5-2015

Three Essays on Income and Wealth Inequality

Damir Cosic
Graduate Center, City University of New York

Three Essays on Income and Wealth Inequality

by

Damir Cosic

A dissertation submitted to the Graduate Faculty in Economics in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York.

2015

This manuscript has been read and accepted for the Graduate Faculty in Economics in satisfaction of the dissertation requirements for the degree of Doctor of Philosophy.

Thom Thurston

Date

Chair of Examining Committee

Merih Uctum

Date

Executive Officer

Thom Thurston

Wim Vijverberg

Yochanan Shachmurove

Supervisory Committee

THE CITY UNIVERSITY OF NEW YORK

Abstract

Three Essays on Income and Wealth Inequality

by

Damir Cosic

Advisor: Thom Thurston

This dissertation consists of three essays on income and wealth inequality. The essays examine various aspects of this complex feature of the economic system.

The first essay shows that the distribution of firm sizes in an economy is an important determinant of wage distribution. I use data from the U.S. Current Population Survey and the ExecuComp between 1992 and 2012 to construct a new dataset and estimate wage distribution and various measures of wage inequality. I decompose differences in wage inequality across firm sizes and over time by using semi-parametric methods. In 1992, wages were distributed more unequally in small than in large firms. A decomposition shows that this was solely due to inequality among workers with the same observed characteristics, i.e. residual inequality. Inequality due to the distribution of observed characteristics and returns to those characteristics was higher in large firms at that time. By 2012, inequality in small firms grew further, but not as fast as in large firms. Over the same period, employment share of large firms increased but this had little effect on changes in overall wage distribution.

The second essay proposes a general equilibrium model suitable for studying wealth distribution. One of the challenges in modeling wealth distribution is reproducing the high concentration of wealth observed in the U.S. I build upon the benchmark model developed by Krusell and Smith (1998) by introducing firm heterogeneity and managerial class. I develop a model, propose a numerical method for solving it, simulate it, and compare the results with the data and the benchmark model. The simulated distribution fits the data well in the upper tail of the distribution.

The third essay explores the role that changes in income inequality may play in households balance sheets. This is particularly important considering the high levels of household debt that was a key contributing factor in the financial crisis of 2008. I use a family of error-correction models to estimate long-run relationship between income inequality and household debt. Tests based on Westerlund (2007), Pedroni (1999, 2004) and Johansen (1988) are used on a panel of 14 developed countries. No evidence of cointegration between the time series for income inequality and household debt is found.

Acknowledgements

I would like to thank my dissertation advisor, Prof. Thurston, who has been a continuous presence throughout the last nine years of my studies in economics. His macroeconomics course that I took while in the Hunter College master's program cemented my decision to become a macroeconomist. As the executive officer of the program at Graduate Center, he accepted me into the program. And as my advisor he offered invaluable input and support throughout my work on this dissertation.

Equally continuous presence in my graduate studies has been Prof. Vijverberg. Even though our research interests have not always overlapped, he has always been available and has provided thoughtful advice. I am extremely grateful for that.

I also extend my deepest gratitude to Prof. Uctum, who as the executive officer of the program has always provided nothing but encouragement and support; Prof. Vachadze, who taught one of the most interesting courses I took and has remained one of my biggest supporters; Prof. Shachmurove, who invited me to present my job market paper at City College and provided extremely valuable feedback.

I would also like to thank Prof. Filer who ushered me into the master's program at Hunter College, which marked the beginning of my career in economics, and to Prof. Deb who advised me in writing my master's thesis and taught me my first steps in economic research by offering me an opportunity to collaborate with him.

Finally, I thank my parents and my sister Dijana for their love and support throughout this ordeal. I also thank my girlfriend Letisha for her love and understanding, and also for her masterful editing skills.

Contents

List of Figures

List of Tables

Chapter 1

Wage Inequality and Firm Sizes

1.1 Introduction

A large increase in wage inequality observed in the U.S. during the 1980's and 1990's has generated a substantial literature on wage distribution. Most of this research focused on changes in labor supply (composition of workforce with respect to education and experience) and demand (due to technological progress), and changes in labor market institutions (e.g. rate of unionization and regulatory framework).[1] However, a large fraction of the increase in wage inequality, the so-called residual inequality, i.e. inequality among workers with same observed characteristics, remains unexplained. Some of the early studies, such as Juhn, Murphy, and Pierce (1993) and Katz and Autor (1999) estimate that the residual inequality accounts for more than a half of the overall increase in wage inequality. Lemieux (2006) and Melly (2005) find more significant effects of changes in composition of the workforce and, with improved methodology and accounting for measurement errors, they reduce the increase in the residual inequality to around still substantial 20% of the total increase in inequality.

One factor that has been largely absent from the literature on wage inequality is the relationship between firm size and wage. It is well established that, on average, large employers pay higher wages than small employers. Brown, Hamilton, and Medoff (1990) find that large firms in the U.S. in the 1980s paid significantly higher wages than small firms for the same position, occupation and observed characteristics of employees. In addition to higher wages, large firms provided better working conditions and more generous benefits. Multiple explanations of this firm size-wage premium have been proposed, such as more advanced technology, higher workers' productivity, lower costs of other inputs, more efficient

[1] Studies that examine the effects of increased trade volume on wage inequality, such as Borjas et al. (1997), find them relatively insignificant.

monitoring, but no conclusive evidence has been produced. The literature on wage and firm size (for a survey see Oi and Idson, 1999), however, falls short of discussing the broader issue of wage inequality in the context of firm size.

This study aims to fill that void. I estimate wage distributions in small and large firms in 1992 and 2012, and decompose the differences between the two into effects of observed characteristics, returns to observed characteristics and residual inequality. I find that wages have been distributed more unequally among workers employed by small firms than among workers employed by large firms. The gap in wage inequality between small and large firms was particularly pronounced in 1992, the beginning of the observed period, when Gini coefficient of wages in firms with less than one hundred employees was 0.399. At the same time, in firms with more than one thousand employees the Gini coefficient was lower by five points at 0.343. For comparison, Gini of overall pre-tax income in the U.S. increased by five points from 0.436 to 0.486 between 1984 and 2005, two decades that saw a rapid growth in income inequality.[2] Percentile ratios reveal that this difference in wage inequality between small and large firms mostly originated in the upper half of the wage distribution. Over the next two decades, overall wage inequality increased, but it grew at a higher rate in large firms. By 2012, the end of the observed period, Gini was 0.44 in small and 0.415 in large firms.

These findings have important implications for understanding the forces that shape the wage distribution in an economy. The literature on wage inequality attributes most of its recent growth to a change in distribution of individual characteristics of workers and a change in rewards to those characteristics. Education figures most prominently in these estimates. Higher demand for skilled labor raised its share in the labor force, as well as education

[2] Source: OECD.Stat, http://stats.oecd.org.

premium. However, this is not a dominant factor in the difference in wage inequality between small and large firms. In fact, large firms have a more educated labor force and pay a higher education premium. Consequently, they have higher wage inequality due to observed characteristics. It is solely due to a higher residual inequality that small firms have higher overall wage inequality than large firms.

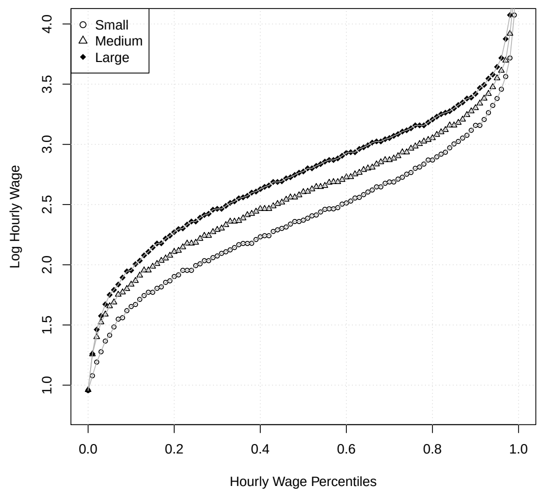
A more educated labor force and higher returns to education in large firms are consistent with large firms using more advanced technology, a hypothesis offered by Idson and Oi (1999), and complementarity of capital and high-skilled labor, a view advanced by Krusell et al. (2000) in their study of an increase in wage inequality in the U.S. over time. Due to more advanced technology, large firms have higher demand for skilled labor because of capital-skill complementarity. Skilled labor is more productive in large than in small firms implying higher returns to education and, therefore, higher wage inequality due to observed characteristics.

On the other hand, higher residual inequality in small firms is at odds with the literature. Even though it is not discussed explicitly, hierarchical models of the firm, as well as some empirical studies at the firm level imply higher residual inequality in large firms. Williamson (1967) and Rosen (1982) define the firm as a hierarchical organization where more competent managers occupy higher levels of the hierarchy, have higher marginal productivity and consequently earn higher wages than less competent ones or production workers. Hierarchies in large firms are more extended than in small firms and require more competent managers to populate top levels, both of which should generate a more unequal distribution in large firms. Furthermore, considering that managerial competence is not directly observed by researchers, it should show up through residual inequality. This view of the firm is confirmed in an empirical study by Baker, Gibbs, and Holmstrom (1994a) who find that around 70 percent of the variance in wages across employees in a large U.K. firm can be attributed to hierarchical levels they occupy. Gabaix and Landier (2008) show that compensation of chief executive officers (CEO) is positively correlated with firm size. Both of these findings imply greater residual inequality in larger firms.

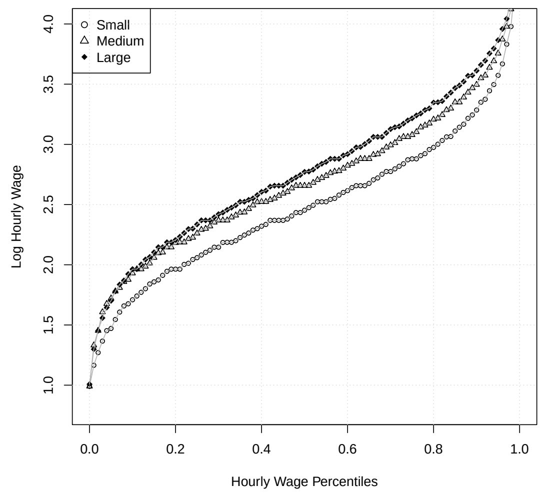
Over the observed period of time, between 1992 and 2012, distributions of wages in small and large firms converged. Figure 1.1 makes this more apparent. The top row shows empirical distributions in 1992 and 2012 for small, medium and large firms. Consistent with findings in the literature, the distribution in large firms dominates the one in medium firms, which dominates the one in small firms, but overall, the three distributions are much closer in 2012 than they were in 1992. The firm size-wage premium shrank, and so did the difference in inequalities. But, even though overall inequality increased in both small and large firms, the increase was the most dramatic in large firms. The extent of this change is more evident from the bottom row of Figure 1.1, which shows changes in log real hourly wages between 1992 and 2012 for each percentile. Because increases at the top are of an entirely different magnitude than the changes in the rest of the distribution, the bottom-right panel shows changes for only the top five percent of wage distribution. There are two things that separate wage dynamics in large firms: first, wages in the bottom of the distribution fell or remained the same, while in other firms they increased almost uniformly by around ten percent; second, wages at the very top (top 0.1%) increased by 300%, more than double of the equivalent increase in medium and small firms.

Figure 1.1: Percentiles of log real hourly wage in the U.S. for small, medium, and large firms in 1992 and 2012

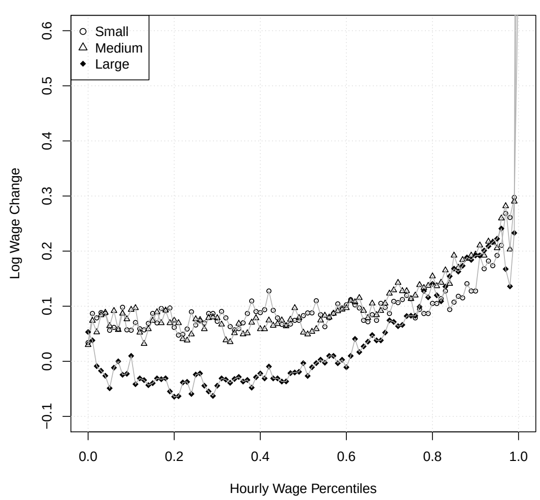
(a) Percentiles, 1992



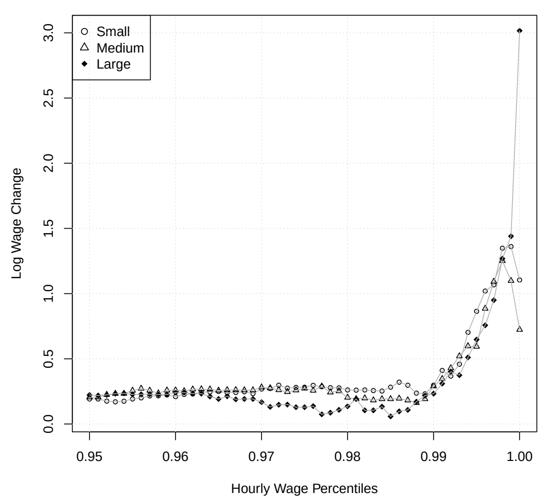
(b) Percentiles, 2012



(c) Changes between 1992 and 2012 by percentile, entire distribution



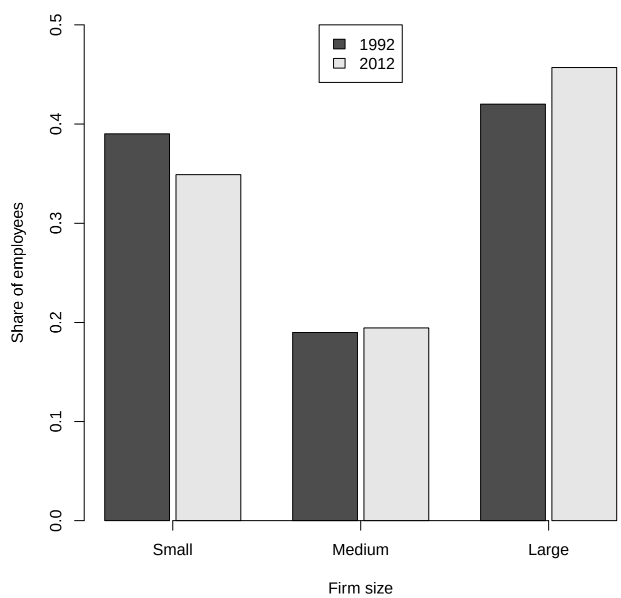
(d) Changes between 1992 and 2012 by percentile, top five percent



Source: March CPS and ExecuComp for 1992 and 2012, full-time, full-year workers ages 16 to 64 whose longest job in the observed year was in private sector. Full-time, full-year workers are those who usually worked 35-plus hours per week and worked forty plus weeks in the previous year. Hourly earnings are calculated as annual earnings divided by weeks worked and usual number of hours per week. Firms with 1-99 employees are defined as small, those with 100-999 employees as medium and those with 1000 and more employees as large. Wages are deflated using the consumer price index (CPI). Observations with earnings of below one half of minimum real wage and those with allocated earnings are dropped.

Another change that occurred between 1992 and 2012 is an increase in employment share in large firms. Figure 1.2 shows that the share of employees working in firms with over 1,000 employees increased from 42% in 1992 to 45.7 % in 2012, while the share of employment in small firms dropped from 39% to less than 35%. Considering that inequality in large firms is lower, an increase in their employment share should slow down the increase in overall wage inequality. A decomposition of the changes in overall inequality shows this indeed to be the case, although this effect was relatively small.

Figure 1.2: Employment shares in the U.S. for small, medium and large firms.



Source: Business Dynamics Statistics (BDS) released in September 2014 by the U.S. Census Bureau. Small firms employ 1-99 employees, medium firms 100-999, and large firms more than 1,000 employees.

The main source of data in this study (and virtually all other studies of wage distribution in the U.S.) is the Current Population Survey (CPS). Even though its May Supplement has been preferred in the literature since Lemieux (2006) reported a problem with estimation of hourly wages in the CPS March Supplement, I use the latter because it is the only source that contains information on firm size. Because income in CPS data is top-coded[3] and top incomes are undersampled, it is bound to underestimate the upper tail of the wage distribution. To mitigate this drawback, I augment the CPS data with the data from the database on executive compensation, ExecuComp, and the associated database of information about firms, CompuStat. Because the ExecuComp starts in 1992, I focus on the period between 1992 and 2012. Merging with CPS is not without challenges, as the ExecuComp sample covers only a part of the population in the upper tail of income distribution. To verify validity of the combined dataset, I compare the wage distribution in that sample with the distribution estimated by Piketty and Saez (2003) from the U.S. tax data.[4]

The next section describes the data and the construction of the new dataset. Section 1.3 presents key facts about wage distribution and factors that affect it. Section 1.4 describes the econometric methods. Section 1.5 introduces results. The last section concludes.

1.2 Data

This section describes the data sources, the CPS[5] and the ExecuComp, the process of merging them, and it evaluates the resulting data set. The dataset covers the time period between 1992 and 2012.[6]

I calculate the main variable of interest, real hourly wage, by dividing the annual wage income by the number of hours worked in that year and deflating it by the Consumer Price Index (CPI).[7] In the CPS sample, I use the total income from the longest-held job in the

[3]Because of privacy concerns, the U.S. Census Bureau does not show actual values of high incomes, but assigns to them some maximum value, i.e. top code. For more on this procedure see the following section.

[4]The latest update of their data was made public in 2013 on Saez's website

[5]The CPS sample is obtained from King et al. (2010).

[6]Because information about the total income received in a given year is included only in the following year of the CPS survey, the CPS samples are from 1993 to 2013.

[7]Specifically, I use the multiplier CPI99 provided by King et al. (2010).

previous year, the number of weeks worked on that job and the usual number of weekly hours as reported by the participants in the survey. In the ExecuComp sample, weeks and hours are not reported, so I use 49 weeks a year and 50 hours per week, the average numbers of weeks per year and hours per week, respectively, reported by Birley and Norburn (1987) in their survey of Fortune 500 executives.

The main shortcoming of the CPS data with respect to estimation of income distribution is its use of top-codes for very high incomes. This is done to prevent identification of individuals with extremely high incomes, but it also prevents estimation of the upper tail of the income distribution. The top-coding values and procedures have changed over time. Between 1992 and 1995, the top-code was $99,999. All incomes above that value were set to $99,999. In addition to the effect that this procedure has on the income distribution in the sample, it also misrepresents the aggregate income. Researchers commonly deal with the latter problem by multiplying top-coded incomes by a multiplier, which is usually derived from aggregate data on labor income (e.g. national income accounts). For example, Autor, Katz, and Kearney (2008) use 1.5 as the multiplier.

The top-coded income increased to $150,000 in 1996, and then again to $200,000 in 2003. Additionally, a new procedure was introduced in 1996. Rather than using the top-code as the replacement value, individuals with top-coded incomes are classified into dozen groups according to their individual characteristics (gender, race, full-time status). For each group, the average income is calculated and assigned to each individual in the group. This procedure preserves the aggregate income and requires no adjustment in that regard. However, it still hides the actual values of top incomes and thus introduces bias into estimates of the upper tail of the income distribution.

Another potential problem with estimating wage distributions from CPS is that CPS probably undersamples very high wages. Not only is the wage distribution highly skewed, but top earners also tend to be geographically concentrated, and area-based sampling used in CPS is likely to underestimate top wages. A similar problem that exists in the Survey of Consumer Finances (SCF) with respect to the distribution of wealth is addressed by an additional sample that samples the top five percent of the distribution at a higher rate. CPS, however, makes no such an adjustment. To address undersampling and top-coding I use the ExecuComp database in a similar way that SCF uses their additional sample.

There are a number of challenges to combining the ExecuComp with the CPS; the most critical is the limited coverage of the U.S. firms by the ExecuComp database. The ExecuComp contains information on top executives from firms whose stocks are included in the Standard & Poor's 1500. While this sample provides a relatively good coverage of publicly traded firms,[8] it contains no information about privately held firms, which make a majority of the U.S. firms.

The exclusion of privately held firms from the ExecuComp results in a sample that is heavily biased towards large firms. Because majority of small firms are privately held, they are severely underrepresented in the sample. For example, in 2012 the employment share in small firms (1-99 employees) was around 35% according to the Business Dynamics Statistics (BDS), while its estimate from the ExecuComp sample is below one percent. This problem is addressed by adjusting the weights in the sample. This procedure is described below.

