized risky share, while acknowledging that the underlying risk preferences are unobservable. Furthermore, financial risk taking captures other personal traits than pure risk aversion, e.g. financial literacy and self-confidence.

In fact, it is difficult to imagine how it would be like to establish the causal effect of risk attitudes on labor market outcomes, because it is difficult to vary *preferences* randomly or quasi-randomly. To learn about this kind of sorting behavior – at least outside laboratory environments – we therefore believe that this is the right way to go.[32] In comparison to self-reported data on risk aversion, a clear advantage of our administrative wealth data is that it reflects realized choices (not stated preferences), and the non-response rate is zero.

---

[32]Another way to examine the association between risk attidues and labor marker outcomes would be to predict future labor market outcomes of students from lab measures of risk preferences. Such a research design could potentially allow researchers to discriminate betweeen e.g. risk and ambiguity aversion, which we cannot do here. On the other hand, generalizibility would be more limited than in our representative sample.

References

Albrecht, J., A. Bjorklund, and S. Vroman (2003, January). Is There a Glass Ceiling in Sweden? *Journal of Labor Economics 21*(1), 145–177.

Albrecht, J., P. S. Thoursie, and S. Vroman (2015). Parental Leave and the Glass Ceiling in Sweden. In *Research in Labor Economics, Volume 41 - Gender Convergence in the Labor Market*, Chapter 3, pp. 89–114.

Almenberg, J. and A. Dreber (2015). Gender, stock market participation and financial literacy. *Economics Letters 137*(C), 140–142.

Altonji, J. G. and R. M. Blank (1999). Chapter 48 race and gender in the labor market. Volume 3 of *Handbook of Labor Economics*, pp. 3143 – 3259. Elsevier.

Angelov, N., P. Johansson, and E. Lindahl (2016). Parenthood and the gender gap in pay. *Journal of Labor Economics 34*(3), 545–579.

Azmat, G. and B. Petrongolo (2014). Gender and the labor market: What have we learned from field and lab experiments? *Labour Economics 30*(C), 32–40.

Barber, B. M. and T. Odean (2001). Boys will be boys: Gender, overconfidence, and common stock investment. *The Quarterly Journal of Economics 116*(1), 261–292.

Bertrand, M. (2011). *New Perspectives on Gender*, Volume 4 of *Handbook of Labor Economics*, Chapter 17, pp. 1543–1590. Elsevier.

Betermier, S., T. Jansson, C. Parlour, and J. Walden (2012). Hedging labor income risk. *Journal of Financial Economics 105*(3), 622 – 639.

Black, S. E., P. J. Devereux, P. Lundborg, and K. Majlesi (2015, March). Learning to take risks? the effect of education on risk-taking in financial markets. Working Paper 21043, National Bureau of Economic Research.

Bonin, H., T. Dohmen, A. Falk, D. Huffman, and U. Sunde (2007). Cross-sectional earnings risk and occupational sorting: The role of risk attitudes. *Labour Economics 14*(6), 926 – 937.

Borghans, L., J. J. Heckman, B. H. H. Golsteyn, and H. Meijers (2009, 05). Gender Differences in Risk Aversion and Ambiguity Aversion. *Journal of the European Economic Association 7*(2-3), 649–658.

Calvet, L. E., J. Y. Campbell, and P. Sodini (2007, October). Down or Out: Assessing the Welfare Costs of Household Investment Mistakes. *Journal of Political Economy 115*(5), 707–747.

Cho, I. S. (2011). Do gender differences in risk preferences explain gender differences in labor supply, earnings or occupational choice? WP Nr. 11022 Economics Working Papers. Iowa State University.

Christiansen, C., J. S. Joensen, and H. S. Nielsen (2007). The risk-return trade-off in human capital investment. *Labour Economics 14*(6), 971 – 986. Education and Risk.

Croson, R. and U. Gneezy (2009, June). Gender differences in preferences. *Journal of Economic Literature 47*(2), 448–74.

Dillon, E. W. (2018). Risk and return trade-offs in lifetime earnings. *Journal of Labor Economics 36*(4), 981–1021.