Another consequence of the exclusion of privately held firms from the sample is a potential bias due to different distributions of executive compensations in privately held and publicly

[8]The S&P 1500 fact sheet states that the index covers around 90% of the U.S. market capitalization (http://us.spindices.com/indices/equity/sp-composite-1500).

traded firms. Because of lack of data on compensation in privately held companies, this bias cannot be easily addressed by weighting. There are, however, some reasons for moderate optimism. One of the few studies that compares CEO compensation in privately held and publicly traded companies is by Ke, Petroni, and Safieddine (1999). They focus on the insurance industry and find a relatively small difference between the two. The main source of the difference is the fact that compensation of CEOs of publicly traded insurance companies is based more on company's performance than compensation of COEs of privately held insurance companies.[9]

I use a subset of the CPS dataset that includes all persons who were employed full time (more than 35 hours per week) and full year (more than 40 weeks) in the previous year, and who were between 16 and 64 years of age. The sample does not include government employees and unpaid family workers. This represents around 44% of the total civilian population of the above age who was employed for any amount of time in 1992, and around 48% in 2012.

The CPS provides the number of employees in a firm as a categorical variable that represents one of several bins. Because the sizes of the bins changed at some point between 1992 and 2012, I reduce their number to make the sample information compatible over time. I define three firm sizes: small, with 1-99 employees, medium, with 100-999 employees, and large with more than 1,000 employees.

Educational attainment is also represented by a categorical variable. I convert it to five dummy variables that represent the following categories: no high-school, high-school diploma, one to three years of college, four to five years of college and a post-graduate

[9]Specifically, they use net income over total assets (ROA) as a measure of performance, which, when interacted with a dummy for public ownership, has a significant positive coefficient. The dummy for public ownership itself is statistically insignificant.

degree. Experience is calculated as age minus years of education minus seven, and divided into four categories: 0-9, 10-19, 20-29, and more than 30 years.

1.2.1 Merging the ExecuComp and the CPS

The ExecuComp contains exhaustive information about executives and their compensation, but little information about firms. To obtain information about firm size and industry I merge the ExecuComp and the CompuStat databases. Some observations in the latter are missing firm size information. If a firm is missing firm size for all years in the sample, it is dropped. This is true for a small number of observations, typically between 0.1% and 0.4% of the sample for any given year. If a firm is missing firm size values for only some years, missing values are linearly interpolated between years, and extrapolated on both ends by using the value of the earliest or the latest non-missing observation, respectively.

The weights are calculated in two steps: first, weights in each sample separately; then, combined weights. No intervention is required in the first step for the CPS sample; I use the individual survey weights. The following two subsections describe the calculation of the first-stage weights in the ExecuComp sample, and the calculation of the combined weights.

The ExecuComp Weights

The ExecuComp provides information about up to ten executives from a single firm. For each firm k included in the dataset, there is one observation per year for each of its n_k executives whose information is present in the dataset, where $1 \leq n_k \leq 10$. Executives from the same firm are ranked by their compensation in each year. The number of executives of rank r in all firms of size i in the sample is n_{ri}. There are m_i firms of size i in the sample and M_i of them in the population. The latter is obtained from BDS.

I calculate the ExecuComp weight of an observation as the inverse of the probability that the observation is selected by the sampling procedure. The probability that an executive of rank r from a firm of size i is selected, $P(r, i)$, equals the product of the probability that a firm of size i is selected, $P(i)$, and the conditional probability, $P(r|i)$, that an executive of rank r is selected given the firm size is i:

$$P(r, i) = P(i)\, P(r|i)$$

Assuming random sampling among all firms[10], probabilities $P(i)$ and $P(r|i)$ can be calculated as follows:

$$P(i) = \frac{m_i}{M_i} \text{ and } P(r|i) = \frac{n_{ri}}{m_i}$$

Thus, the probability that an executive of rank r from firm of size i is selected is:

$$P(r, i) = \frac{n_{ri}}{M_i}$$

and, subsequently, the weight of such an observation is:

$$w_{r,i} = \frac{M_i}{n_{ri}}$$

Combined Weights

To calculate combined weights for the merged sample, I follow Kennickell, Woodburn, and McManus (1996) who describe calculation of weights for the Survey of Consumer Finances (SCF). The SCF consists of two samples: the area-probability (AP) sample, which covers the overall population, and the list sample, whose purpose is to oversample wealthy households. Because the two samples are using different sampling procedures, the weights need to be

[10] This is clearly not the case, as the ExecuComp database contains only publicly traded firms, but no better alternative is available.

adjusted to ensure that the combined sample is still representative of the population.

The weights are usually adjusted in reference to a set of demographic variables so that the distribution of these variables in the combined set reflects their distribution in the population. Demographic information in the ExecuComp dataset, however, is of relatively low quality. There is no information on education, and data on gender and age are missing in many observations. Consequently, I use firm size as a reference, similar to the calculation of the first-stage ExecuComp weights.

The formula for the combined weight of an observation j that represents an employee in a firm of size i is:

$$w_j = R_{i_{CPS}} w_{i_{CPS}} + R_{i_{EC}} w_{i_{EC}}$$

where CPS and EC subscripts refer to the CPS and ExecuComp samples, w_{i_s} is the weight for an employee in a firm of size i in sample s. The term R_{i_s} is calculated by the following formula:

$$R_{i_s} = \frac{n_{i_s}/N_{i_{CPS}}}{n_{i_{CPS}}/N_{i_{CPS}} + n_{i_{EC}}/N_{i_{EC}}}$$

where n_{i_s} and N_{i_s} are the non-weighted and weighted, respectively, numbers of observations of firms of size i in sample s.

To take into account variability in the amount of labor actually supplied, this combined weight for each individual is multiplied by the number of weeks worked last year and the usual number of hours per week as reported by the participants in the survey.

To validate the combined dataset, I compare it with the CPS along two dimensions: firm size and industry. For each year, I calculate the frequencies in these two variables in both, the original CPS and the combined dataset. The test reports no significant discrepancies.[11]

As another validation of the combined dataset, I estimate the wage distribution and compare its bottom part to the CPS-based estimates and its upper tail with the estimates obtained by Piketty and Saez (2003).[12] The goal is to have a big part of the bottom of the wage distribution estimated from the combined dataset the same as the one estimated from the CPS. Ideally, only in the upper tail there should be divergence from the CPS and convergence toward estimates by Piketty and Saez (2003).

Table 1.1 shows estimates of wage income shares for bottom 10%, 25%, 50% and 75% in years 1992, 2002 and 2012 estimated from the CPS and the combined dataset. The shares estimated from the combined dataset do not differ much from those estimated only from the CPS, which indicates that the merging process did not significantly affect the bottom of the wage distribution in the CPS sample. This is one of the objectives because one of the assumptions is that the CPS represents the bottom of wage distribution well.

Table 1.1: Wage shares in percentages for bottom income groups from the CPS and the combined CPS and ExecuComp datasets.

Year	Data source	P0-10	P0-25	P0-50	P0-75
1992	CPS	2.6	9.0	25.0	49.1
	CPS+ExecuComp	2.5	8.9	24.6	49.1
2002	CPS	2.5	8.6	23.4	45.4
	CPS+ExecuComp	2.4	8.5	23.0	45.4
2012	CPS	2.7	8.4	22.8	45.8
	CPS+ExecuComp	2.6	8.1	22.0	44.4

To see how much, if any, improvement is achieved by combining the CPS and the ExecuComp, I compare its top income shares to those estimated by Piketty and Saez (2003). Table 1.2 shows estimates of top income shares (top 10%, top 1% and top 0.1%) in 1992, 2002 and

[11]I use R package for sample reweighing, "anesrake", for this purpose.

[12]It is worth noting that the universe covered by Piketty and Saez (2003) differs somewhat in that it includes government employees. Their estimates, however, remain the best available benchmark.

2012 estimated from the CPS, the combined dataset and those estimated by Piketty and Saez (2003). Overall, the estimates from the combined set for all three years represent an improvement relative to the CPS estimates. They are the farthest from the estimates by Piketty and Saez (2003) for 1992. This is most likely due to the low coverage of the ExecuComp sample in that year. This is the first year of the database and it contains significantly fewer observations than later years. The estimates for 2002 and 2012 are remarkably close to those obtained by Piketty and Saez (2003).

Table 1.2: Wage shares in percentages for top income groups from the CPS, the combined CPS and ExecuComp, and Piketty and Saez (2003), whose latest update from 2013 contains data up to 2011.

Year	Data source	P90-100	P99-100	P99.9-100
1992	CPS	28.3	7.4	0.8
	CPS+ExecuComp	29.2	8.7	1.8
	Piketty & Saez	32.5	9.6	3.3
2002	CPS	32.4	8.5	1.2
	CPS+ExecuComp	33.9	10.3	2.5
	Piketty & Saez	33.4	10.3	3.8
2012	CPS	32.3	9.6	1.9
	CPS+ExecuComp	34.7	12.3	4.1
	Piketty & Saez	34.9	11.0	4.0

1.3 Key Facts

Table 1.3 shows Gini coefficients of wage distributions in 1992 and 2012. Panel (a) shows Gini coefficients for three different firm sizes and for all firms together. Panel (b) shows a decomposition of Gini into between, within and overlap components. Both panels show values of Gini estimated from the CPS alone and from the dataset that combines the CPS and the ExecuComp data.

It is immediately clear from panel (a) that wages in small firms are distributed more unequally than wages in medium and large firms. This is the case in both observed years and

Table 1.3: Gini coefficient for wages in 1992 and 2012 decomposed by firm sizes and Gini components. Source: the CPS March supplement and the ExecuComp.

(a) Decomposition by firm sizes

Firm size	CPS		CPS+ExecuComp	
	1992	2012	1992	2012
Small	0.394	0.423	0.399	0.440
Medium	0.339	0.381	0.354	0.415
Large	0.335	0.401	0.343	0.415
All	0.370	0.412	0.378	0.431

(b) Decomposition by Gini components

Component	CPS		CPS+ExecuComp	
	1992	2012	1992	2012
Between	0.046	0.048	0.048	0.047
Within	0.140	0.155	0.142	0.161
Overlap	0.184	0.209	0.188	0.223

regardless of the dataset used for estimation. This is somewhat unexpected. In theoretical models of the firm, like those proposed by Williamson (1967) and Rosen (1982), a deeper hierarchy, which is associated with a larger firm, generates a more unequal wage distribution. Furthermore, the finding by Gabaix and Landier (2008) that CEO compensation is positively correlated with the firm size also points in the direction of greater inequality in larger firms.

In the case of Gini estimated only from the CPS data, one reason for this result could be a bias resulting from the top censoring of the CPS earning data. If large firms had significantly higher wages in the top tail of the distribution, their Gini would be affected disproportionally by top censoring. However, considering the Gini estimates based on the combined data, it is hard to find support for this argument. Combining the CPS and the ExecuComp revises upward all estimates by similar amounts and Gini coefficients for small firms remain significantly higher. Furthermore, estimates of percentile ratios in Table 1.4 show the same relationship between inequality and firm size, even though this measure of inequality excludes the top tail of wage distribution.

The distribution of wages became more unequal between 1992 and 2012 in all firms. The overall Gini increased by five points. The increase in inequality was the biggest in large firms, whose Gini went up by seven points. It increased by six points in medium-size firms and by four points in small firms.

The bottom panel of Table 1.3 also demonstrates that most of the increase in inequality occurred within each firm size category, while the spread of the mean wages for the three firm sizes remained the same or even slightly decreased. The three rows of the panel show the between, within, and overlap component of the Gini decomposition[13]. The between component is the Gini coefficient that would prevail if each employee was assigned the mean wage in his or her firm size category. The within component is a weighted average of Gini coefficients for each category weighted by the product of its employment and income shares. The overlap component is a residual that takes into account overlapping parts of wage distributions for firm size categories.

Table 1.4: Wage percentile ratios: 90-10, 90-50, and 50-10 for small, medium and large firms. Source: CPS March and ExecuComp.

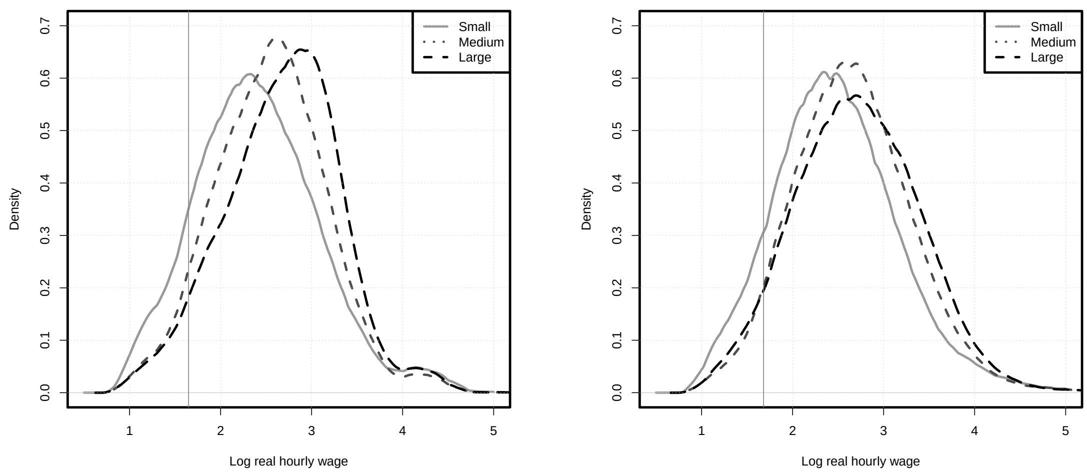
Year	Firm size	90-10	90-50	50-10
1992	Small	1.715	0.904	0.811
	Medium	1.526	0.738	0.788
	Large	1.578	0.682	0.896
	All	1.671	0.833	0.838
2012	Small	1.702	0.916	0.786
	Medium	1.631	0.833	0.799
	Large	1.719	0.879	0.840
	All	1.727	0.916	0.811

It is useful to look again at percentile ratios in Table 1.4 because, unlike Gini, they show

[13]The decomposition has the form: $G = G_B + \sum a_k G_k + G_O$, where G_B is the between component, a_k is the product of population share and income share for group k, G_k is the Gini coefficient for group k, and G_O is the residual due to overlap of distributions of the groups. (see Lambert and Aronson, 1993)

where the action is. The 90/50 percentile ratio indicates that inequality of wages in the top-half of the distribution increased between 1992 and 2012 in firms of all sizes. This was particularly pronounced in medium and large firms. In contrast, the 50/10 percentile ratio shows that inequality in the bottom half of the distribution decreased in small and large firms, and only modestly increased in medium firms.

Figure 1.3: Distribution densities of log real wages in 1992 and 2012 for small, medium and large firms.



Densities are normalized to sum to one. Solid line represents wage distribution for all firms. Vertical line represents the log of real minimum wage. Source: March CPS combined with ExecuComp.

Plots of wage distribution estimates provide additional insight into changes in wage distribution. Figure 1.3 shows estimates of overall wage distribution density, and its decomposition to densities for small, medium and large firms. The estimates for the three firm sizes are weighted so that they add up to the overall kernel density estimate. It can be seen in the left panel that in 1992 wage distribution for small firms had thicker tails and wider spread than distributions for medium and large firms. The mass of the distribution for large firms is shifted to the right, which is consistent with the findings of other studies that wages in large firms were higher than those in small firms for any level of employees' education, experience and other individual characteristics. Distributions for 2012 shown in the right panel are more closely aligned. While the distribution for small firms still has a thicker lower tail, its spread is not much wider than the spread of the distribution for large firms, which extended toward lower wages.

In summary, wage inequality increased overall in the U.S. firms between 1992 and 2012. This increase was mostly concentrated in the top half of the distribution. Wages have been distributed more unequally in small than in large firms, but this gap decreased over the observed period of time.

The rest of this section discusses some of the key determinants of wage, such as individual characteristics of employees and industry employment shares. Table 1.5 shows sample means of log real wage, key demographics variables and industry dummies. The sample mean of each variable is shown for 1992 and 2012, for all firms and separately for small, medium and large firms.

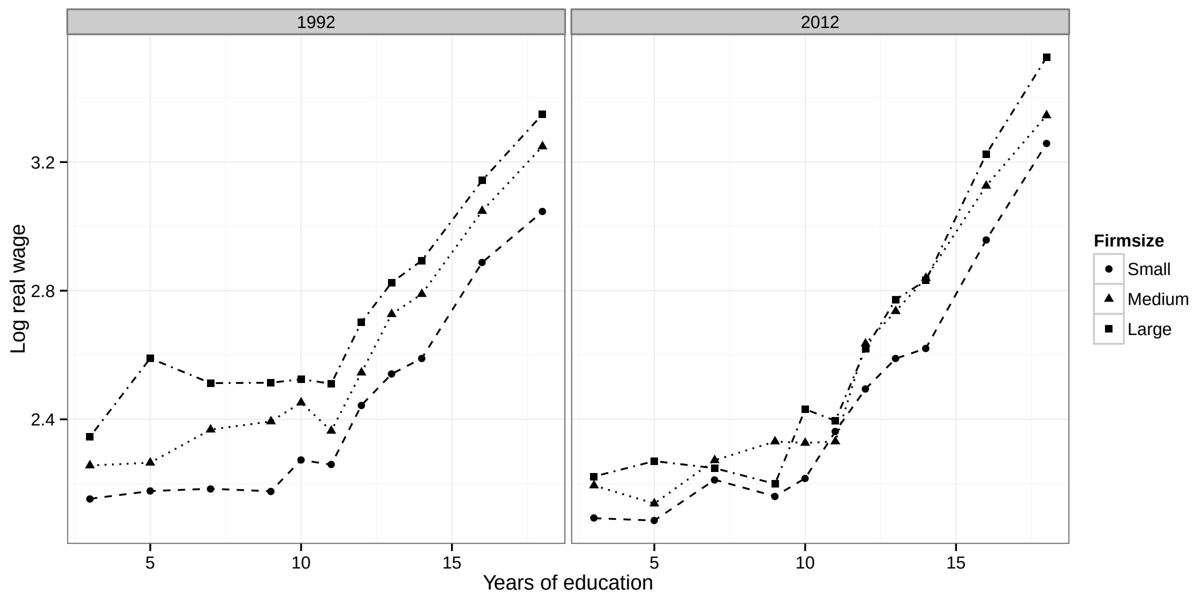
The table shows that larger firms had more educated labor force in both years. All levels of educational attainment above high school diploma have higher share in larger firms. Over time, the average educational attainment of employees in all firms increased. For example, percentage of employees with four to five years of college increased from 15 percent in 1992 to 21.5 percent in 2012. Except for post-graduate education, however, the gap between small and large firms did not change much.

In addition to the change in composition of labor force with respect to education, the wage premium to education increased between 1992 and 2012. Figure 1.4 shows log of average real wage conditional on years of education for the three firm sizes. The left and right charts show data for 1992 and 2012, respectively. A striking feature of these two charts is the fall in real wages in large firms for all but college-educated employees. While in 1992 wages in

Table 1.5: Sample means for 1992 and 2012. Source: March CPS.

	1992				2012			
		Firm size				Firm size		
Variable	Total	Small	Medium	Large	Total	Small	Medium	Large
Log real wage	2.446	2.342	2.479	2.579	2.547	2.458	2.604	2.635
Female	0.445	0.421	0.473	0.464	0.458	0.437	0.474	0.477
White	0.865	0.885	0.852	0.845	0.802	0.828	0.793	0.772
Education								
No high school	0.132	0.159	0.127	0.095	0.088	0.115	0.076	0.059
High school	0.373	0.378	0.373	0.364	0.293	0.315	0.286	0.267
College, 1-3 yr	0.283	0.271	0.288	0.299	0.309	0.298	0.309	0.325
College, 4-5 yr	0.150	0.133	0.151	0.175	0.215	0.189	0.231	0.241
Post-college	0.062	0.060	0.061	0.066	0.095	0.083	0.099	0.109
Experience								
1-9 yr	0.287	0.281	0.281	0.300	0.260	0.241	0.247	0.291
10-19 yr	0.295	0.289	0.312	0.291	0.220	0.214	0.224	0.226
20-29 yr	0.225	0.227	0.220	0.225	0.227	0.233	0.233	0.217
30+ yr	0.194	0.203	0.187	0.184	0.294	0.313	0.297	0.267
Industry								
Agriculture	0.028	0.052	0.011	0.004	0.027	0.049	0.015	0.003
Mining	0.007	0.004	0.008	0.010	0.009	0.006	0.015	0.011
Construction	0.068	0.114	0.044	0.016	0.074	0.124	0.056	0.019
Durables mfg.	0.118	0.067	0.169	0.161	0.073	0.048	0.112	0.088
Non-durables mfg.	0.082	0.048	0.122	0.106	0.048	0.032	0.073	0.058
Transportation	0.041	0.034	0.035	0.056	0.036	0.034	0.037	0.038
Utilities	0.023	0.005	0.016	0.053	0.019	0.011	0.016	0.031
Trade, wholesale	0.043	0.050	0.050	0.030	0.027	0.027	0.033	0.026
Trade, retail	0.207	0.204	0.137	0.254	0.207	0.182	0.124	0.282
Finance	0.057	0.033	0.060	0.090	0.057	0.032	0.053	0.092
Real estate	0.018	0.025	0.017	0.007	0.018	0.026	0.017	0.009
Business services	0.059	0.074	0.056	0.037	0.092	0.098	0.095	0.083
Personal services	0.060	0.090	0.030	0.033	0.053	0.081	0.028	0.028
Information	0.004	0.003	0.006	0.004	0.004	0.001	0.004	0.009
Entertainment	0.017	0.023	0.015	0.010	0.027	0.031	0.028	0.021
Health	0.049	0.068	0.056	0.016	0.075	0.083	0.096	0.053
Hospitals	0.040	0.004	0.083	0.065	0.041	0.007	0.060	0.077
Education	0.032	0.035	0.038	0.022	0.049	0.049	0.066	0.041
Professional services	0.048	0.064	0.049	0.023	0.061	0.079	0.073	0.032

Figure 1.4: Log real wage for small, medium and large firms in 1992 and 2012 by years of education. Source: March CPS.



large firms dominated wages in small and medium firms at virtually every level of education, in 2012 average wages in the three firm sizes are almost indistinguishable for all educational levels below college. On the other hand, the average wage for employees with at least some college increased in firms of all sizes.

As can be seen, there were two important shifts related to education and its effect on wages. First, there was a change in composition toward more educated labor force. Second, there was a in increase in relative rewards to education. This increase, however, did not follow the same pattern in firms of all sizes. While college premium increased in all firms, wages for employees with less than college education decreased only in large firms.

A look at distribution of experience does not reveal any distinct patterns. The shares of all four experience groups are similar across the three firm sizes. There is a shift, however, over time toward more experienced workers. While in 1992 the share of employees with over 30 years of experience is around 20 percent, in 2012 this number is closer to 30 percent.

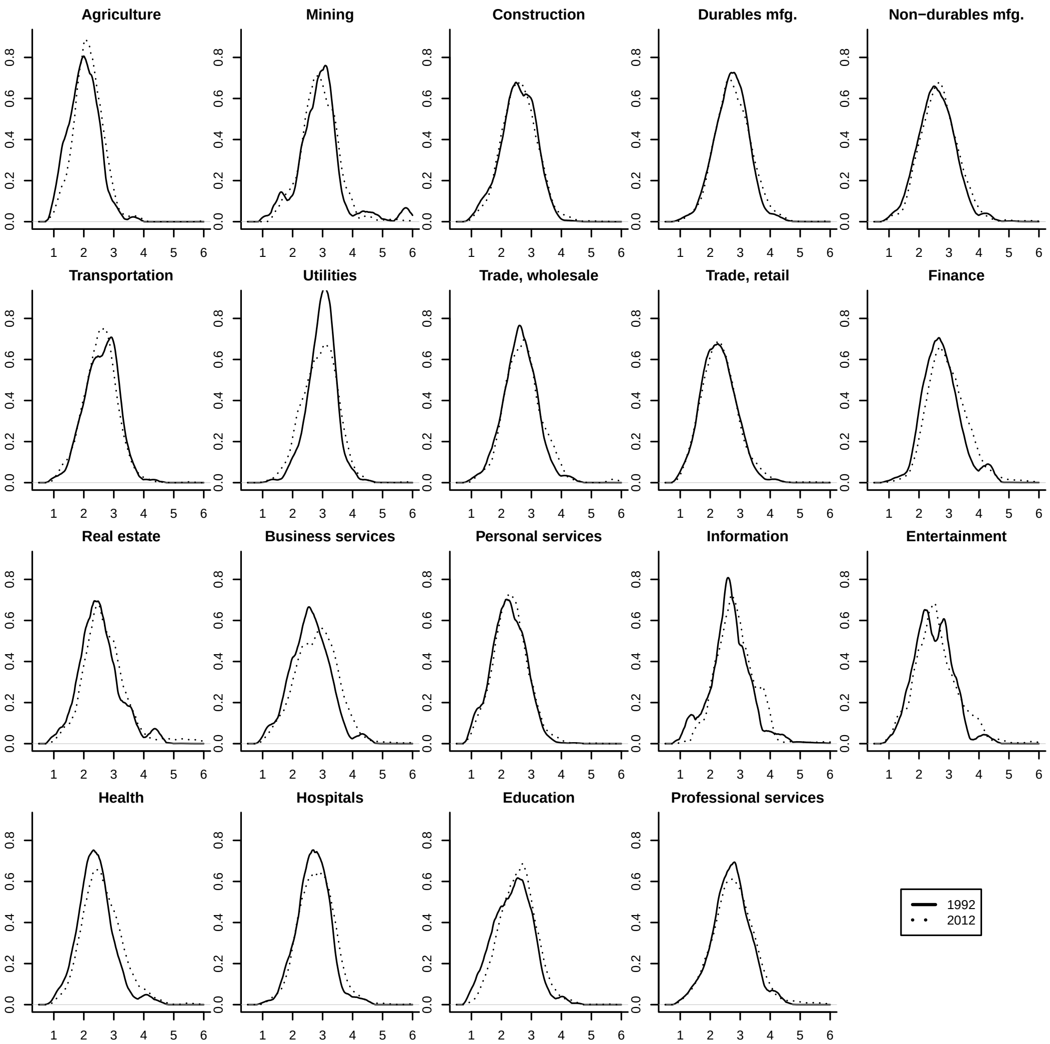
Finally, to see if these changes in wage distribution are driven by a particular industry or group of industries, Table 1.6 shows Gini coefficients for the 19 industries in 1992 and 2012, and changes over that period of time. The industries are sorted in the order of decreasing change in Gini.

Table 1.6: Gini coefficient for wages in 1992 and 2012 by industry. Source: CPS and ExecuComp.

Industry	1992	2012	Change
Real estate	0.409	0.660	0.251
Trade, wholesale	0.341	0.422	0.082
Utilities	0.256	0.336	0.080
Entertainment	0.401	0.470	0.069
Non-durables mfg.	0.362	0.421	0.059
Finance	0.387	0.434	0.047
Hospitals	0.301	0.348	0.047
Agriculture	0.363	0.409	0.046
Durables mfg.	0.320	0.363	0.043
Information	0.514	0.557	0.043
Trade, retail	0.349	0.388	0.039
Professional services	0.396	0.434	0.038
Business services	0.398	0.432	0.034
Construction	0.330	0.358	0.028
Personal services	0.333	0.352	0.020
Transportation	0.327	0.344	0.017
Health	0.443	0.445	0.003
Education	0.383	0.378	-0.005
Mining	0.495	0.433	-0.062
All firms	0.378	0.431	0.053

The only outlier is the real estate industry. Not only does it have the highest Gini in both years, but the increase in Gini in that industry is more than three times greater than the second highest increase. Changes in most other industries are within a close distance from the change in overall Gini. It is worth noting that this dramatic change in the wage distribution in the real estate industry is observed only in the data set that combines the CPS and the ExecuComp data; changes estimated from only the CPS data are relatively small.

Figure 1.5: Log real wage distribution densities in 1992 and 2012 by industry. Source: March CPS and ExecuComp.



To verify validity of the data, I estimated basic statistics such as sample mean, range, and quantiles, and compared them with the statistics of other industries. There were no significant differences that might indicate measurement errors. As an additional test, I estimated kernel densities of wage distributions across industries. Figure 1.5 shows these densities for 1992 and 2012 for all industries. Visually, there are no dramatic changes in the density wage distribution in the real estate industry. The main difference between the distributions in two years, apart from being slightly shifted relative to each other, is that the distribution from 2012 has a long upper tail. Considering its size, however, the overall trend is unlikely to be driven by the changes in the real estate industry. As it can be seen from Table 1.5, this industry employs less than two percent of all workers.

1.4 Econometric Methods

This section provides an overview of the methodology, lists identification issues and the necessary assumptions, and lays out the estimation strategy. I mostly use notation from a study by Fortin, Lemieux, and Firpo (2011) who provide an excellent survey of decomposition methods. Because most such methods are limited to binary variables, i.e. decomposition of differences between two groups, or decomposition of "treatment effect", I divide firm sizes into only two groups: small and large, where the threshold is 1000 employees.

There are two principal questions about the relationship between firm size and wage distribution that I aim to answer: i) what are the sources of difference in wage distribution between small and big firms at a point in time, and ii) how much did this difference between small and big firms contribute to overall changes in wage distribution over time.

The first question is similar to the one answered by the method for decomposition of the difference in means of a variable originally proposed by Oaxaca (1973) and Blinder (1973). The difference in means of a variable in two subsets, which may represent subsets of the population (e.g. union and non-union) or two points in time, is decomposed into a part that is due to differences in covariates and a part that is due to differences in coefficients. Depending on the specific formulation of the decomposition, it may also be possible to identify a part that is due to differences in unobservables.

In this study the variable of interest is the log of real hourly wage, the two groups are small and large firms, covariates are standard human capital characteristics, and coefficients represent returns to these characteristics. When it comes to inequality, though, rather than using the mean, we decompose summary measures of inequality, such as Gini, Theil, or percentile ratios. In general, for some such measure ν, the decomposition can be written as

$$\Delta^\nu = \Delta^\nu_S + \Delta^\nu_X + \Delta^\nu_\varepsilon$$

where the three terms on the right hand side represent differences in ν due to wage structure, observed characteristics and unobserved characteristics, respectively.

The second question adds another level of complexity. In this case, we are dealing with changes in wage distribution between two points in time, and we want to estimate the effect of one of the covariates – firm size in this case – on those changes. The decomposition of the effects of a covariate k estimates a part of change in wage distributions that is due to the change in returns to the covariate k, $\Delta^\nu_{S_k}$, and a part that is due to the change in distribution of the covariate k, $\Delta^\nu_{X_k}$. Because it seeks to isolate the effect of a single covariate on differences in wages between two groups, this type of decomposition is known in the literature as "detailed decomposition", as opposed to the type that answers the first question, which is called "aggregate decomposition".

All methods for distributional decomposition rely on estimating counterfactual distributions. For example, in the case of wage distributions in small and large firms, a counterfactual distribution would be a distribution of wages that would prevail in large firms if the workforce composition was the same as in small firms. Knowing such a counterfactual would allow a decomposition of differences in distributions into a part that is due to differences in covariates and a part due to differences in wage structures.

The definition of counterfactual distribution follows from a key relationship between conditional and marginal probabilities:

$$F(w) = \int F(w|X)dF(X)$$

where $F(w)$ and $F(X)$ are marginal cumulative distributions of wage and covariates X, respectively, and $F(w|X)$ is the conditional CDF of wage conditional on covariates. In case of two groups, A and B, the above equation can be written for each of them:

$$F_g(w) = \int F_g(w|X)dF_g(X) \tag{1.1}$$

where $g = A, B$, and $F_g(w|X)$ represents a conditional distribution of wage for individuals in group g conditional on their characteristics given by X. A counterfactual distribution can be obtained by combining the distribution of characteristics of one group with the wage structure of the other. The conditional distribution defined by

$$F_A^{(B)}(w) = \int F_B(w|X)dF_A(X) \tag{1.2}$$

represents the counterfactual wage distribution that would prevail in group A if the wage structure was the same as in group B. This interpretation is valid, however, only if we can assume that the wage structure is independent from the manipulation of the distribution of characteristics. In the case where the two groups represent small and large firms, it is reasonable to assume that the wage structure in large firms would not significantly change if workforce composition was replaced by the composition observed in small firms. This assumption has routinely been made in the literature.

The difference in some measure of wage distribution between groups A and B that is due to the difference in wage structure, Δ_S^ν, now can be calculated as:

$$\Delta_S^\nu = \nu(F_B) - \nu(F_A^{(B)})$$

It is important to note, however, that wage distribution depends on unobserved characteristics as well. To make that explicit, Fortin, Lemieux, and Firpo (2011) write the wage function as $w = m_g(X, \varepsilon)$, $g = A, B$, which makes it clear that the conditional wage distribution depends not only on $m_g(\cdot)$, but also on the distribution of unobservables:

$$F_g(w|X) = Pr\left(m_g(X, \varepsilon) < w|X\right), \quad g = A, B$$

The estimation of the counterfactual distribution requires this conditional distribution to be replaced, a procedure which involves replacement of unobservables as well. The authors show that only if the distribution of unobservables ε conditional on X is the same in both groups can we attribute the difference $\nu(F_B) - \nu(F_A^{(B)})$ solely to the difference in wage structures.

This assumption about conditional independence is also routinely made in the decomposition literature, even though in some cases it may be too strong. In the case of groups defined by firm size, it may be violated because of self-selection, as people may have preferences about the size of firms where they work, or because of selection by firms, as employers may be able to observe some employees' characteristics that are not available in the dataset. However, given the lack of a better alternative and in keeping with the literature, I assume that the conditional independence holds and verify the results by alternative estimation methods.

In the case of detailed decomposition, even more stringent assumptions need to be made. They are discussed in relation to specific methods.

1.4.1 Aggregate Decomposition of Wage Distribution

To decompose differences in wage distributions between small and large firms at a point in time, I use an approach proposed by Machado and Mata (2005) and further refined by Melly (2005). It relies on the conditional quantile regression to estimate counterfactual distributions and then decompose the differences in wage distribution into the differences due to characteristics, coefficients and residuals.

Conditional quantile regression assumes a linear relationship between the τth conditional quantile, $Q_\tau(w|X)$ and the vector of K covariates X:

$$Q_\tau(w|X) = X\beta(\tau)$$

Because the conditional quantile function $Q_\tau(w|X)$ is an inverse of the conditional distribution function, $F(w|X)$

$$Q_\tau(w|X) = F^{-1}(\tau|X)$$

estimating quantile regressions for a large number of values for $\tau = [0, 1]$ provides a way to estimate the conditional distribution function, which, in turn, allows an estimation of a counterfactual condition distribution.

Koenker and Bassett (1978) proposed a quantile regression estimator as

$$\widehat{\beta}(\tau) = \min_{b \in \mathbb{R}^K} \frac{1}{N} \sum_{i=1}^{N} (w_i - X_i b)(\tau - 1(w_i \leq X_i b))$$

where i represents one of N observations. Once the conditional distribution is estimated, equations (1.1) and (1.2) can be used to calculate the unconditional and counterfactual unconditional distributions of wages, respectively.

Calculation of a conditional distribution, however, may be problematic because it involves calculating an inverse of a function. This requires that $X\widehat{\beta}(\tau)$ be monotonic, i.e. $\tau_j \leq \tau_k \Rightarrow X_i\beta(\tau_j) \leq X_i\beta(\tau_k)$. But quantile regression does not guarantee this. To solve it, Melly (2005) proposes the following estimator of the inverse of the unconditional distribution of wages, $w = F^{-1}(\theta)$:

$$\hat{w}(\hat{\beta}, X) = \inf\left\{ w : \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{J}(\tau_j - \tau_{j-1})1(X_i\hat{\beta}(\tau_j) \leq w) \geq \theta \right\} \tag{1.3}$$

where J is the number of values of τ for which the conditional quantile regression is estimated. By combining $\hat{\beta}$ and X from two different groups it is possible to estimate a counterfactual distribution. For example, $\hat{w}(\hat{\beta}^{(A)}, X^{(B)})$ represents wage distribution that would prevaile in group A if the covariates were distributed as in group B.

To isolate the effects of residuals, the author uses the following estimator of the τth quantile of the residual distribution conditional on X:

$$\hat{\varepsilon}_g(\tau) = X_g\beta_g(\tau) - X_g\beta_g(0.5)$$

By combining estimates of residuals from group A with the coefficient of the median regression from group B, it is possible to construct a counterfactual distribution that would prevail if residuals were as in A and median returns to characteristics were as in B. Such a distribution can be estimated as $\hat{w}(\hat{\beta}^{(B,A)}, X)$ where $\hat{\beta}^{(B,A)} = \hat{\beta}^{(B)}(0.5) + \hat{\beta}^{(A)}(\tau) - \hat{\beta}^{(A)}(0.5)$.

A summary statistic of inequality ν can be decomposed as follows:

$$\begin{aligned}\nu(w^{(B)}) - \nu(w^{(A)}) =& \nu(\hat{w}(\hat{\beta}^{(B)}, X^{(B)})) - \nu(\hat{w}(\hat{\beta}^{(B,A)}, X^{(B)})) + \\ & \nu(\hat{w}(\hat{\beta}^{(B,A)}, X^{(B)})) - \nu(\hat{w}(\hat{\beta}^{(A)}, X^{(B)})) + \\ & \nu(\hat{w}(\hat{\beta}^{(A)}, X^{(B)})) - \nu(\hat{w}(\hat{\beta}^{(A)}, X^{(A)}))\end{aligned} \tag{1.4}$$

where the first term in parentheses represents the contribution of difference in residuals, the second the contribution of difference in coefficients, and the third the contribution of difference in covariates.

1.4.2 Detailed Decomposition of Wage Distribution

Detailed decomposition of differences in distribution requires not only decomposition into the wage structure part, Δ_S^ν, and the composition part, Δ_X^ν, but also estimation of contributions of individual covariates X_k to each of these components. In this case we want to estimate how much of the change in wage distribution between two points in time was due to a change in wage structure in small and large firms, and how much due to a change in employment shares in small and large firms. The method based on the conditional quantile regression outlined in the previous subsection allows estimation of the former, but not of the latter. This subsection outlines some challenges specific to detailed decomposition and then describes the method used in this study.

There are two important requirements for detailed decomposition which are difficult to satisfy simultaneously. First, we would want that the estimates of contributions of all covariates add up. In other words, if we denote contributions of covariate k to the wage structure and to the compositional effect as $\Delta_{S_k}^\nu$ and $\Delta_{X_k}^\nu$, respectively, then this requirement can be written as $\Delta_S^\nu = \sum_{k=1}^K \Delta_{S_k}^\nu$ and $\Delta_X^\nu = \sum_{k=1}^K \Delta_{X_k}^\nu$. Second, it would be desirable that the result of decomposition does not depend on the order in which contributions of individual covariates were estimated. This requirement is called path independence.

A class of detailed decomposition methods that satisfy the adding-up requirement but not the path independence estimate counterfactual distribution for covariate k by replacing only the distribution of X_k in group B with the distribution of X_k in group A, while keeping distributions of other covariates intact. Subsequently, in an estimation of the counterfactual for some covariate $j \neq k$ only distribution of X_j is replaced while distribution X_k is restored to the original one.

Another class of methods also proceeds sequentially by replacing distributions of individual covariates, but the difference is that distributions of the previously estimated covariates are not being restored. At the end of the procedure, when the last covariate is replaced, the distribution of all covariates is replaced by their distributions in A. This class of methods, in general, does not satisfy the adding-up requirement but results are path-independent.

A method proposed by DiNardo, Fortin, and Lemieux (1996) (DFL) falls in the second category. It uses a reweighing approach to estimate a counterfactual distribution $F_A^{(B,X_1)}$ that would prevail in group A if the conditional distribution of a binary covariate X_1 was distributed as in group B. In this study, groups A and B represent two time periods and the covariate of interest is a binary variable that indicates if the firm is large. If we denote the vector of all other covariates X_2 such that $X = [X_1 X_2']'$, the equation (1.1) can be written as

$$F_A(w) = \int \int F_A(w|X)\, dF_A(X_1|X_2)\, dF_A(X_2) \tag{1.5}$$

The counterfactual distribution of wage $F_A^{(B,X_1)}(w)$ can be calculated by replacing the conditional distribution of X_1 in A with the conditional distribution of X_2 in B:

$$F_A^{(B,X_1)}(w) = \int \int F_A(w|X)\, dF_B(X_1|X_2)\, dF_A(X_2) \tag{1.6}$$

By introducing a reweighing function

$$\Psi(X_1, X_2) = \frac{dF_B(X_1|X_2)}{dF_A(X_1|X_2)} = X_1 \frac{P_B(X_1 = 1|X_2)}{P_A(X_1 = 1|X_2)} + (1 - X_1) \frac{P_B(X_1 = 0|X_2)}{P_A(X_1 = 0|X_2)} \tag{1.7}$$

the above expression becomes

$$F_A^{(B,X_1)}(w) = \int\int F_A(w|X)\,\Psi(X_1,X_2)\,dF_A(X_1|X_2)\,dF_A(X_2) \tag{1.8}$$

The conditional probabilities in the reweighing function can be estimated from a probit regression of X_1 on the vector of other covariates X_2. This regression is estimated twice, once for each time period. Predictions of these two regressions can be used as estimates of conditional probabilities $P_g(X_1 = 1|X_2)$ and $P_g(X_1 = 0|X_2)$ for $g = A, B$. The last step in calculating the counterfactual distribution is integration.