Dohmen, T., A. Falk, D. Huffman, U. Sunde, J. Schupp, and G. G. Wagner (2011). Individual risk attitudes: Measurement, determinants, and behavioral consequences. *Journal of the European Economic Association 9*(3), 522–550.

Edin, P.-A. and P. Fredriksson (2000). LINDA - Longitudinal INdividual DAta for Sweden. Working Paper Series 2000:19, Uppsala University, Department of Economics.

Flood, L. (2004). Formation of wealth, income of capital and cost of housing in sesim. Technical report, University of Gothenburg.

Guiso, L., T. Jappelli, and D. Terlizzese (1996). Income risk, borrowing constraints, and portfolio choice. *The American Economic Review 86*(1), 158–172.

Hartog, J. (2011). *Chapter 4 A Risk Augmented Mincer Earnings Equation? Taking Stock*, Chapter Research in Labor Economics, pp. 129–173.

Jung, S. (2017). The gender wage gap and sample selection via risk attitudes. *International Journal of Manpower 38*(2), 318–335.

King, A. G. (1974). Occupational choice, risk aversion, and wealth. *ILR Review 27*(4), 586–596.

Koenker, R. W. and G. Bassett (1978). Regression quantiles. *Econometrica 46*(1), 33–50.

Krueger, A. B. and A. I. Mueller (2016, February). A contribution to the empirics of reservation wages. *American Economic Journal: Economic Policy 8*(1), 142–79.

Le, A., P. Miller, W. S. Slutske, and N. G. Martin (2011). Attitudes towards economic risk and the gender pay gap. *Labour Economics 18*(4), 555–561.

Levhari, D. and Y. Weiss (1974). The effect of risk on the investment in human capital. *The American Economic Review 64*(6), 950–963.

Lundborg, P., E. Plug, and A. W. Rasmussen (2017, June). Can Women Have Children and a Career? IV Evidence from IVF Treatments. *American Economic Review 107*(6), 1611–37.

Lusardi, A. and O. S. Mitchell (2014, March). The economic importance of financial literacy: Theory and evidence. *Journal of Economic Literature 52*(1), 5–44.

Merton, R. (1971). Optimum consumption and portfolio rules in a continuous-time model. *Journal of Economic Theory 3*(4), 373–413.

Nielsen, H. S. and A. Vissing-Jorgensen (2006). The impact of labor income risk on educational choices: Estimates and implied risk aversion. Working paper.

Säve-Söderbergh, J. (2012). Self-directed pensions: Gender, risk, and portfolio choices. *The Scandinavian Journal of Economics 114*(3), 705–728.

Shaw, K. L. (1996, October). An Empirical Analysis of Risk Aversion and Income Growth. *Journal of Labor Economics 14*(4), 626–653.

Sundén, A. (2006). The Swedish Experience with Pension Reform. *Oxford Review of Economic Policy 22*(1), 133.

Thörnqvist, T. and A. Vardardottir (2015). Bargaining over risk. the impact of decision power on household portfolios.

Vissing-Jorgensen, A. (2002, April). Towards an explanation of household portfolio choice heterogeneity: Nonfinancial income and participation cost structures. Working Paper 8884, National Bureau of Economic Research.

Williams, J. T. (1979). Uncertainty and the accumulation of human capital over the life cycle. *The Journal of Business 52*(4), 521–548.

Wiswall, M. and B. Zafar (2015). Determinants of college major choice: Identification using an information experiment. *The Review of Economic Studies 82*(2), 791–824.

A Appendix

A.1 Summary statistics

Summary statistics are reported in Table A1.

A.2 Model appendix

Our empirical model draws on Shaw (1996). However, in contrast to Shaw (1996) we derive an equation for the wage *level* rather than wage growth. Shaw (1996) builds on Williams (1979), who, in turn, augmented the continuous-time portfolio-model of Merton (1971) to include human capital investments. The purpose of this appendix is to sketch the derivations of equations (3) and (4) of Section 5.2. For complete derivations we refer the reader to Williams (1979), who analyzed a very rich model structure. Here we make some simplifications of his model. First, we assume that wage growth (conditional on human capital investments) is deterministic and given by ω. Hence, given a realized value θ the individual perfectly observes future wage growth. Second, we set depreca-tion of human capital (modeled by Williams as being stochastic) to zero. Finally, following Shaw (1996) we assume, for notational convenience, that there is only one risky asset.