This counterfactual distribution allows us to calculate the composition effect of X_1 between two time periods, i.e., the contribution of the change in covariate X_1 between times A and B, assuming that all other covariates and wage structure remain as in A:

$$\Delta^{\nu}_{X_1} = \nu(F_A(w)) - \nu(F_A^{(B,X_1)}(w)) \tag{1.9}$$

To estimate the effect of the change in wage structure, I follow the approach proposed by Dinardo and Lemieux (1997) in their estimate of the effect of unions on wage distribution. They estimate the following counterfactual distribution, one for each time period:

$$F_g^{(S_1)}(w) = \int F_g(w|X_1 = 0, X_2)\,\Psi_{g,S_1}(X_2)\,dF_g(X_2, X_1 = 0) \tag{1.10}$$

where

$$\Psi_{g,S_1}(X_2) = \frac{P_g(X_1 = 0)}{P(X_1 = 0|X_2)}$$

The counterfactual distribution $F_A^{(S_1)}(w)$ represents distribution of wages that would prevail in period A if all firms were small (i.e. $X_1 = 0$). The difference in change in wage structures between small and large firms contributed to the change in wage distribution between years A and B can be calculated as

$$\Delta^{\nu}_{S_1} = [\nu(F_A(w)) - \nu(F_A^{(S_1)}(w))] - [\nu(F_B(w)) - \nu(F_B^{(S_1)}(w))] \tag{1.11}$$

1.5 Estimation and Results

This section presents results obtained by three different methods. Each of them gives a slightly different perspective of wage distribution and its changes, although there is some overlap. The section starts with a Melly aggregate decomposition of differences in wage distribution between large and small firms at a point in time. The decomposition is estimated for 1992 and 2012. It is followed by the DFL detailed decomposition, which estimates contributions of changes in wage structure and changes in the composition of firms to the changes in wage distribution between 1992 and 2012. And finally, there is a Melly aggregate decomposition of changes in wage distribution estimated for each firm size. All estimation in this section is performed on the CPS dataset, rather than the combined dataset, because the ExecuComp lacks some demographic information that is used in the estimation methods.

1.5.1 Effects of Size on Wage Distribution

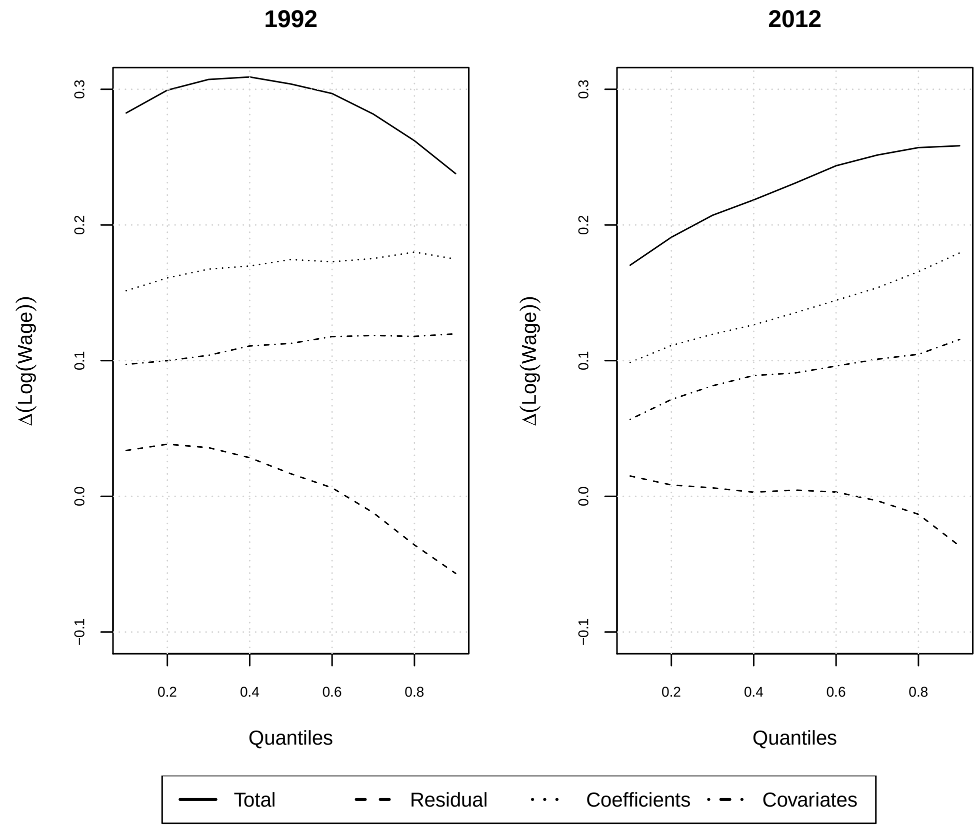
Table 1.7 shows differences in measures of wage inequality between small and large firms and results of the aggregate decomposition of these differences in 1992 and 2012. One hundred quantile regressions for each year are estimated with covariates that include dummies for gender and race, five dummies for education, four dummies for experience, the interactions of education and experience, and 19 dummies for industries. Observations are weighted by the CPS weights multiplied by hours worked. Actual and counterfactual distributions, were estimated by the estimator defined in (1.3). These distributions were then used to decompose differences in measures of inequality between small and large firms according to (1.4), where group B represents large firms (1000+ employees) and group A small firms.

Table 1.7: Differences in wage distribution between large (> 1000 employees) and small firms in 1992 and 2012. Covariates include gender, education, experience, industry and interactions between education and experience. Source: CPS 1993 and 2013.

Year	Statistics	Total	Coefficients	Covariates	Residuals
1992	90/10	-0.073	0.063	0.017	-0.152
	90/50	-0.109	0.033	-0.002	-0.140
	50/10	0.036	0.029	0.019	-0.012
2012	90/10	0.053	0.115	0.046	-0.108
	90/50	0.013	0.068	0.029	-0.084
	50/10	0.040	0.047	0.016	-0.023

The last three columns represent the three terms in parentheses on the right hand side of (1.4). In particular, *Coefficients* represents the second term, *Covariates* the third term, and *Residuals* the first term. There are a few interesting results. First, the bottom part of wage distribution was more unequal in large firms in both years. The total difference in the 50/10 percentile ratio between small and large firms, as well as the estimated individual components did not change much over the two decades. Second, the top part of wage distribution was less unequal in large firms in 1992. Overwhelmingly, this was due to lower residual inequality in large firms. Returns to observed characteristics are in fact distributed more unequally in large firms. Workforce composition had a very small effect on this difference. Third, in 2012 large firms caught up with small firms with respect to inequality in the top half of wage distributions. Both wage structure and workforce composition contributed to this. Residual inequality remained lower in large firms, but not lower enough to cancel the effects of coefficients and covariates. Figure 1.6 shows this decomposition for each percentile.

Figure 1.6: Aggregate decomposition: the difference in log-wage distributions between large and small firms in 1992 and 2012, and its components.



1.5.2 Effects of Size on Changes in Distribution over Time

Table 1.8 shows results of a detailed decomposition of changes in four measures of wage distribution between 1992 and 2012. The first two columns show values of various measures of inequality in 1992 and 2012, respectively, and the third column shows their difference. Columns four and five show estimates of contributions to this difference by the change in employment shares for small and large firms, and by the change in wage structures of small and large firms, respectively.

The *Composition* column is estimated by (1.9) and the *Wage Structure* column by (1.11).

Groups A and B in this case represent years 1992 and 2012, respectively. Probability density functions are estimated with a Gaussian kernel density estimator at 2000 points. Weights used in estimation of the actual distributions are the CPS weights multiplied by hours worked. For the counterfactual distributions, these weights are multiplied by the appropriate reweighing function.

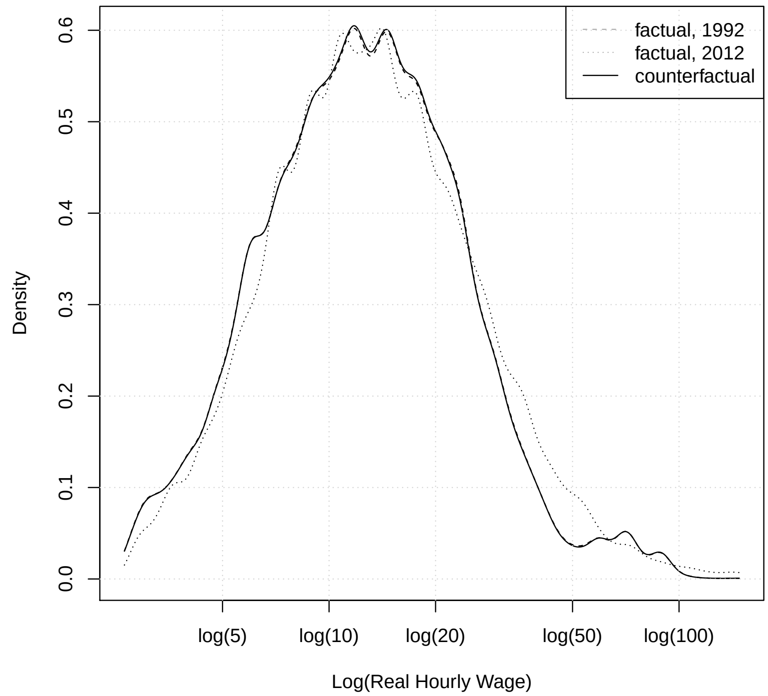
To estimate the reweighing functions, I estimated two binary probit models, one for 1992 and one for 2012. In each, the dependent variable is a binary variable that indicates whether the employer is a large firm. Independent variables are binary variables for gender and race (white/non-white), five dummies for education, four dummies experience and 18 dummies for industries. Observations are weighted by their CPS weights. Predictions of the probit models are used to calculate conditional probabilities $P_g(X_1 = x|X_2)$. The unconditional probability $P_g(X_1 = x)$ is simply the proportion of observations in group g for which $X_1 = x$.

Table 1.8: Estimates of changes in wage distribution due to the change in composition (employment shares) and wage structures in small and large firms.

Measure	1992	2012	Difference	Composition	Wage Structure
Gini	0.368	0.412	0.044	-0.001	0.007
				-1.6 %	15.2 %
90-10	1.658	1.747	0.089	-0.005	0.023
				-5.3 %	26.3 %
90-50	0.817	0.920	0.103	-0.005	0.047
				-4.5 %	45.5 %
50-10	0.841	0.827	-0.014	0.000	-0.023
				0.0 %	166.7 %

The results show that almost all of the increase in inequality of wages between 1992 and 2012 that can be attributed to firm sizes arises from the change in the wage structure in large firms. Interestingly, most of that change occurred in the top half of the distribution. The change in the 50/10 percentile ratio is negligible and has the opposite sign.

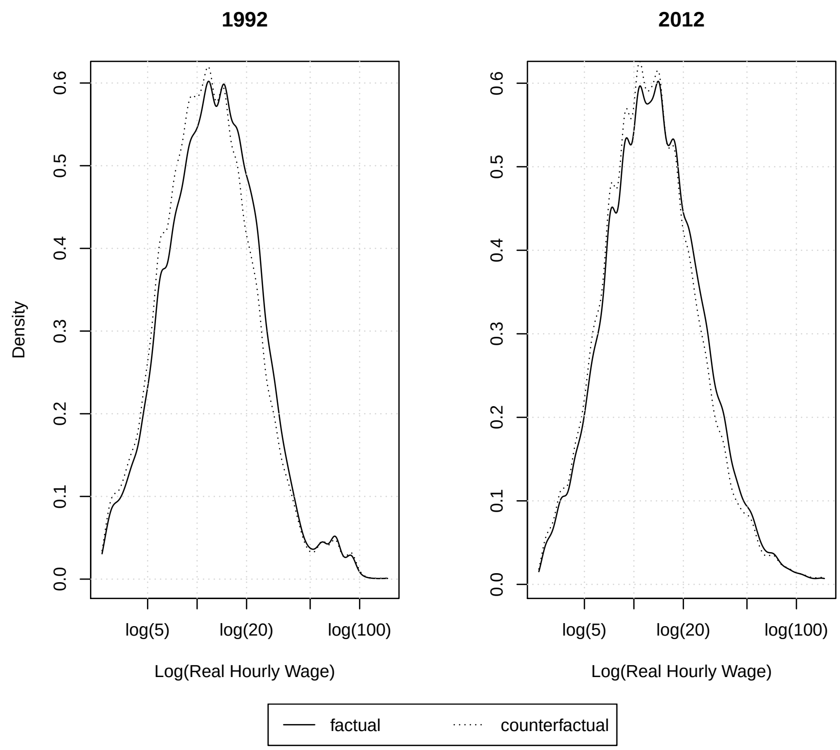
Figure 1.7: Detailed decomposition: log wage distribution in 1992, 2012, and the counterfactual distribution that would prevail in 1992 if the composition of firms with respect to size was as in 2012.



Figures 1.7-1.9 show these results graphically. Figure 1.7 shows the actual wage distributions in 1992 and 2012, and the counterfactual distribution $F_A^{(B,X_1)}$ that would prevail in 1992 if employment shares of small and large firms were as in 2012. The difference between the actual distribution for 1992 and the counterfactual distribution represents the effect of change in employment shares in small and large firms. The counterfactual distribution is almost indistinguishable from the actual distribution in 1992, which shows that, even though employment share in large firms significantly increased over time, its effect has been negligible.

Figure 1.8 shows distributions used to estimate the effect of changes in the wage structure. The left panel shows the actual distributions in 1992 and the counterfactual distribution that would prevail if wage structure in all firms was as it was in small firms. The right panel,

Figure 1.8: Detailed decomposition: actual and counterfactual distributions for 1992 and 2012. The counterfactual distributions are constructed as if wage structure in all firms was as in small firms for a given year.



which shows the same two distributions for 2012, indicates much smaller gap between the two distributions. By subtracting the counterfactual from the actual distribution for both years, and then calculating the difference between the two we can calculate the effect of change in wage structure.

Finally, Figure 1.9 shows the differences between distributions that represent the two effects. The solid line represents the effect of wage structure, and the dotted line represents the effect of change in employment shares. The graph makes it clear that wage structure effects are an order of magnitude bigger than the compositional effects.

To look at the changes over time from a different angle, I also estimate an aggregate

Figure 1.9: Detailed decomposition: the effects on change in wage distribution of changes in composition of firms with respect to size and wage structure.

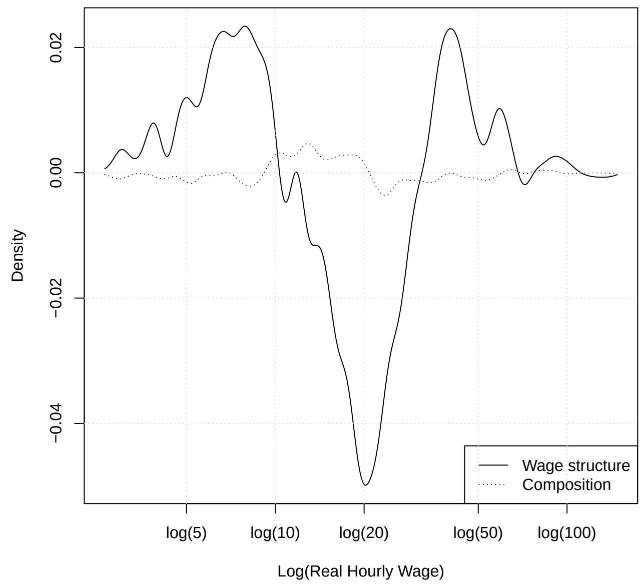


Table 1.9: Changes in 90/10, 90/50 and 50/10 percentile log wage differences between 1992 and 2012 for small, medium and large firms. The changes are decomposed into parts due to changes in coefficients, covariates and residuals.

Firm size	Statistics	Total	Coefficients	Covariates	Residuals
Small	90/10	-0.004	-0.022	0.062	-0.045
	90/50	0.003	-0.010	0.023	-0.010
	50/10	-0.007	-0.011	0.039	-0.035
Medium	90/10	0.081	0.013	0.060	0.009
	90/50	0.081	0.017	0.017	0.047
	50/10	0.000	-0.005	0.043	-0.037
Large	90/10	0.174	0.066	0.080	0.028
	90/50	0.160	0.062	0.048	0.049
	50/10	0.014	0.004	0.031	-0.021
All	90/10	0.065	0.010	0.071	-0.016
	90/50	0.080	0.020	0.030	0.030
	50/10	-0.015	-0.010	0.041	-0.045

decomposition proposed by Melly (2005) applied to changes in wage distribution over time. Group A now represents year 1992 and group B year 2012. I decompose the change in the wage distribution for all firms together and separately for small, medium and large firms. Table 1.9 shows the results. The last four columns represent the total change in the inequality statistics, a part contributed by a change in returns to observed characteristics, a part contributed by a change in the labor force composition with respect to observed characteristics, and a part contributed by a change in residual inequality, respectively.

Total changes in wage inequality reveal the familiar pattern: the biggest change occurred in large firms, a moderate one in medium firms, and there was almost no change in small firms. Most of the increase in inequality in medium and large firms was located in the upper half of the distribution. Estimates of individual components show that changes in covariates caused similar amount of change in wage distribution in all firms: the 90/10 percentile ratio increased by around 6 points in small and medium firms, and by 8 points in large firms, although the changes are distributed between the upper and lower tail somewhat differently.

The most significant differences among the firms, however, are in the changes of coefficients, i.e. rewards to the observed characteristics. Inequality in the top half of wage distribution increased by 6.2 points in large firms, 1.7 points in medium firms, and decreased slightly in small firms. There were no significant changes in the bottom half of distribution due to coefficients.

Residuals represent the unexplained part of the change that combines changes in unobserved characteristics of employees and returns to these characteristics. This residual inequality increased in the top half of distribution in large and medium firms, and decreased elsewhere. Considering that the aggregate decomposition in subsection 1.5.1 shows lower residual inequality in large firms, this result may be interpreted as convergence in residual inequality.

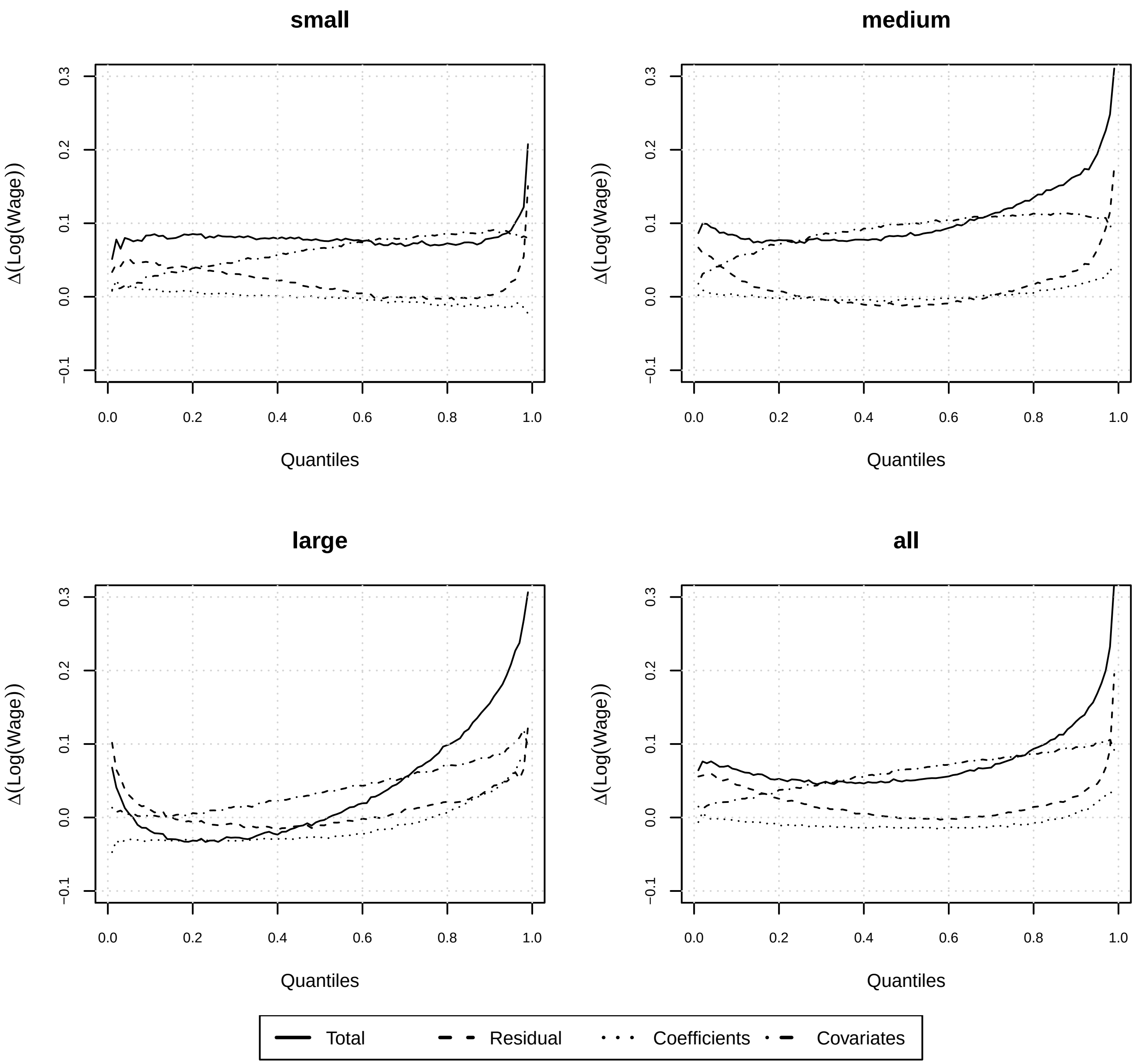
It is important to note similarities and differences between this decomposition and the detailed decomposition presented in Table 1.8. Numbers in the *Total* column for all firms correspond to the numbers in *Difference* column in Table 1.8. They differ because they are estimated by different methods. While the numbers slightly differ, both methods show a large increase in inequality in the top half and a small decrease in inequality in the bottom half of the distribution.