Suppressing the individual index i life-time utility in continuous time can be written

$$\max E\{\int_0^T U[c_t, l_t, k_t, t]dt + B[W(T), T]\} \tag{9}$$

, where c_t is consumption at time t, l_t is leisure, k_t is human capital and $B[W(T), T]$ is a bequest function (non-decreasing and concave in terminal financial wealth $W(T)$). Instantaneous utility U is strictly concave in consumption and leisure. The agent maximizes (9) subject to human capital constraint

$$dk_t = [\omega + \mu_\theta s_t]k_t dt + \sigma_\theta s_t k_t dZ \tag{10}$$

, where the notation follows Section 5.2. The financial wealth constraint can be

43

Table A1: Summary statistics, by gender, in total sample and wage sample

	(1) Complete Sample Males Mean	(2) Complete Sample Females Mean	(3) Difference (1)-(2)	(4) Wage Sample Males Mean	(5) Wage Sample Females Mean	(6) Difference (4)-(5)
Age	45.24	45.28	-0.04	45.27	45.71	-0.44**
Foreign born	0.15	0.15	-0.00	0.11	0.11	-0.00
Married or cohabiting	0.61	0.64	-0.04**	0.66	0.67	-0.02**
College degree	0.28	0.32	-0.04**	0.37	0.39	-0.02**
Having children in household	0.51	0.59	-0.08**	0.55	0.60	-0.06**
Earnings	225,988.08	150,802.62	75,185.47**	294,109.11	197,565.66	96,543.45**
Earnings > 0	0.87	0.86	0.01**	1.00	1.00	-0.00
Earnings > 1PBA	0.82	0.78	0.05**	0.99	0.98	0.01**
Earnings > 2PBA	0.79	0.72	0.07**	0.98	0.94	0.04**
Public sector employee	0.15	0.44	-0.29**	0.31	0.67	-0.36**
Monthly wage (full time equivalent)				23,982.04	18,930.61	5,051.43**
Contracted hours (in % of full time)				97.53	86.99	10.54**
Real assets	518,009.35	397,000.02	121,009.33**	495,366.12	415,554.86	79,811.26**
Financial assets (bank acc. imputed)	209,742.58	144,873.23	64,869.35**	193,367.52	149,120.10	44,247.42**
Debt	279,618.35	200,661.75	78,956.59**	261,502.65	207,719.50	53,783.15**
Net wealth	477,497.84	345,152.84	132,345.00**	437,105.05	357,782.12	79,322.93**
Ratio stocks to financial assets	0.11	0.07	0.04**	0.12	0.07	0.05**
Ratio stocks to financial assets, residuals	0.02	-0.02	0.04**	0.02	-0.02	0.04**
Indicator for stocks	0.36	0.25	0.11**	0.40	0.27	0.13**
Observations	62717	60026	122743	24859	31254	56113

Note: The total sample includes all individuals aged 31-60 who were sampled in LINDA in year 2000. Unweighted summary statistics. PBA denotes "price base amounts", SEK 36,600 in year 2000. In Section 4.1 we explain "Ratio stocks to financial assets, residuals".

44

written

$$dW_t = [(rW_t + y_t - c_t) + (\mu_f - r)RISK_t]dt + W_t\sigma_f RISK_t dZ \qquad (11)$$

, where W is financial wealth and dZ is the increment of a standardized Wiener process with mean zero. Using stochastic dynamic programming Williams defines the following indirect utility function corresponding to this problem:

$$J(k_t, W_t, t) \equiv \max E\{\int_t^T U[c_\tau, l_\tau, k_\tau, t]d\tau\} + B[W(T), T]\} \qquad (12)$$

, which is strictly concave in k and W. $J(k_t, W_t, t)$ is expected utility conditional on the values of the state variables k_t (human capital) and W_t (non-human capital) at time t. Imperfect substitutability between human and non-human wealth arises since human capital, unlike financial capital, is neither reversible nor marketable.