The all-firms part of column *Coefficients* corresponds somewhat to the column *Wage Structure* in Table 1.8. Both represent differences in returns to individual characteristics, but the former measures only returns to the observed characteristics, while the latter includes returns to both observed and unobserved characteristics. Again, the numbers differ, but the two methods agree on the direction and relative magnitude of changes. The last two columns, however, have no analogue in Table 1.8.

Figure 1.10 shows results of this decomposition graphically. A cursory glance reveals the already established result that the biggest change in wage distribution between 1992 and 2012 occurred in large firms, where most wages in the bottom half of distribution decreased, while all of those in the top half increased, some of them significantly. The decomposition shows that a change in the wage structure played the most significant role in the reduction of wages at the bottom, while all three factors work together in increasing wages in the top part of the distribution.

In small firms, on the other hand, changes in wages were remarkably consistent across most of the distribution. Changes in wage structure were insignificant, while contributions of residuals and covariates vary across the distribution, but in a way that keeps their sum approximately constant.

Figure 1.10: Decomposition of changes in the distribution of log hourly real wages into the changes due to individual covariates, coefficients, and residuals, for small, medium, and large firms between 1992 and 2012.



1.6 Conclusion

Hierarchical organization of the firm implies, under some commonly made assumptions, higher dispersion of wages in large firms. It comes as a surprise then to find that in the real world wages are more unequally distributed in small than they are in large firms. It is even more surprising to find the reason for this difference. Precisely because of this hierarchic effect, which is unobserved, it is reasonable to expect that residual inequality is higher in large firms. In other words, if the distributions of observed characteristics and returns to them were the same in small and large firms, inequality would be higher in large firms. Contrary to the expectations, I find that residual inequality is significantly higher in small firms. Even though inequality due to observed characteristics is higher in large firms, the effect of residual inequality exceeds it and makes overall inequality higher in small firms.

Besides offering a new perspective, this result has broader implications. In the context of research of wage inequality, it points in a new direction where sources of wage inequality should be explored. The literature has found that a large portion of the increase of wage inequality in the 1970s and 1980s was due to an increase in residual inequality. Lemieux (2006) attributed some of it to a measurement error, but there are few other explanations being suggested. Even though it is customary to mention that this may be due to changes in rewards to unobserved skills, this is usually where the discussion ends. Firm size could be one way to resume this discussion. Of course, the fact that firm size plays a role in wage inequality does not by itself represent an explanation. But, especially considering the effect of firm size on residual inequality, it does open a new questions and perhaps a new avenue of research.

Another discussion to which these results might contribute is the one about causes of firm size-wage premium. Even though the existence of this premium has been known since the study by Moore (1911), there is still no agreement about its cause. Economists have estimated the difference in average wage between small and large firms and explained a part of it by differences in the composition of labor force. Still, Brown, Hamilton, and Medoff (1990) report that regressions that control for observable characteristics of employees and employers leave around 15% of the difference unexplained. Multiple explanations have been proposed but none accepted so far. This study sheds some light on this issue.

One hypothesis, favored by Brown, Hamilton, and Medoff (1990), is that the size premium is a result of lower costs of other inputs that large firms face. Discount on large purchases and lower interest rates on loans are two examples of such costs. However, it is not obvious why the resulting gains should benefit the workers rather than the owners of the firm. One suggested explanation is that large firms pay higher wages to reduce shirking. If this is true, the question is how this excess wage is distributed. If reduction of shirking is the main reason for higher average wage in large firms then it also needs to account for lower residual inequality in these firms. The measures of inequality used in this study measure relative inequality; Gini and percentile ratios do not change if all wages are multiplied by the same factor. Therefore, to reduce residual inequality, the excess wage would need to be distributed in a way that rewards workers with lower wages relatively more than workers with higher wages, which would be difficult to justify. Assuming that, apart from an anti-shirking incentive, workers are paid their marginal product, there is no reason to provide higher incentive to less productive workers.

The findings in this study are more in line with the findings by Idson and Oi (1999), but not with their main conclusion. The authors provide some evidence of higher capital-worker ratio in large firms, as well as more advanced technology and higher labor productivity, which is certainly one reason why wages are higher in large firms. According to the capital-skill complementarity theory, this would also generate higher demand for skilled labor, higher education premium, and thus higher inequality due to observed characteristics. However, as authors also note, this is not the answer to the firm size-wage premium puzzle, which is a difference in wages among workers with the same observed characteristics. To explain this difference, the authors propose that large firms have ability to better match workers' unobserved characteristics with jobs; for example, a worker that puts in more effort may be matched with a better machine than another worker with the same observed characteristics. This would make the harder working workers more productive, but it should also make their wages higher and thus increase residual inequality. However, a lower residual inequality in large firms found in this study does not support this hypothesis.

Another contribution of this study is a construction of a new dataset. Researchers of wage inequality in the U.S. have mostly relied on the CPS data. Estimates of inequality based on it, however, are downward biased because of top-coding and under-sampling of high incomes. By merging the CPS data with the ExecuComp, I obtain a dataset that better represents the actual wage distribution.

Chapter 2

Firm Heterogeneity and Wealth Distribution

2.1 Introduction

Studying distribution of wealth is notoriously challenging. Not only is wealth difficult to measure and data on its distribution are lacking, but mathematical models that are capable of capturing the dynamics of wealth distribution are difficult to construct. The complexity of wealth dynamics does not lend itself easily to simplifications required for constructing tractable economic models. Furthermore, wealth distributions tend to be extremely skewed. Two main features of the wealth distribution in the U.S. are that around 40% of the population holds no wealth, and a small number of people hold a huge part of the total wealth. Constructing a general equilibrium model that endogenously generates that kind of wealth distribution is not a simple task.

This study proposes one such model. I construct a dynamic general equilibrium model with heterogenous agents and firms. Agents are assigned to one of the two occupations: production workers or managers. Combined with the firm heterogeneity, this generates substantial inequality in distribution of labor income, which, in turn, yields a distribution of wealth whose upper tail comes close to the upper tail of the U.S. wealth distribution. The fit in the bottom of the wealth distribution is less satisfactory.

Firm heterogeneity is based on a model by Lucas (1978), in which a firm's size is positively correlated to its manager's ability. Each agent has some randomly assigned, time invariant managerial talent. The most efficient assignment of agents to firms assigns the most talented agents to managerial positions, while others are employed as workers. This assignment determines labor income distribution. Workers receive the equilibrium wage, while a manager receives the profit of the firm he manages. Managerial rent is higher than wage and positively related to the size of the firm. To solve the model in a dynamic general equilibrium framework with aggregate shocks I use a method proposed by Krusell and Smith (1998). I calibrate the model to reflect the evolution of firm size distribution in the U.S.

To understand general challenges of constructing a model for studying income and wealth distributions, it is necessary to understand the key components of these two distributions and how they interact with each other. An individual's income consists of labor and capital income. Labor income depends on many factors, such as the individual's occupation and education, and state of the economy, which also includes aggregate capital or wealth. Capital income represents the compensation for lending one's wealth over a period of time; it is determined by the individual's wealth and the prevailing interest rate. An individual's wealth, on the other hand, consists of all assets that she accumulated over her lifetime. One of the main determinants of this asset accumulation process is the individual's past income.

Therefore, past income causes current wealth, and current wealth causes current income. This two-way causality between income and wealth requires that a model aiming at explaining wealth distribution represents important features of both distributions and their interaction.

Several models that achieve this goal can be found in the literature. The one proposed by Krusell and Smith (1998) (KS henceforth) has become a workhorse in the field, exhibiting both parsimony of design and richness of behavior. It consists of a number of ex-ante identical agents who receive stochastic idiosyncratic shocks to their labor supply. The shock determines whether an agent is employed or unemployed in any given time period, generating a bimodal distribution of labor income. This simple but unrealistic income process generates a surprisingly complex distribution of wealth in the model.

The wealth distribution generated by the benchmark KS is not complex enough, though, as it fails to account for the fraction of population with no wealth, and for the big concentration at the top. The authors improved its performance by two modifications: (1) they obtained a large fraction of agents with zero wealth by introducing a minimum income; (2) they increased the concentration of wealth at the top by introducing heterogeneity in agents' time-discount factor.

It is this solution for the high concentration of wealth at the top where the KS model is the most open to criticism. In effect, it implies that the tremendous inequality in wealth that we have been seeing in the U.S. and, in particular, its concentration at the top are primarily caused by different rates at which individuals save. While heterogeneity in saving rates certainly plays a role in the shape of wealth distribution, an important determinant of wealth distribution is the distribution of income. This relationship is not present in the KS model.

One successful approach to having a realistic labor income distribution was proposed by Iacoviello (2008) who uses an exogenous labor income process. The process consists of a time-invariant individual-specific component, economy-wide time-varying component, and an idiosyncratic stochastic component. The author calibrates it to match the evolution of labor income distribution in the U.S. and achieves a fairly realistic wealth distribution that is determined largely by the labor income distribution. One problem is that the labor income process is completely exogenous to the model, and thus the effect of wealth accumulation on labor income is missing. In general, an increase in aggregate capital stock raises wages, and it may have a complex effect on their distribution. Iacoviello's approach does not include this relationship.

The model presented in this study contains a somewhat realistic distribution of labor income, while preserving the relationship between it and aggregate capital. It achieves that by assigning agents to two occupations and allowing firms to vary by size. All production workers receive the same, market-determined wage, which is a function of the aggregate capital-labor ratio. The agents who are not production workers are managers, each of which manages exactly one firm and receives the firm's profit as compensation. It is this managerial compensation that creates a skewed labor income distribution. Managers with high managerial talent manage larger firms, which generate more profit and thus yield higher managerial salaries. Since wages and distribution of firm sizes depend on the aggregate capital-labor ratio, the overall distribution of labor income does as well.

This particular approach to modeling heterogeneous firms was first proposed by Lucas (1978), and it has been used in other studies in slightly different forms to study distribution of firm sizes and managerial income (see Rosen, 1982; Gabaix and Landier, 2008; Banerjee and

Moll, 2010). A question that poses itself, however, is about the identification of managerial talent. It is not clear what that talent is, but most likely it represents some combination of innate talent, education and experience. More importantly, it is clearly unobservable. The Lucas model, however, provides a one-to-one relationship between the distribution of managerial talent and the distribution of firm sizes, the latter of which is observable. By estimating the distribution of firm sizes from the data, one can infer the distribution of managerial talent.

This approach has an added benefit of modeling the relationship between the distribution of firm sizes and the distribution of labor income, which has been documented in the largely empirical literature. Brown, Hamilton, and Medoff (1990) estimate, based on the May 1983 CPS, a 35% wage gap between firms with more than 500 employees and small firms.[1] The gap is even bigger when fringe benefits are taken into account. Abowd, Kramarz, and Margolis (1999) find a strong positive relation between firm size and wages in a large longitudinal sample of firms in France. Baker, Gibbs, and Holmstrom (1994b) provide a close-up analysis of employment patterns at a UK firm. They find that around 70 % of the variance in wage across employees of one U.K. firm is due to hierarchical levels.[2] Gabaix and Landier (2008) find a positive relationship between firm size and executive compensation.

The following section presents the model and defines its equilibrium. Section 3 outlines computational algorithms used for solving and simulating the model, and section 4 is about calibration of the model. Section 5 presents results and section 6 concludes.

[1] When differences in worker quality are taken into account, the size of the wage premium is between 10 and 15%.

[2] They identify eight hierarchical levels, level 1 being the lowest and level 8 the CEO. Interestingly, first four levels are of similar size; narrowing begins only at level 5. Also, the firm almost quadrupled in size over the observed period, from 1380 to 5218 employees, but the number of observed levels remained the same. This might be due to previously unused management capacity in the lower levels.

2.2 Model

The model consists of N infinitely lived agents, each of which supplies labor and maximizes the expected lifetime utility by choosing the optimal consumption path. An agent's income consists of wage and the interest paid on his holding of wealth. Agents provide labor to a number of firms of different sizes. There are two occupations, a production worker and a manager, to which workers are assigned in each period according to their managerial talent. The economy follows a Markov process and can be in one of two states, good or bad. This state determines the aggregate supply of labor and firms' productivity.

2.2.1 Households

Agent i maximizes the expected lifetime utility

$$E_0 \sum_{t=0}^{\infty} \beta^t U(c_{it}) \tag{2.1}$$

$$\text{subject to } \quad c_{it} + k_{i,t+1} - (1-\delta)k_{it} = r_t k_{i,t} + y_{i,t} \tag{2.2}$$

$$k_{i,t+1} \geq 0 \tag{2.3}$$

where c_{it} is consumption, k_{it} capital, δ depreciation, and y_{it} labor income. An agent can be employed as a worker or as a manager. Agent's labor income equals the equilibrium wage if he is employed as a worker; it equals firm's profit if the agent is a manager.

Utility function is given by

$$U(c) = \frac{c^{1-\nu} - 1}{1-\nu}$$

Each agent is endowed with a randomly assigned, time-invariant managerial talent $x \in (0,1)$, which is drawn from some distribution $\Gamma(x)$. As shown by Lucas (1978), the competitive equilibrium determines a threshold value z_t such that agent i will be a worker if

$x_i < z_t$, and a manager otherwise. There are N agents and they are indexed in the order of increasing x. Define $n_t = \max_i\{x_i | x_i < z_t\}$. Then, at time t, agents $1, \ldots, n_t$ are employed as workers, and agents $n_t + 1, \ldots, N$ as managers.

In each time period, agents experience idiosyncratic random shock ε_{it}, which determines their labor supply for the period. The shock can have two values. If $\varepsilon_{it} = 1$, agent i provides one unit of labor in period t; if $\varepsilon_{it} = 0$, agent i is unemployed at time t. The probability distribution of the idiosyncratic shocks is determined by the state of the economy.

2.2.2 Firms

Firms produce a single good by using labor, capital and managerial skill. Each firm has one manager and firms are indexed by the same index as their managers. If agent i is a worker in period t, then firm i produces zero output in that period. Production function of the firm managed by agent i is

$$Y_{it} = x_i g(f(K_{it}, L_{it})) = x_i \{A_t[(\lambda K_{it}^{\alpha} + (1-\lambda) L_{it}^{\alpha}]^{\frac{1}{\alpha}}\}^{\gamma} \tag{2.4}$$

where Y_{it} is output of firm i and A_t aggregate technology at time t. The elasticity of substitution between labor and capital is $\sigma = \frac{1}{1-\alpha}$. A strictly increasing and strictly concave function $g : \mathbb{R}^+ \rightarrow \mathbb{R}^+$, where $\mathbb{R}^+$ is the set of positive real numbers, introduces diminishing returns to scale to avoid a situation in which there is only one firm. It is assumed that $\gamma \in (0, 1)$.

Technology A_t follows a Markov process and it can be in one of the two states: good, A_g, or bad, A_b. Transitional probabilities are given by matrix π such that $\pi_{i,j}$ represents conditional probability that A_{t+1} will be j given that A_t is i. Furthermore, the distribution of idiosyncratic shocks, and thus the aggregate labor supply are determined by the aggregate shock.

The efficient allocation of inputs at time t is defined by value z_t and allocation of capital

$\mathbf{K_t}$ and labor $\mathbf{L_t}$ to $N - n_t$ firms that maximize total output of the economy at time t. $\mathbf{K_t}$ and $\mathbf{L_t}$ are N-dimensional vectors, such that $K_{it}, L_{it} = 0$ if agent i is a worker; if agent i is a manager, K_{it}, L_{it} equal the amount of capital and labor employed in the firm he manages. An allocation is feasible if $\sum_{i=1}^{N} K_{it} = \sum_{i=1}^{N} k_{it}$ and $\sum_{i=1}^{N} L_{it} = n_t$. The model does not specify how this allocation takes place, but it assumes that it does in every time period.

2.2.3 Equilibrium

The aggregate state of the economy is given by (Θ_t, A_t), where Θ_t is the current distribution of agents' holdings of capital. Given some initial distribution Θ_0, the model endogenously determines Θ_t. Input prices $w_t(\Theta_t, A_t)$ and $r_t(\Theta_t, A_t)$ are determined competitively. The existence of a competitive dynamic equilibrium requires that a law of motion that defines transition into the next time period exists, even though it is not known in advance. It is denoted as $\Theta_{t+1} = H(\Theta_t, A_t)$.

An individual agent solves the intertemporal optimization problem. In addition to his current capital holding, current income and interest rate, important information necessary for predicting future income and interest rate is the aggregate state and its law of motion. In other words, information relevant to the agent is given by $(x_i, k_{it}, \varepsilon_{it}, \Theta_t, a_t)$. Agent's optimization problem can be stated as

$$v(x_i, k_{it}, \varepsilon_{it}, \Theta_t, a_t) = \max_{c_t, k_{i,t+1}} \{U(c_t) + \beta E[v(x_i, k_{i,t+1}, \varepsilon_{i,t+1}, \Theta_{t+1}, a_{t+1})]\} \tag{2.5}$$

subject to

$$c_{it} + k_{i,t+1} - (1-\delta)k_{it} = r_t k_{it} + y_{it} \tag{2.6}$$

$$\Theta_{t+1} = H(\Theta_t, a_t) \tag{2.7}$$

$$k_{i,t+1} \geq 0 \tag{2.8}$$

where

$$y_{it} = \begin{cases} w_t, & \text{if } x_i < z_t \\ Y_{it} - r_t K_{it} - w_t L_{it}, & \text{if } x_i \geq z_t \end{cases}$$

and

$$\sum_{i=1}^{N} k_{it} = \sum_{i=n_t+1}^{N} K_{it} \tag{2.9}$$

$$n_t = \sum_{i=n_t+1}^{N} L_{it} \tag{2.10}$$

The solution of the problem is a decision rule function φ: $k_{i,t+1} = \varphi(x_i, k_{it}, \varepsilon_{it}, \Theta_t, a_t)$. A recursive competitive equilibrium is a law of motion H, the value function v, decision rule φ, interest rate $r(K_t, L_t, A_t)$, and labor income distribution $y(x_i, K_t, L_t, A_t)$.

The individual decision rule function, given in form of a policy rule for capital, is:

$$\begin{aligned} k_{t+1} &= (1-\delta+r_t)k_t + y_t \\ &\quad - \left[\beta E_t \frac{(1-\delta+r_{t+1})}{[(1-\delta+r_{t+1})k_{t+1} + y_{t+1} - k_{t+2}]^{\nu}} - \mu_{t+1}\right]^{-1/\nu} \end{aligned} \tag{2.11}$$

where μ_t is the Lagrange multiplier for the debt-limit constraint.