Let J_W and J_k denote partial derivatives. Using (12) we define the following quantities:

- $\eta(k(t), W(t), t) \equiv \frac{J(k(t), W(t), t)_W}{J(k(t), W(t), t)_k}$ is the current marginal rate of substitution between financial and human capital.

- $\rho^\theta(k(t), W(t), t) \equiv -\frac{J(k(t), W(t), t)_{kk} \times k}{J(k(t), W(t), t)_k}$ is the Pratt-Arrow index of relative risk aversion for gambles with existing human capital.

- $\rho^f(k(t), W(t), t) \equiv -\frac{J(k(t), W(t), t)_{WW} \times W}{J(k(t), W(t), t)_W}$ is the Pratt-Arrow index of relative risk aversion for gambles with existing financial capital.

For interior solutions of s_t and $RISK_t$, Williams derives the following implicit relationships for optimal allocations:

$$s_t = \frac{\mu_\theta - \eta_t}{\sigma_\theta^2 \rho_t^\theta} \qquad (13)$$

$$RISK_t = \frac{\mu_f - r}{\sigma_f^2 \rho_t^f} \qquad (14)$$

, see equations 10 and 11 in Williams. To further characterize the relationship between ρ_t^θ and ρ_t^f we have to impose more structure on preferences. When the instantaneous utility function and the bequest function exhibit the following properties

$$U[c_t, l_t, k_t, t] = a_t[c_t - \tilde{c}_t]^\gamma [l_t k_t]^\lambda \tag{15}$$

$$B[W(T), T] = a_T[W_T - \tilde{W}_T]^{\gamma + \lambda} \tag{16}$$

, where a_t and $\tilde{c}_t$ are time-varying parameters, the indirect utility function has the following approximate solution

$$J_t \approx m_t[q_t k_t + W_t - \tilde{W}_t]^{\gamma + \lambda} \tag{17}$$

Taking partial derivatives with respect to (17) we obtain

$$\eta_t = \frac{1}{q_t}, \tag{18}$$

$$\rho_t^\theta = (1 - \gamma - \lambda) \frac{q_t k_t}{q_t k_t + W_t - \tilde{W}_t}, \tag{19}$$

$$\rho_t^f = (1 - \gamma - \lambda) \frac{W_t}{q_t k_t + W_t - \tilde{W}_t}, \tag{20}$$

Under these conditions, ρ_t^θ and ρ_t^f are related in the following way:

$$\rho_t^f = \rho_t^\theta \frac{W_t}{q_t k_t} \tag{21}$$

When $W_t = q_t k_t$ the approximation $\rho_t^f \approx \rho_t^\theta$ of equations 3 and 4 holds with equality.
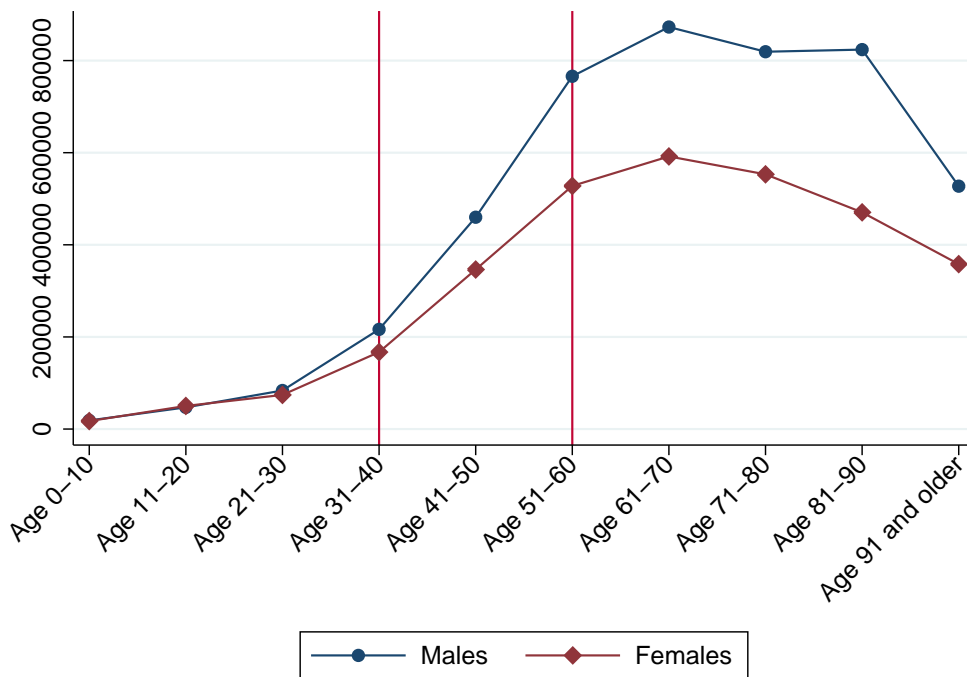
Figure A1: Average net wealth in different age/gender groups. Graph based on all individuals in the LINDA sample. Vertical lines indicate estimation sample in the earnings analysis.

A.3 Graphs by age groups

Figures A1-A3 illustrate some variables as functions of age in the data from year 2000.

A.4 Quantile regressions

In this appendix we present and discuss quantile regression estimates. While a standard OLS regression estimates a conditional mean, a quantile regression estimates a conditional quantile. With quantile regressions we can obtain a view on how e.g. the gender wage gap differes across the wage distribution.[33] Let