There are two features of the equilibrium worth noting. First, to calculate the expected value of future input prices, it is necessary to know the whole distribution Θ_t[3]. Second, the distribution of labor income depends on the time-invariant distribution of managerial talent Γ and the aggregate capital. Both features introduce significant complexity in solving the model. The first one is dealt with by KS. A method to deal with the second one is a contribution of this paper.

[3]Input prices are determined by the shocks and the aggregate capital. But, to calculate the aggregate capital, it is necessary to know its distribution

2.3 Computation

I solve equation (2.5) numerically, but there are two problems associated with this approach. First, the curse of dimensionality arises from the multi-dimensional nature of the state of the model, part of which is the distribution of capital Θ. An individual optimization takes into account not only the current state, but also the expectations of future states, which makes the procedure computationally intensive. Second, the calculation of the model's state, or the expectation thereof, requires solving the efficient assignment of inputs to firms for all possible states. This further exacerbates the problem of computational intensity. KS propose a solution for the first problem[4], and this paper proposes one for the second problem.

To address the problem of dimensionality, KS introduce a type of bounded rationality where agents use only a finite number of moments of the distribution Θ to form expectations of the future and solve their optimization problem. In effect, they transform the law of motion (2.7) into $\mathbf{m}_{t+1} = H_I(\mathbf{m_t}, a_t)$, where $\mathbf{m_t}$ is a matrix of first I moments.[5] The individual optimization problem is solved by iterating the value function on a grid of values for individual and aggregate capital and approximating the function with polynomials for values not on the grid. The approximate law of motion is used to calculate the expected values of future wage and interest rate.

Once optimal behavior of individual agents is found, it is easy to derive behavior of the aggregate state by simulating a large number of agents. This allows a comparison of the simulated moments to the perceived moments. A large discrepancy would indicate that a greater number of moments should be used by the agents. KS show that the first moment is enough for a very good approximation of the aggregate behavior.

[4] I use Matlab code written by Maliar, Valli, and Maliar (2009).

[5] From KS:

> Since current prices depend only on the total amount of capital and not on its distribution, limiting agents to a finite set of moments is restrictive only as far as future prices are concerned. In particular, to know future prices, it is necessary to know how capital stock evolves. Since savings decisions do not aggregate, the total capital stock in the future is a nontrivial function of all the moments of the current distribution.

While KS solve the problem of heterogeneous agents, the current model has an additional complexity that its firms are heterogenous as well. In the case of the representative firm, its output in each period is determined by the amount of inputs and the available technology. With heterogenous firms, output also depends on the assignment of inputs to individual firms. Lucas (1978) solves for the efficient assignment of inputs to firms with production function given by (2.4). Such an assignment maximizes the total output of the economy at time t. Assuming that individual intertemporal optimization results in the efficient assignments in all time periods[6], I apply Lucas' solution, and construct a numerical algorithm for its calculation. The algorithm is described in the following subsection.

The heterogeneity of firms also introduces a nonuniform distribution of wages that depends on the current state, which aggravates the curse of dimensionality. Whereas KS reduce the dimension of the model's state by using only the first moment of capital holding distribution to calculate future wage and interest rate, introduction of wage distribution requires calculation of the entire wage distribution for every permissible future value of aggregate capital. Even though wage distribution is flat for the majority of agents who are employed as workers, such an object would still be very large for population sizes normally used in simulations. To deal with this additional complexity, I calculate wage distribution only for the values of aggregate capital on the grid. For off-grid values, I approximate the wage distribution. The algorithm for calculation is a combination of the policy rule iteration and polynomial approximation. The policy rule is given by (2.11).

2.3.1 Computing the Efficient Assignment

The efficient assignment of workers, managers and capital to firms at time t maximizes total output of the economy at time t given the available technology A_t and total amount of capital

[6]It remains to be proved that the efficient assignment at time t will arise from individual intertemporal optimization problems.

K_t:

$$Y_t = \sum_{i=n_t+1}^{N} x_{it} g(f(K_{it}, L_{it})) \tag{2.12}$$

If we define the intensive form of production function as
$f(K_{it}, L_{it}) = L_{it} f(K_{it}/L_{it}, 1) = L_{it}\phi(\rho_t) = L_{it} A_t[(\lambda\rho_t^{\alpha} + (1-\lambda)]^{\frac{1}{\alpha}}$, then the following system of equations defines the efficient assignment:

$$\rho_{it} = \rho_t, \text{ for all } i > n_t \tag{2.13}$$

$$n_t = \sum_{n_t+1}^{N} L_{it} \tag{2.14}$$

$$\rho_t n_t = K_t \tag{2.15}$$

$$\frac{w_t}{r_t} = \frac{\phi(\rho_t) - \rho_t \phi'(\rho_t)}{\phi'(\rho_t)} \tag{2.16}$$

$$r_t = x_{it} g'(L(x_{it})\phi(\rho_t))\phi'(\rho_t) \tag{2.17}$$

$$z_t g(L(z_t)\phi(\rho_t)) = w_t + (w_t + r_t\rho_t)L(z_t) \tag{2.18}$$

where z_t is the managerial talent of the marginal manager (i.e. the manager with the lowest index i at time t); $\rho_{it} = K_{it}/L_{it}$ ratio of capital to labor for firm i; it follows from (2.16) that this ratio is the same for all firms; $L(x_{it})$ is the amount of labor employed by firm i, and $L(z_t)$ is the amount of labor employed by the firm $n_t + 1$, which is managed by the marginal manager; w_t, r_t are wage and capital rent. Equation (2.14) represents constraints on labor and, together with constraints on capital, implies (2.15). Equations (2.16) and (2.17) are derived from first order conditions for maximization of output. Equation (2.18) is the condition for the wage of the marginal manager; it has to be equal to his opportunity cost, which is worker's wage.

The strategy for solving the above system is to start with some initial value of n_t and iteratively solve the system until it converges to the solution. If we denote the initial value of n_t with ${n_t}^{(0)}$, then the managerial talent of the marginal manager, ${z_t}^{(0)}$, is the managerial talent of the agent ${n_t}^{(0)} + 1$. The value of ${\rho_t}^{(0)}$ can be calculated from equation (2.15). Equation (2.17), which holds for each firm, is used at this stage only for the firm managed by the marginal manager: $r_t = z_t g'(L(z_t)\phi(\rho_t))\phi'(\rho_t)$. This leaves the three non-linear equations, (2.16), (2.17), and (2.18), with three unknowns, w_t, r_t, and $L(z_t)$.

In the specific case of production function given by (2.4), the diminishing returns to scale function given by $g(y) = y^\gamma$, and for $x_{it} = z_t$, equations (2.16), (2.17), and (2.18) look like this:

$$
\begin{aligned}
\frac{w_t}{r_t} &= \frac{(\lambda\rho_t^\alpha + 1 - \lambda)^{\frac{1}{\alpha}} - \rho_t\lambda(\lambda + (1-\lambda)\rho_t^{-\alpha})^{\frac{1-\alpha}{\alpha}}}{\lambda(\lambda + (1-\lambda)\rho_t^{-\alpha})^{\frac{1-\alpha}{\alpha}}} \\
&= \frac{1-\lambda}{\lambda}\rho_t^{1-\alpha} \\
r_t &= z_t\gamma[L(z_t)A_t(\lambda\rho_t^\alpha + 1 - \lambda)^{\frac{1}{\alpha}}]^{\gamma-1}A_t\lambda(\lambda + (1-\lambda)\rho_t^{-\alpha})^{\frac{1-\alpha}{\alpha}} \\
&= z\gamma L(z_t)^{\gamma-1}A^\gamma\lambda(\lambda\rho_t^\alpha + 1 - \lambda)^{\frac{\gamma-\alpha}{\alpha}}\rho_t^{-(1-\alpha)} \\
w_t &= z_t[L(z_t)A_t(\lambda\rho_t^\alpha + 1 - \lambda)^{\frac{1}{\alpha}}]^\gamma - (w_t + r_t\rho_t)L(z_t)
\end{aligned}
$$

Substituting w_t from the first equation into the third one:

$$
\left(\frac{1-\lambda}{\lambda}\rho_t^{1-\alpha}(1 + L(z_t)) + \rho_t^\alpha L(z_t)\right) r_t = z_t\left(L(z_t)A_t(\lambda\rho_t^\alpha + 1 - \lambda)^{1/\alpha}\right)^\gamma \tag{2.19}
$$

Combining it with the second one provides a solution for $L(z_t)$

$$
L(z_t) = \frac{\gamma(1-\lambda)}{(\lambda\rho_t^\alpha + 1 - \lambda)(1-\gamma)} \tag{2.20}
$$

Now r_t and w_t can be calculated as well. Then, use (2.17) to calculate amount of labor demanded by each firm at these prices:

$$
L_{it}(x_{it}, w_t, r_t) = \frac{1}{\phi(\rho_t)}g'^{-1}\left(\frac{r_t}{x_{it}\phi'(\rho_t)}\right) = \frac{1}{\phi(\rho_t)}\left(\frac{r_t}{\gamma x_{it}\phi'(\rho_t)}\right)^{\frac{1}{\gamma-1}} \tag{2.21}
$$

Add all L_{it} to calculate total demand for labor and denote it ${L_t}^{(0)}$. If ${L_t}^{(0)} > {n_t}^{(0)}$, increase n_t; otherwise decrease it. Repeat the procedure.

To define the stopping criteria, it is important to note that equation (2.18) is derived for the case of a continuum of agents. In the discrete case, the marginal manager's wage will, in general, be somewhat higher than worker's wage. This means that there will be some $n_t^{(i)}$ for which

$$
\begin{aligned}
&L_t(w(n_t^{(i)} - 1), r(n_t^{(i)} - 1)) > n_t^{(i)} - 1 \quad \text{and} \\
&L_t(w(n_t^{(i)}), r(n_t^{(i)})) < n_t^{(i)}
\end{aligned}
$$

In other words, perform the algorithm until the latest increment is one and the difference between aggregate supply and demand changes the sign.

Select $n_t = n_t^{(i)}$ and recalculate $w_t, r_t, L(z_t)$ so that total demand for labor matches the supply. This is straightforward. Expression (2.21) represents an individual firm's labor demand as a function of interest rate. All other parameters are either exogenous or determined by the choice of n_t. Hence, scaling the interest rate by $(n_t/L_t(w(n_t), r(n_t)))^{\gamma-1}$ makes the aggregate labor demand equal to the aggregate labor supply.

2.4 Calibration

Because the main goal of this study is to compare the results of the proposed model with those of the benchmark model by KS, I try to match the calibration of their model as closely as possible. I use the same transition matrix for the Markov process that determines the state of the economy, as well as the values of parameters that define the two states. The technology A_t takes values of 1.01 and 0.99 in the good and bad state, respectively, and the unemployment rates that correspond to the two states are 4% and 10%. The discount factor β is set to 0.99 and the depreciation rate δ to 0.025. All calculations and simulations are performed with the population of 1,000 agents.

One of the requirements for the production function is that the elasticity of substitution between labor and capital be less than 1. According to Lucas (1978), this is a sufficient condition for existence of a unique solution for the value of managerial talent of the marginal manager, z_t. This requirement is the main reason for using the CES production function, unlike KS who use the Cobb-Douglas production function. For the coefficient of substitution between labor and capital, α, I use the value of -0.4, which guarantees that the elasticity of substitution between labor and capital, $\sigma = 1/(1-\alpha)$, is within the range estimated in the literature and less than 1.

A critical part at this stage is calibration of the managerial talent, which is unobserved. To estimate it, I use the relationship between it and the distribution of firm sizes. This relationship is given by (2.21). For a given aggregate state of the economy, (A_t, K_t, L_t), (2.21) can be written as $L_{it} = \varkappa_t x_i^{\frac{1}{1-\gamma}}$, and its inverse $x_i = (L_{it}/\varkappa_t)^{1-\gamma}$, where $\varkappa_t$ is a function of the capital to labor ratio and the interest rate.[7] The relationship between the distribution of the managerial talent and the distribution of firm sizes can be derived as follows:[8]

$$P(X > x) = P((L/\varkappa)^{1-\gamma} > x) = P(L > \varkappa x^{\frac{1}{1-\gamma}}) \tag{2.22}$$

Assuming that the distribution of firm sizes follows a Pareto distribution, as shown by Axtell (2001), such that $P(L > l) = l^{-\omega}$, it follows that

$$P(X > x) = (\varkappa x^{\frac{1}{1-\gamma}})^{-\omega} \tag{2.23}$$

Therefore, the distribution of managerial talent can be inferred from the distribution of firm sizes. I fit the Pareto distribution to the data on firm size in the U.S. for each year

[7]From $L_{it} = \frac{1}{\phi(\rho_t)}\left(\frac{r_t}{\gamma x_{it}\phi'(\rho_t)}\right)^{\frac{1}{\gamma-1}}$, it follows that $\varkappa_t = \frac{1}{\phi(\rho_t)}\left(\frac{r_t}{\gamma\phi'(\rho_t)}\right)^{\frac{1}{\gamma-1}}$, where $\phi'(\rho_t) = \lambda A_t \rho_t^{\alpha-1}[(\lambda\rho_t^{\alpha} + (1-\lambda)]^{\frac{1}{\alpha}-1}$.

[8]I drop subscript t for simplicity.

between 1977 and 2009.[9] The data is from the Business Dynamics Statistics released by the U.S. Census Bureau. It is provided in the form of tabulated frequencies with unequal bin sizes. The smallest bin contains firms with 1-4 employees and the biggest contains all firms with more than 10,000 employees.

To estimate ω, I use the tail CDF:

$$P(L > l) = 1 - P(L < l) = \left(\frac{l_0}{l}\right)^{\omega} \tag{2.24}$$

in the log-log form:

$$\log(P(L > l)) = \omega \log(l_0) - \omega \log(l) \tag{2.25}$$

where l_0 is the minimum size of the firm. For $l_0 = 1$, this simplifies to

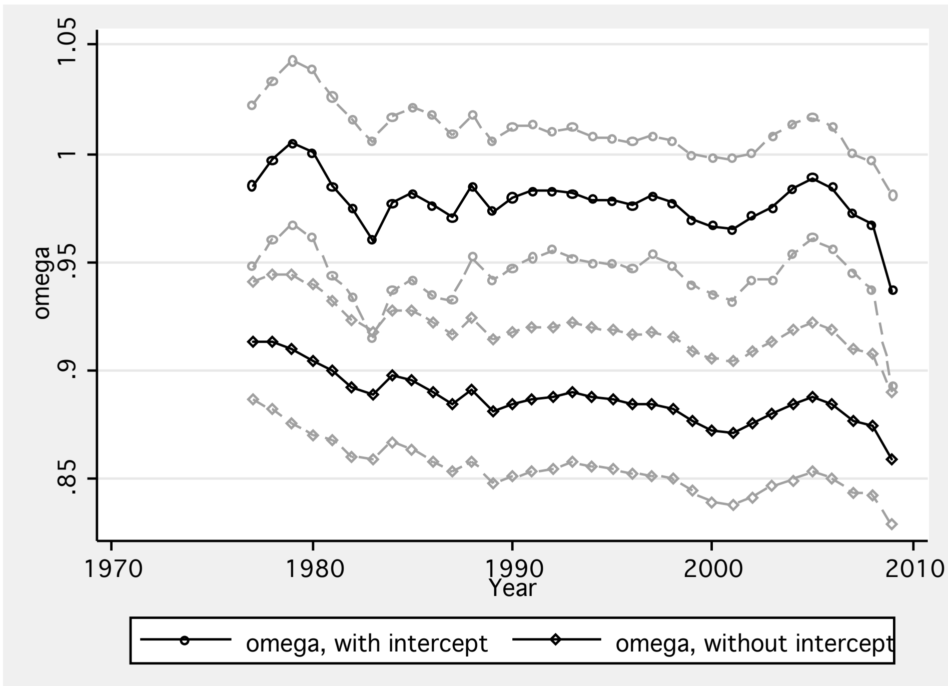
$$\log(P(L > l)) = -\omega \log(l) \tag{2.26}$$

I estimate both (2.25) and (2.26) by using OLS for each year in the sample. The results are shown in Table 2.2 and in Figure 2.1. When value of the intercept is not restricted to zero, as in (2.25), the estimate of ω is not statistically different from 1 for most years in the sample. In the second case, given by (2.26), the estimate is not statistically different from 0.9 for all but the last year in the sample. A statistical test for equality of the estimated values of ω for different years fails to reject the hypothesis that the coefficients are equal in all years in the sample. This is the case for both estimation equations. Despite of that, for both specifications Figure 2.1 shows a decreasing trend in ω, which means a shift toward a more skewed distribution of firm sizes.

There are no strong arguments for preferring either specification. Considering that the goodness of fit as measured by R^2 is similar for both specifications, I choose the more

[9] For a detailed treatment of fitting a power law see Hill (1975) and Clauset, Shalizi, and Newman (2009). The latter only considers the case $\omega > 1$.

Figure 2.1: Estimates of the parameter ω for the Pareto distribution fitted to the data on number of employees in the U.S. firms between 1977 and 2009: $P(L < l) = 1 - l^{-\omega}$. The estimates were obtained by OLS regressions on the log-log form of the above equation, with and without the restriction on intercept being zero. The gray dashed lines represent 95% confidence intervals.



parsimonious one. Also, the smallest firm size in the sample is one, which is consistent with (2.26).

Another concern in estimation of Pareto distribution is the upper limit of the distribution's support. Equations (2.25) and (2.26) imply an infinite support: $L \in [l_0, \infty)$. In case of a finite upper limit l_1, the expression for the CDF implied by (2.24) should be divided by the value of the CDF at l_1, yielding $P(L < l) = \frac{1-\left(\frac{l_0}{l}\right)^{\omega}}{1-\left(\frac{l_0}{l_1}\right)^{\omega}}$. However, the order of magnitude of this bound is 10^6, which makes the denominator indistinguishable from one for all practical purposes.

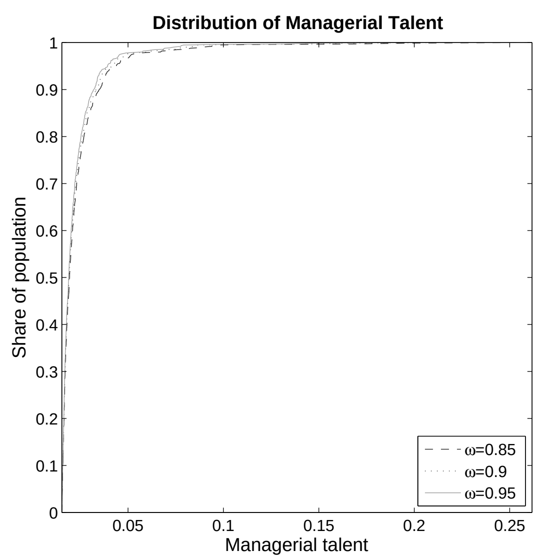
However, limits on the distribution's support cannot be ignored in the estimation of the distribution of managerial talent. By assumption, this distribution is bounded between 0 and 1. To truncate the distribution of x from above at 1, the expression for the CDF implied by (2.23), $P(X < x) = 1 - (\varkappa x^{\frac{1}{1-\gamma}})^{-\omega}$, needs to be divided by its value at 1, which is $1 - \varkappa^{-\omega}$.

This obtains the following expression for the PDF of the managerial talent:[10]

$$f(x) = \frac{\omega}{1-\gamma}\frac{\varkappa^{-\omega}}{1-\varkappa^{-\omega}}x^{\frac{-\omega}{1-\gamma}-1}$$

The last parameter that remains to be calibrated is γ. Ideally, the value would be closer to one than to zero, to limit the effect of decreasing returns to scale in the production function. Furthermore, it would ensure that income shares of labor and capital in the steady state match those found in the data. However, the choice of values is limited because the algorithm to solve the model does not converge for some values of γ, so I have chosen the optimal value of 0.7 heuristically. Figure 2.2 shows three distributions of managerial talent randomly generated for values of $\omega = 0.85, 0.9, 0.95$.

Figure 2.2: Distributions of managerial talent randomly generated for $\omega = 0.85, 0.9, 0.95$.



As a result of these limitations in the calibration process, the capital share of income produced by the model in the steady state is lower than the one estimated in the U.S.