[33]Albrecht et al. (2003) did a related exercise on LINDA data from 1998, see Table 2 therein. They investigated the gender wage gap at different quantiles of the wage distribution. The samples are not fully comparable, because we have a more narrow age restriction in our paper. Still, their results are reasonably similar to ours.
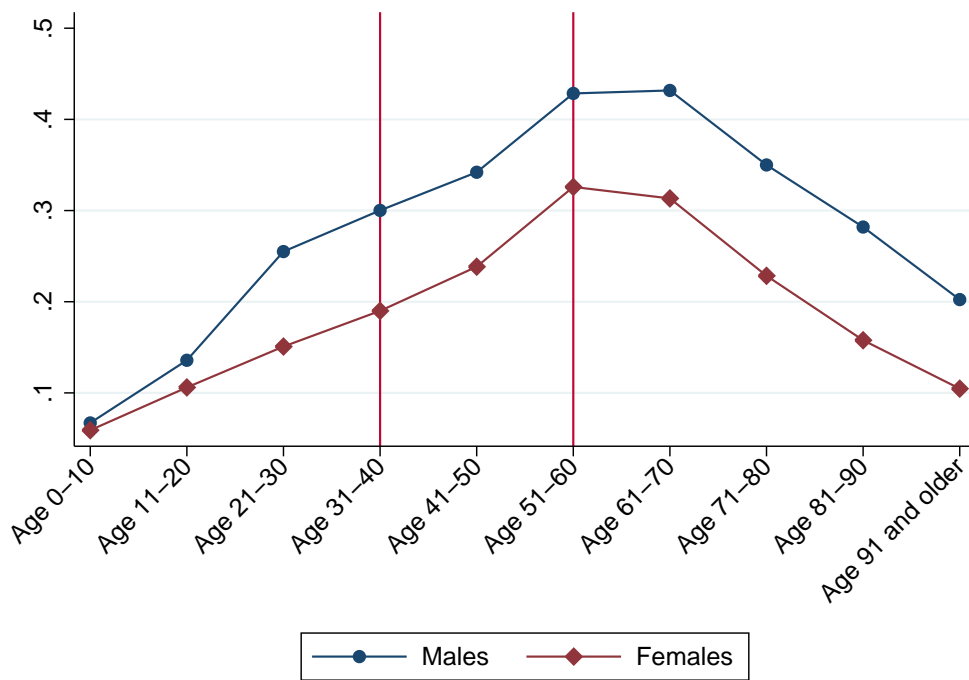
Figure A2: Share with individual stocks in different age/gender groups (extensive margin). Graph based on all individuals in the LINDA sample. Vertical lines indicate estimation sample in the earnings analysis.
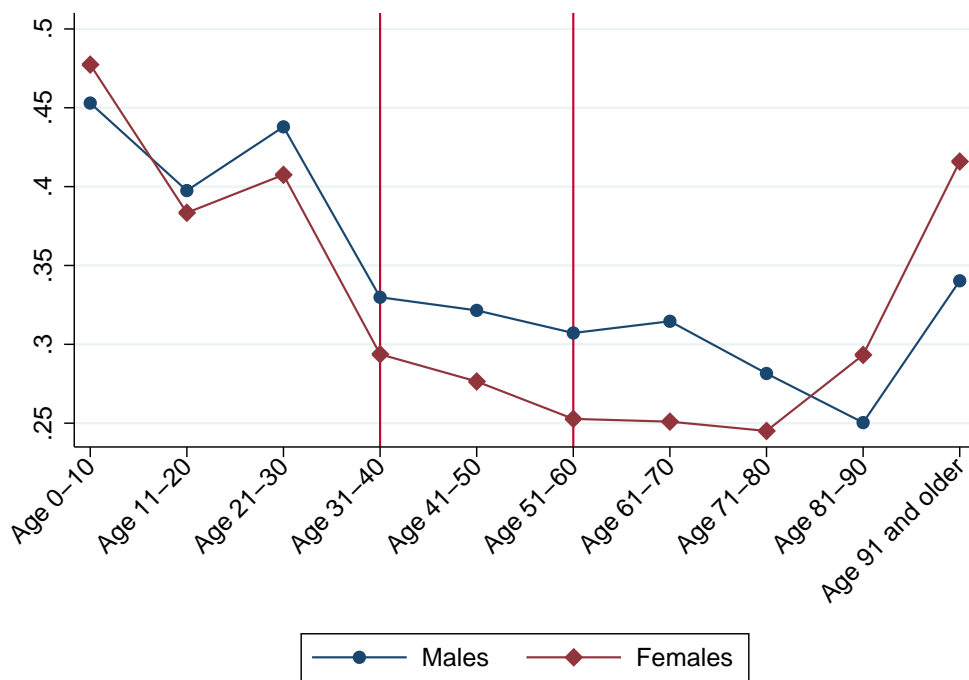
Figure A3: Mean ratio of individual stocks to financial assets in different age/gender groups conditional on holding individual stocks (intensive margin). Graph based on all individuals in the LINDA sample. Vertical lines indicate estimation sample in the earnings analysis.

Table A2: The wage and no wage sample, excluding the bottom log earnings decile

	(1) Wage sample	(2)	(3) No wage sample	(4) Wage sample	(5)	(6) No wage sample
	Log wage	Log earnings	Log earnings	Log wage	Log earnings	Log earnings
Risky share (res)	0.172**	0.175**	0.162**	0.110**	0.0880**	0.0716**
	(0.0115)	(0.0137)	(0.0178)	(0.00987)	(0.0150)	(0.0266)
Observations	24,545	24,545	26,116	30,128	30,128	14,923
Adjusted R-squared	0.058	0.051	0.046	0.039	0.037	0.032

Note: Robust standard errors in parenthesis. * indicates significance at 5% level and ** at 1% level. All specifications include a set of predetermined control variables: a quadratic in age, county of birth, and country of birth dummies. Risky share residuals were obtained by regressing the raw risky share on 100 percentile dummies in financial wealth in the full LINDA sample of individuals aged 31-60. In these regressions the total sample (the sum of the wage and no wage sample) consists of individuals earning more than first decile. We have defined decile based on *log* earnings. This implies that the first earnings decile does not include individuals with zero earnings.