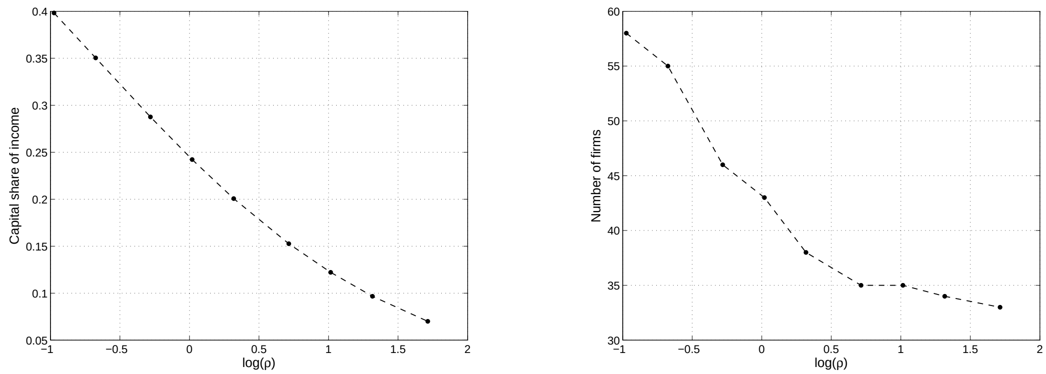
[10] To generate a random variable X from a uniformly distributed random variable U:

$$u = \frac{1-\varkappa^{-\omega}x^{\frac{-\omega}{1-\gamma}}}{1-\varkappa^{-\omega}}$$

$$x = (\varkappa^{\omega} - u(\varkappa^{\omega}-1))^{-\frac{1-\gamma}{\omega}}$$

Figure 2.3: Number of firms and capital share of income as a function of the capital-labor ratio.

(a) Capital share of income (b) Number of firms

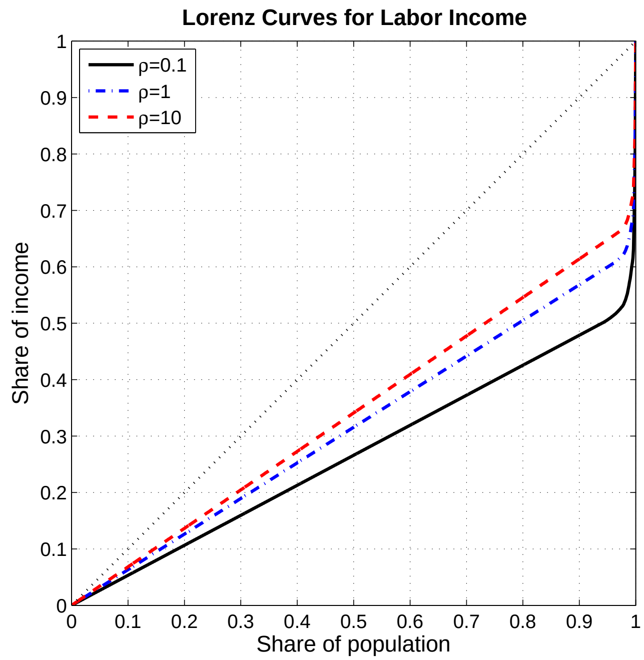


economy. Figure 2.3a shows capital share of income as a function of the log capital-labor ratio. In the steady state, the log of this ratio is close to one and corresponding capital share of income is around 12%, which is below the usual estimates of 30-35%.

Figure 2.3b shows the number of firms in the economy as a function of the log capital-labor ratio. As the Lucas model predicts, the number of firms decreases as the capital stock increases. Since the amount of labor remains fixed, this is equivalent to saying that the average firm size increases with capital stock. In the steady state, there are around 35 firms, which implies that out of 1,000 agents, 35 are employed as managers, and the rest as production workers.

The capital-labor ratio also determines the distribution of wages in the model. In Figure 2.4, the dashed curve represents the Lorenz curve for labor income when capital-labor ration $\rho = 10$, which is close to its steady state value. The long linear segment of the curve represents production workers, which are all paid the same wage. The short steep part at the top of the distribution represents the managers. A decrease in capital has two effects on the distribution: it increases inequality by reducing the wage of the production workers, and it decreases inequality at the top because more agents receive higher, managerial salaries. However, the former effect dominates and the inequality of wage distribution decreases with increase in capital.

Figure 2.4: Lorenz curves for distribution of labor income for three values of the capital-labor ratio.



2.5 Results

This section describes results and compares them with the U.S. data and the results of the benchmark KS model. The main result is the wealth distribution generated by the model proposed in this study. I evaluate it by its fit to the data, and in particular, by its fit to the upper tail of the wealth distribution in the U.S.

The wealth distribution of the model is obtained by simulating the model for 1000 periods. Aggregate shocks are generated randomly from a distribution that is similar to the distribution of expansions and recessions in the U.S. after World War II. Individual shocks are randomly generated in each period to reflect average levels of unemployment during expansions and recessions.

The Lorenz curve for the wealth distribution is shown by the solid line on Figure 2.5. Similar to the labor income distribution, there are two distinct segments; a long, linear one that represents agents who spend most periods as production workers, and a short, steep one that represents agents who spend most periods as managers.

The evaluation of the model can also be divided along these two line segments: the model fails in the first one, and succeeds in the second one. As it can be seen from Figure 2.5, the actual distribution of wealth in the U.S., which is represented by the dashed line, is extremely skewed with bottom 40% holding almost no wealth and the next 40% of the population holding around 15%. On the other end of the distribution, the top 5% hold 62% and top 1% hold almost 35% of total wealth.[11]

The chart for the simulated distribution, on the other hand, indicates fairly equal wealth distribution among a big part of the population, and large inequality only at the very top. For example, the coefficient of variation[12] for the bottom 95% agents is only 0.022. This is significantly less variation in the bottom than in the benchmark Krusell-Smith distribution, whose coefficient of variation for the same part of the population is 0.5.

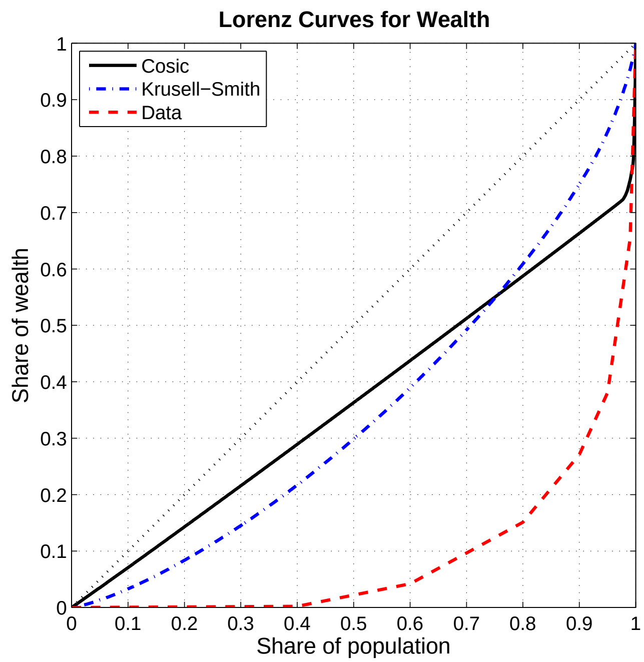
It is the upper tail, however, where this model performs well. As it can be seen in Figure 2.5, the simulated distribution tracks the actual distribution relatively well. The top 1% of the simulated population holds around 24% compared with the actual 35%, and benchmark Krusell-Smith of 4.6%. Even though the model does not fit the data exactly, it improves the fit by an order of a magnitude relative to Krusell-Smith, with some potential for further improvement.

The biggest problem with the proposed model, however, is the sensitivity of the solution algorithm to the parameter values. For some values of parameters γ and ν, the algorithm never reaches a solution. This problem limited the parameter space in the calibration stage. It also prevented some simple extensions of the model. For example, KS achieve significant improvement in the goodness of fit of the bottom of the distribution by adding a minimum income that agents receive while unemployed. When the minimum income is added to the present model, however, the solution stops converging.

[11] Data on the U.S. wealth distribution in 2007 is from Wolff (2010).

[12] The coefficient of variation is calculated as a quotient of the standard deviation and the mean.

Figure 2.5: Lorenz curves for three distributions of wealth: the one generated by the model, the one generated by the benchmark Krusell and Smith (1998) and the one estimated from the data.



Finally, I examine the relationship between the distribution of firm sizes and the wealth distribution generated by the model. Table 2.1 shows the share of wealth held by the top 1% wealthiest agents and the coefficient of variation for the bottom 95% agents at three different values of the coefficient that determines the distribution of firm sizes, ω. The values of ω included in the table cover the range of values that have occurred in the U.S. between 1977 and 2009 with 95% probability, i.e. the values within the confidence intervals in Figure 2.1. Wealth inequality is positively correlated with inequality of the distribution of firm sizes. Based on the two measures of wealth inequality, this is the case both at the top and bottom of the distribution. The two measures of inequality, however, vary little with the distribution of firm sizes.

Table 2.1: Share of wealth held by the top 1% and coefficient of variation for the bottom 95% for three values of ω.

ω	Top 1%	$CV_{0.95}$
0.85	24.4	0.023
0.90	24.2	0.022
0.95	24.2	0.021

2.6 Conclusion

My main goal for this paper was to propose a general equilibrium model that is able to reproduce the actual wealth distribution found in the data, while including the most important mechanisms that determine how wealth is distributed. This means, in particular, introducing labor income inequality as one of the main generators of wealth inequality. In addition, one of the main objectives was to explore the effect of distribution of firm sizes on the distribution of wealth.

The study has been a partial success. The constructed model includes some mechanisms important for explaining the wealth distribution, including effects of the labor income distribution and the distribution of firm sizes. The proposed method for solving the model reaches a unique solution in most cases within a reasonable amount of time. Simulation of the model generates a distribution of wealth whose upper tail matches the upper tail of the U.S. wealth distribution. The relationship between the distribution of firm sizes and the distribution of wealth has the expected sign: an increase in inequality in the former causes an increase in inequality in the latter.

On the other hand, the model fails in some respects. Two issues are particularly important. As a technical matter, the numerical algorithm for solving the model is fairly sensitive to parameter values and it does not converge for some of them. This has made the calibration of the model difficult and imprecise. As a more substantial issue, the wealth distribution obtained from the simulation follows the labor income distribution too closely. It appears that a person's wealth is almost completely determined by his or her labor income. This is particularly worrying considering that the benchmark KS model generates significantly more variation in the distribution of wealth with a simpler labor income process. Furthermore, the ability to extend the model to include the minimum income is critical in achieving realistic wealth holdings in the bottom half of the distribution.

Despite these failings, the model presented in this study shows a possible way forward in the research on wealth inequality. Previous literature has mostly focused on the role of incomplete credit markets and heterogenous preferences as generators of wealth inequality. While they certainly play a significant role, there is no doubt that one of the biggest factors in the increase in wealth inequality has been rising labor income inequality. The proposed model introduces this factor as a key determinant of wealth distribution. Moreover, it ties labor income inequality to the divergence of wages in managerial and non-managerial positions, which has been among primary drivers of rising wage inequality. Lastly, it introduces the distribution of firm sizes as one of the determinants of wage distribution, another relationship that has been observed empirically.

Further extensions to the model may include the distinction between skilled and unskilled labor, as well as the cost of capital that decreases with firm size. Each of these extensions would help achieve a more realistic distribution of wages of production workers.

Table 2.2: Estimates of the parameter ω for the Pareto distribution fitted to the data on number of employees in the U.S. firms between 1977 and 2009: $P[S < s] = 1 - \left(\frac{1}{s}\right)^{\omega}$. The estimates were obtained by OLS regressions on the log-log form of the above equation, with and without the restriction on intercept being zero.

	With Intercept					Without Intercept				
Year	ω	SE	Conf. Interval		R^2	ω	SE	Conf. Interval		R^2
1977	0.9851	0.0165	0.9477	1.0225	0.9972	0.9136	0.0124	0.8861	0.9412	0.9980
1978	0.9969	0.0162	0.9602	1.0337	0.9974	0.9131	0.0139	0.8821	0.9440	0.9975
1979	1.0048	0.0167	0.9670	1.0426	0.9972	0.9103	0.0154	0.8760	0.9445	0.9969
1980	1.0001	0.0171	0.9615	1.0387	0.9971	0.9046	0.0155	0.8700	0.9393	0.9968
1981	0.9848	0.0183	0.9435	1.0261	0.9966	0.9003	0.0144	0.8683	0.9323	0.9972
1982	0.9749	0.0180	0.9341	1.0157	0.9966	0.8919	0.0141	0.8604	0.9234	0.9972
1983	0.9599	0.0201	0.9144	1.0054	0.9956	0.8883	0.0132	0.8589	0.9177	0.9976
1984	0.9769	0.0178	0.9368	1.0171	0.9967	0.8975	0.0136	0.8671	0.9278	0.9975
1985	0.9815	0.0176	0.9415	1.0214	0.9968	0.8953	0.0145	0.8630	0.9275	0.9971
1986	0.9759	0.0184	0.9344	1.0175	0.9965	0.8900	0.0146	0.8576	0.9225	0.9971
1987	0.9704	0.0168	0.9323	1.0085	0.9970	0.8848	0.0142	0.8532	0.9165	0.9972
1988	0.9850	0.0146	0.9521	1.0179	0.9978	0.8912	0.0149	0.8580	0.9244	0.9969
1989	0.9735	0.0141	0.9416	1.0054	0.9979	0.8810	0.0147	0.8484	0.9137	0.9970
1990	0.9796	0.0144	0.9469	1.0122	0.9978	0.8843	0.0151	0.8507	0.9180	0.9968
1991	0.9825	0.0135	0.9519	1.0131	0.9981	0.8868	0.0150	0.8533	0.9202	0.9969
1992	0.9826	0.0120	0.9556	1.0097	0.9985	0.8875	0.0147	0.8547	0.9203	0.9970
1993	0.9817	0.0133	0.9515	1.0119	0.9982	0.8899	0.0145	0.8576	0.9221	0.9971
1994	0.9788	0.0129	0.9496	1.0079	0.9983	0.8878	0.0143	0.8560	0.9196	0.9972
1995	0.9780	0.0128	0.9491	1.0070	0.9983	0.8863	0.0144	0.8542	0.9183	0.9971
1996	0.9759	0.0130	0.9464	1.0054	0.9982	0.8841	0.0144	0.8519	0.9162	0.9971
1997	0.9806	0.0122	0.9531	1.0081	0.9985	0.8847	0.0149	0.8516	0.9178	0.9969
1998	0.9772	0.0127	0.9484	1.0060	0.9983	0.8825	0.0148	0.8496	0.9154	0.9969
1999	0.9693	0.0132	0.9394	0.9992	0.9981	0.8766	0.0146	0.8441	0.9091	0.9970
2000	0.9665	0.0140	0.9348	0.9981	0.9979	0.8724	0.0149	0.8393	0.9055	0.9968
2001	0.9646	0.0145	0.9318	0.9975	0.9977	0.8710	0.0149	0.8379	0.9042	0.9968
2002	0.9710	0.0129	0.9417	1.0002	0.9982	0.8751	0.0150	0.8417	0.9084	0.9968
2003	0.9748	0.0146	0.9417	1.0079	0.9978	0.8802	0.0150	0.8467	0.9137	0.9968
2004	0.9837	0.0132	0.9538	1.0135	0.9982	0.8840	0.0155	0.8495	0.9186	0.9966
2005	0.9886	0.0122	0.9609	1.0163	0.9985	0.8879	0.0155	0.8533	0.9225	0.9966
2006	0.9841	0.0124	0.9560	1.0122	0.9984	0.8845	0.0154	0.8502	0.9188	0.9967
2007	0.9722	0.0121	0.9447	0.9996	0.9984	0.8766	0.0148	0.8437	0.9096	0.9969
2008	0.9669	0.0132	0.9371	0.9966	0.9981	0.8749	0.0145	0.8427	0.9071	0.9970
2009	0.9365	0.0196	0.8922	0.9807	0.9957	0.8593	0.0137	0.8287	0.8899	0.9972

Chapter 3

Income Inequality and Household Debt

3.1 Introduction

It is mostly uncontroversial that the financial crisis of 2008 in the U.S. was caused primarily by an unsustainable amount of debt. It is less obvious how this debt was created. Why were borrowers willing to take on more debt than they could pay back and why were lenders willing to lend as much? Liberalization of financial markets and innovations in financial instruments certainly played a role on the supply side, but what is the source of the increase in demand for credit? One candidate for such a role that recently started receiving more attention is income inequality. This paper looks at this possibility.

The motivation for looking at income inequality as a cause of a financial crisis stems from the view that higher income inequality suppresses consumption of low-income households and that the households whose incomes fall seek to maintain the same level of consumption by increasing borrowing. Supply of credit to low-income households is normally limited, but under certain conditions it can increase. If income inequality keeps increasing over a long period of time, lending to consumers may become necessary to maintain economic growth. However, it also leads to unsustainable levels of debt, high loan-default rates and, eventually, a financial crisis.

The goal of this paper is to study long-run dynamics between income inequality and level of debt. The paper follows up on a study by Bordo and Meissner (2011), which finds no significant relationship between changes in income inequality and the growth rate of debt in a cross-country panel. To examine long-run effects, I use level variables to examine cointegration between debt level, GDP, and income inequality measured as a fraction of income received by the top 1% of income distribution. On a panel of fourteen OECD countries, I use Westerlund (2007) and Pedroni (1999) tests for panel data to examine cointegration between debt-to-GDP ratio and top 1% income share, and Johansen (1988) to test for cointegration between debt, GDP, and top 1% income share for individual countries. Results of all tests indicate no cointegration.

The following section lays out the theory that supports the link between income distribution and debt. It presents two different mechanisms that might play a role in enabling that link and reviews some previous literature. The baseline econometric model and its extensions are presented in Section 3.3, data in Section 3.4, and estimation results in Section 3.5. Concluding remarks are in Section 3.6.

3.2 Theory and Literature

This section presents a theoretical view of the relationship between demand for and supply of credit, and income distribution, and previous literature related to this topic. It first lays down conditions for income inequality to raise demand for credit. Then it presents two competing views on how supply of credit can rise in response to rising income inequality.

The idea that growing income inequality can raise demand for credit dates back to the Great Depression. It was dormant for a while, but it has been revived in the aftermath of the financial crisis of 2008. The idea follows from the assumption that, when higher income inequality reduces income of low-income households, these households resist reduction in consumption by dissaving. If a household's assets are already zero or negative, dissaving means more borrowing.

According to the life-cycle consumption theory, however, this is true only if the increase in income inequality, and therefore, the fall in income of low-income households, is temporary. In that case, these households smooth their consumption by reducing their saving rate. Those with zero or negative assets increase their borrowing. If, on the other hand, the fall in income is permanent, low-income households reduce their consumption without an increase in borrowing.

The problem is that data for some countries show both income inequality and household debt rising over long periods of time. For example, income inequality and household debt in the United States grew for more than three decades prior to the financial crisis. This kind of pattern can hardly be considered temporary. How to reconcile data with the theory?

There are several mechanisms that could influence a household to deviate from the behavior predicted by the life-cycle consumption theory. First, some low-income households might be borrowing constrained. The only reason they don't borrow more is that credit market is imperfect and lenders ration credit. But if credit supply increases and the borrowing constraint relaxes, these households will increase their borrowing. There is plenty of evidence that a fraction of households is borrowing constrained, both from microeconomic panel data studies (Hall and Mishkin, 1982; Zeldes, 1989) and studies on aggregate data (DeLong and Summers, 1986; Campbell and Mankiw 1989).

Second, households might perceive a permanent shock to income distribution as temporary. This argument is usually not accepted because it contradicts the notion of a rational agent. Even when a shock is unexpected, the agent should be able to learn its nature quickly. But in the case of changes in income distribution this argument may be more sensible. In the United States, the median real wage stagnated between 1970 and 2005, while that for the bottom 10% of the income distribution decreased significantly (see Heathcote, Perri, Violante, 2010). At the same time, a tremendous technological progress has driven growth in labor productivity. A rational agent would have a hard time reconciling these two facts and could be forgiven for believing that such a development is a temporary phenomenon.

Two more factors that have similar effects on household behavior are habit formation and relative income, both of which can be traced back to Duesenberry (1949). The former gets rid of time-separability of household utility function and posits that current household's utility depends not only on its current but also on its past consumption. The latter adds relative income as a determinant of household consumption. A household would consume a different amount of goods if income distribution changed, even if the income process of that household remained the same.

Given that low-income households increase their demand for borrowing in the face of falling income, it is important to see what happens with the supply of loanable funds. If supply is inelastic, or if it does not shift, an increase in demand alone would have little impact on the total amount of credit. In the aftermath of the financial crises of 2008, two competing narratives have emerged.