50

$q_\theta = x\beta(\theta)$ be the θ:th quantile of the log wage distribution. This implies that the quantile is assumed to be a linear function of the independent variables x. The θ:th quantile regressor estimator minimizes the following expression over $\beta(\theta)$ (Koenker and Bassett, 1978):

$$\sum_{i:y_i \geq x_i\beta(\theta)}^{N} \theta|y_i - x_i\beta(\theta)| + \sum_{i:y_i < x_i\beta(\theta)}^{N} (1-\theta)|y_i - x_i\beta(\theta)| \tag{22}$$

We have esitmated $\beta(\theta)$ at 19 quintiles, starting at $\theta = 0.05$, ending at $\theta = 0.95$, each increment is 0.05. In Figure A4(a)-(b) the elements of the x vector are the same as in the baseline OLS regression. It includes a dummy for being female, the risky share residual, a quadratic in age, dummies for county and country of birth. Figure A4(a) illustrates the gender wage gap across the wage distribution. We observe the same pattern as Albrecht et al. (2003) and Albrecht et al. (2015): the gender pay gap is the largest at the top ("the glass ceiling").

It is more of an open issue how the risky share residual estimates behave across the wage distribution. Figure A4(b) reveals that the risky share explains more at higher than lower quantiles.

In Figure A4(c) we have calculated the percentage impact on the gender wage gap at different quantiles. More specifically, for the θ:s quantile we compute

$$\frac{\hat{\beta}_1(\theta) - \hat{\beta}_1(\theta)^{restricted}}{\hat{\beta}_1(\theta)^{restricted}} \times 100 \tag{23}$$

, where $\hat{\beta}_1(\theta)$ is the gender dummy estimate reported in Figure A4(a) and $\hat{\beta}_1(\theta)^{restricted}$ is the quantile regression estimate from a specification where the coefficient for the risky share residual is constrained to be zero. The curve drawn in Figure A4(c) is less smooth than in (a) and (b). Sill, the impact is obviously larger at the top than at the bottom of the wage distribution.
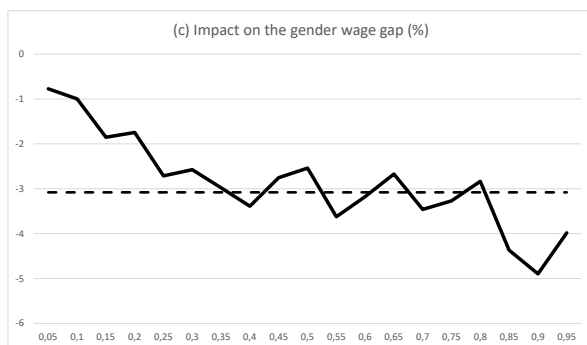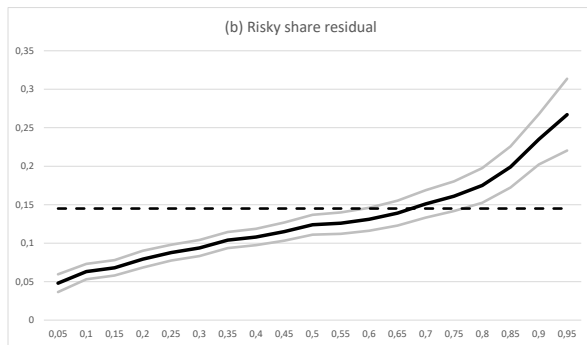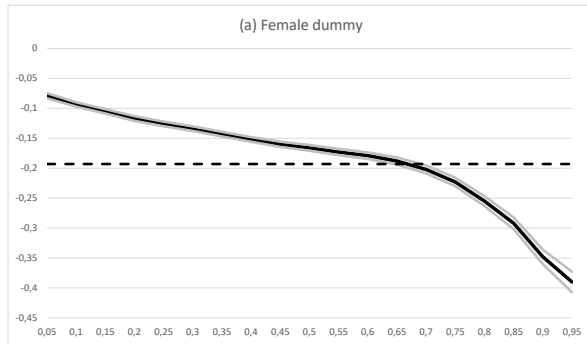
Figure A4: Quantile estimates. The dashed horisontal line represents the OLS estimate.