On one side, Rajan (2011) views the government as the main link between income inequality and credit. He argues that rising income inequality creates pressure on political elites to redistribute income. Political constraints prevent them from directly intervening in income distribution through fiscal policy. Instead, they encourage, and in some cases force, lenders to increase lending to low-income households.[1] In this model, the government plays a central role. Even though demand for more borrowing exists, it would have not materialized if politicians hadn't forced the banks to increase supply. Lenders play a passive role.

The other narrative follows from the Keynesian concept of underconsumption and overinvestment. It relies on an additional assumption, made initially by Keynes (1937) and later confirmed by Dynan, Skinner and Zeldes (2004), that, in absence of credit, rising income inequality reduces total consumption.[2] Because lower total consumption means higher total saving and thus an increase in the supply of loanable funds, this process also results in a higher rate of investment. Falling consumption and rising investment eventually lead to a gap between aggregate demand and aggregate supply.

This gap can be closed in two ways: either output must fall, or consumption must rise. But consumption can rise only if low-income households increase their consumption; high-income households are already consuming at their optimal point. And low-income households can increase their consumption only if they increase borrowing. Normally, they are borrowing constrained, but with aggregate supply already exceeding demand, returns to investment fall and loanable funds get redirected to their alternative use, loans to low-income households.

It is important to note that, in this model, government plays no role. The rising household debt is entirely a result of market forces. As total consumption falls because of rising income inequality, the rate of return to investment falls with it, and a competing use of

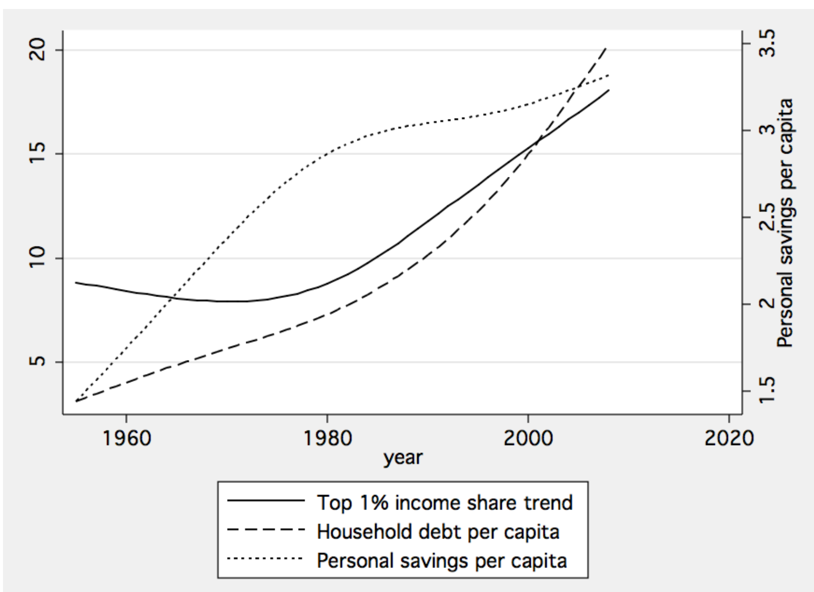
[1] E.g. in the U.S.A, Community Reinvestment Act of 1977. prohibits lending institutions from denying their service to low-income neighborhoods in the areas in which they operate.

[2] This is a corollary of marginal propensity to consume being a decreasing function of income.

loanable funds, lending to low-income households, becomes more attractive. This causes consumption to grow again, the gap between aggregate demand and supply closes, and returns to investment recover.

In this model, credit is the mechanism that allows aggregate demand to keep rising at the same rate as aggregate supply. Figure 3.1 shows U.S. data that are consistent with this mechanism. It shows trends[3] of top 1% income share, as a measure of income inequality, household debt per capita, and household savings per capita for the U.S. between 1955 and 2008.[4] A surge in inequality starts in the late 1970s. Immediately following it, the trend in personal savings experiences a drop, while soon afterwards the trend in household debt increases.

Figure 3.1: Household debt and savings per capita in thousands of 1984 dollars, and top 1% income share, U.S.A.



While a number of authors have promoted this view of the cause of the financial crises of 2008, the most convincing formal argument was put forward by Kumhof and Ranciere

[3]Created with the Hodrick-Prescott filter from annual data.

[4]Data for household debt, savings, population and CPI are obtained from the Fed's website. Data for top 1% income share from The World Top Incomes Database.

(2010). They show a plausible mechanism of a credit bubble in which income inequality plays a major role. The population of their DSGE model consists of workers, who provide labor inelastically and receive only labor income, and investors, whose income comes from investment in physical capital and deposits. Workers negotiate their wage with investors. Its value is proportional to workers' negotiating power. Investment in physical capital yields marginal product of capital, and deposits yield the prevailing interest rate. Deposits can be used to make loans to workers.

In the steady state, workers consume all of their income and investors invest all of their savings in capital. There is no debt. A negative shock to the negotiating power of workers causes income inequality to start rising. Real wages start falling and workers' consumption follows. Investors' income and consumption grow. In the aggregate, however, consumption falls because the fall of workers' consumption is greater than the rise of investors' consumption. The associated rise of aggregate saving raises investment, but as workers' consumption keeps falling due to increasing inequality, an increasing part of saving is diverted to consumer loans to workers. This creates a dynamic that keeps the economy growing, but only because a large part of its population keeps borrowing to maintain their consumption. Eventually, however, the level of debt becomes unsustainable, and growing default rates lead to a financial crisis.

Another study that obtains similar results is Iacoviello (2008). It adds a financial accelerator by introducing real estate as an asset that households use as insurance against idiosyncratic shocks, but also from which they draw utility and which they can use as borrowing collateral subject to a maximum loan-to-value ratio. This ratio is a key element that represents the economy-wide supply of credit. It is introduced as an exogenous shock and estimated from the U.S. data between 1963 and 2003. Income inequality is introduced as an exogenous, idiosyncratic shock, also estimated from the data. The third shock is the standard aggregate productivity shock. The paper estimates the effects that these three shocks have on trend and cyclical components of debt in the U.S. The main finding is that income inequality shock accounts for almost all of the trend in debt, and the other two shocks account for the cyclical component.

A different, non-structural approach, however, yields different results. A paper by Bordo and Meissner (2011) uses a panel of 14 countries to study the relationship between income inequality and the likelihood of a financial crises. The question is answered in two stages. First, the authors estimate the effect of changes in income distribution on the rate of credit growth. Then, they estimate the effect that credit growth has on likelihood of a financial crises. While they find that a higher rate of credit growth increases the probability of a financial crises, they find no significant effect of income distribution on rate of credit growth.

Another way to look at determinants of credit growth is an event study. A large upward deviation of credit from the trend is identified as a credit boom. The analysis focuses on finding variables which deviate from their trends within some window of time before or after a credit boom, indicating a causal relationship. This approach is proposed by Gourinchas, Valdes, and Landerretche (2001) in their paper on determinants of credit booms. They analyze a wide range of macroeconomic variables in 91 countries over the period 1960-1996. For each variable, they calculate its deviation from the long-run trend in the period leading to a credit boom. They find that credit booms start while investment booms are already under way. Investment grows further after the onset of a credit boom. Consumption, on the other hand, is usually below its trend when a credit boom starts. It grows moderately toward its trend throughout the credit boom. Unfortunately, it is impossible to say whether the excess of credit is used for financing consumption.

One problem in Gourinchas Valdes, and Landerretche (2001) lies in the definition of the trend as rolling, backward looking, which means that the trend at any point is determined only by the past data and not by the whole available sample. While such a definition of trend may represent the policy maker's view better and may be more useful in analyzing potential policy responses, it is misplaced in a causal analysis. This is reflected in the paper's identification of credit booms. For example, the U.S. has no credit booms between 1960 and

1996, while Syria experiences one long boom for the most of the sampled period. Another potential problem in the methodology concerns the use of the credit-to-GDP ratio as a measure of credit, as it may not be able to distinguish structural changes in the financial system and deviations from the trend.

These issues are addressed in Mendoza and Terrones (2008). They use the conventional Hodrick-Prescott filter with smoothing factor 100 to decompose variables into the trend and short-run fluctuations. As a measure of credit, they use real credit per capita, rather than the credit-to-GDP ratio. They also use a country-specific definition of the threshold for detection of booms, which is more accommodating toward heterogeneity of financial systems. This improves identification of credit booms.

However, a major flaw of this approach remains: it is focused only on short-run deviations from the trend.[5] The problem that the present paper studies is the trend itself. Do changes in income distribution affect it and can it lead to an unsustainable level of debt?

3.3 Econometric Model

To estimate relationship between income inequality and debt, Bordo and Meissner (2011) proposes a dynamic panel data model with fixed country and time effects:

$$\Delta L_{i,t} = \alpha \Delta L_{i,t-1} + \Delta X'_{i,t-1}\beta + \mu_i + \nu_t + \varepsilon_{i,t} \tag{3.1}$$

where $L_{i,t}$ is log of total outstanding household debt i in period t, $X_{i,t-1}$ matrix of covariates that, in various specifications, includes the top 1% income share as a measure of income inequality, the log of real GDP, an index of investment relative to the price level, the log of the real money supply with money measured according to the M1 definition, and a short-term nominal interest rate. Because all monetary variables enter the model in logarithmic

[5]When Mendoza and Terrones's (2008) method is applied to my data, the United States appears to have no credit booms between 1990 and 2008.

form, differences of those variables represent variables' growth rates. The terms μ_i and ν_t represent country and time fixed effects, respectively. The model is estimated for one-year and five-year periods.[6]

This model measures the effect that a change in income inequality may have on the growth rate of total household debt. The nature of the dependent variable justifies the use of a dynamic panel model. Debt has a unit root for most countries and its growth exhibits autoregressive properties. Debt is, to a large extent, determined by country's financial system and regulations, which is represented by the country fixed effects. Time fixed effects account for changes in the world economy over time.

Bordo and Meissner's model, however, leaves out some potentially important parts of the relationship between debt and income inequality. The theory presented in Section 3.2 implies that an increase in income inequality moves aggregate demand below its steady state value, and an increase in credit can bring it back up. Credit adjusts to maintain the long-run equilibrium of aggregate demand and supply. But model (3.1) is not capable of capturing such a relationship. By including only differences, it ignores the long-run relationship between the variables and focuses only on the relationship between their changes.

To see why this may be important consider a case where aggregate demand is above its steady state value. An increase in income inequality, according to the theory, affects aggregate demand negatively, but as long as demand is above supply, there would be no change in credit. If anything, credit would be falling to push demand further down toward the steady state. However, model (3.1) cannot capture such dynamics because it imposes a unidirectional relationship between changes in credit and income inequality.

To take this behavior into account, a model needs to be able to represent the long-run equilibrium relationship between debt, GDP and income inequality. A rise in income inequality that moves the economy out of the long-run equilibrium should cause an increase in credit, a decrease in GDP, or some combination thereof. A model that can represent such

[6]Results that Bordo and Meissner (2011) reports for five-year period samples are from regressions that use contemporaneous, rather than lagged, regressors.

a relationship is the error-correction model. Three variations of such a model for panel data, described in Pesaran, Shin and Smith (1999), can be represented by the equation:

$$\Delta L_{i,t} = \phi_i(L_{i,t-1} - X'_{i,t-1}\beta_i + \alpha_i) + \sum_{j=1}^{p} \gamma_j \Delta L_{i,t-j} + \sum_{j=0}^{q} \Delta X'_{i,t-j}\delta_{ij} + \mu_i + \nu_t + \varepsilon_{i,t} \quad (3.2)$$

Comparing with (3.1), the main difference is the long-run term, the first term on the right hand side. The expression in the parentheses represents the long-run equilibrium relationship between $L_{i,t}$ and $X_{i,t}$. When the variables are not in the long-run equilibrium, the speed of adjustment of the dependent variable that brings the system back to the equilibrium is defined by the error-correction parameter ϕ_i.

The above model nests model (3.1) and three others. It reduces to (3.1) for $\phi_i = 0$, $p = 1$, $q = 1$ and $\delta_0 = 0$. When $\phi_i \neq 0$ and $\beta_i \neq 0$, it represents a family of error correction models that differ in degree of heterogeneity they allow across countries. The panel estimate of a heterogenous parameter is calculated as the mean of cross-country parameters. The three models discussed in Pesaran, Shin and Smith (1999) are: *dynamic fixed effects* (DFE), which imposes the same coefficients for all countries; *pooled mean group estimator* (PMG), which allows heterogeneity for the error correction term coefficient ϕ_i and other short-run coefficients on covariates δ_j, while imposing homogeneity on coefficients β_i that define the long-run relationship between the variables; and *the mean group estimator* (MG), which imposes no homogeneity restrictions. I consider only the last one.

Error-correction models for panel data impose one restriction that is not present when dealing with an individual country. Because variables are expressed in levels, rather than in relative changes as in (3.1), care needs to be taken that variables and parameters in the model are compatible across the panel. A common source of incompatibility is in cross-country difference in variables' units. For example, in the data used in this paper, debt is expressed in terms of national currencies. Therefore, the unit of the coefficient β_i for the top 1% income share variable, national currency per one percentage point of top 1% income share, is different for each country. Clearly, β_is for different countries cannot be compared or combined (e.g. by computing their mean) without further transformations.

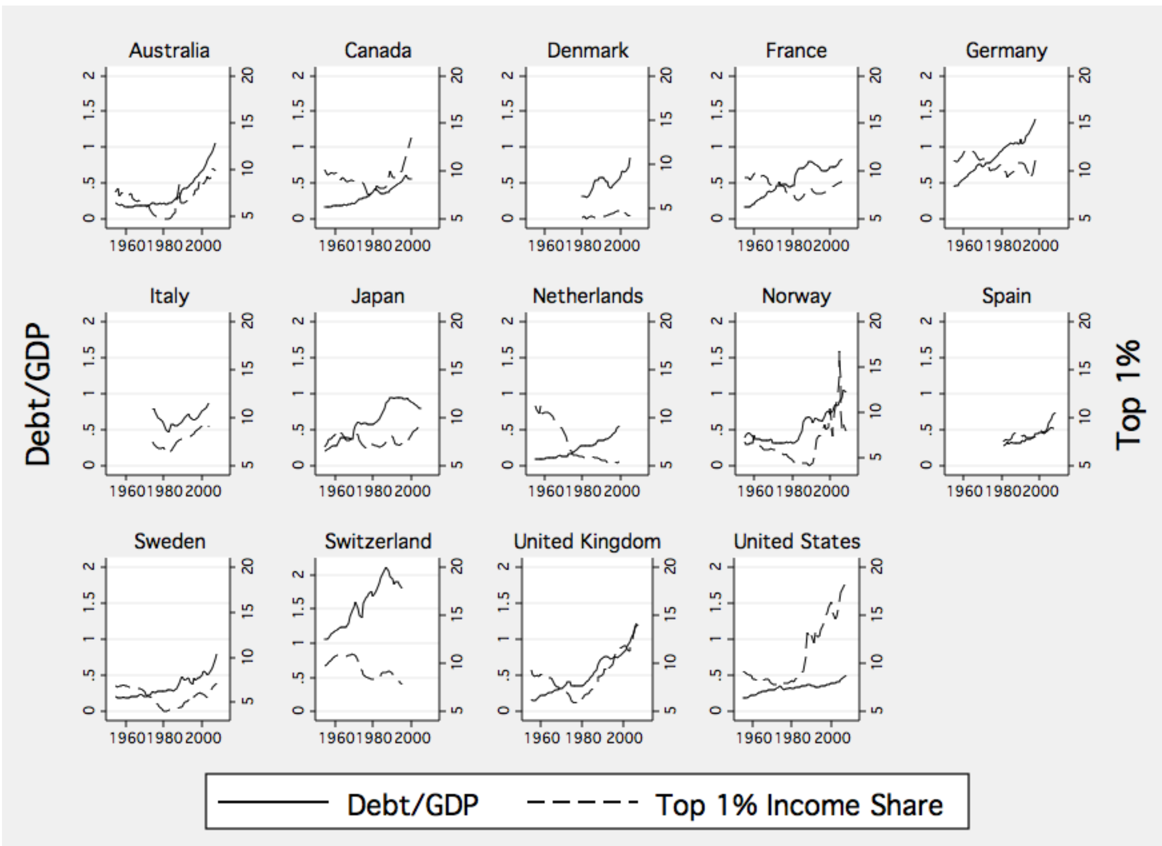
To deal with this restriction, I use debt-to-GPD ratio as the dependent variable in model (3.2), and the only covariate on the right-hand side is top 1% income share. This is an imperfect solution, since it imposes a one-to-one restriction on the long-run relationship between debt and GDP. To see this, note that a cointegrating vector between debt-to-GDP ratio and top 1% income share implies that a constant top 1% income share requires debt-to-GDP ratio to remain constant as well. To verify how restrictive this model is, I also estimate (3.2) for each country individually. Finally, in one variation of the model I include investment-to-GDP ratio as a covariate.

Another problem in modeling the relationship between income inequality and debt is that it is asymmetrical. While theory offers a plausible explanation of upward adjustment of credit to an increase in income inequality, it is not clear that this relationship should exist when income inequality moves in the opposite direction. In general, the process of deleveraging (especially after a financial crises) is slower than the process of credit accumulation. Some kind of of hysteresis in this relationship can be expected.

3.4 Data

The data used in this study is a subset of the panel dataset from Bordo and Meissner (2011). Their data on debt, GDP and investment originate from Schularick and Taylor (2011) and data on top 1% income share from Alvaredo et al. (2012). Schularick and Taylor (2011) define the debt variable as the “end-of-year amount of outstanding domestic currency lending by domestic banks to domestic households and non-financial corporations (excluding lending within the financial system)”. It also excludes all credit card debt. Banks include saving and postal banks, credit unions, mortgage associations, and building societies, but not brokerage houses, finance companies, insurance firms, and other financial institutions. Since the goal of this study is to estimate the effect of income inequality on household debt, a variable that excludes credit card debt and includes loans to firms is less than perfect. Ideally, the debt variable should include only total debt owed by households. The data for debt/GDP and top 1% income share is plotted in Figure 3.2 and shown in Table 3.1.

Figure 3.2: Debt/GDP (left y-axis) and Top 1% income share (right y-axis).



The panel is unbalanced, with individual series length in the range between 26 for Denmark and 54 for Norway, Sweden, and the U.S.A. Even though the original data contains long time series, going back to the 1800s for some countries, I use only observations since 1955. The time series before 1955 contain many large gaps that are difficult to deal with in error-correction models. Also, as Schularick and Taylor point out, transitions between historical stages in the evolution of the world financial system most likely manifest as structural breaks in the data. Dealing with them requires estimation of additional parameters.

The gain from longer time series would be canceled by the loss in degrees of freedom.

Table 3.1: Summary statistics.

	Time		Top 1%				Debt/GDP			
Country	Years	Obs	Mean	SD	Min	Max	Mean	SD	Min	Max
Australia	1955-2007	53	6.8	1.5	4.6	10.1	0.4	0.2	0.2	1.1
Canada	1955-2000	46	9.3	1.2	7.6	13.6	0.3	0.1	0.2	0.6
Denmark	1980-2005	26	4.2	0.2	3.8	4.7	0.5	0.1	0.3	0.9
France	1955-2005	52	8.4	0.8	7.0	9.9	0.5	0.2	0.2	0.8
Germany	1955-1998	44	10.8	0.8	9.2	12.2	0.9	0.2	0.5	1.4
Italy	1974-2004	31	7.7	0.9	6.3	9.4	0.7	0.1	0.5	0.9
Japan	1955-2005	51	7.7	0.6	6.8	9.2	0.6	0.2	0.2	0.9
Netherlands	1955-1999	45	7.4	2.1	5.2	11.3	0.2	0.1	0.1	0.6
Norway	1955-2008	54	6.7	2.2	4.2	16.8	0.5	0.2	0.3	1.1
Spain	1981-2008	28	8.2	0.4	7.5	9.0	0.4	0.1	0.3	0.7
Sweden	1955-2008	54	5.6	1.0	4.0	7.1	0.3	0.1	0.2	0.8
Switzerland	1955-1995	41	9.5	1.0	7.8	11.0	1.6	0.3	1.1	2.2
United Kingdom	1955-2007	53	8.8	2.2	5.7	14.6	0.5	0.3	0.1	1.2
United States	1955-2008	54	10.9	3.4	7.7	18.3	0.3	0.1	0.2	0.5
Total	-	632	8.1	2.4	3.8	18.3	0.5	0.4	0.1	2.2

3.5 Estimation

To establish the baseline and to verify the results of Bordo and Meissner (2011) I first estimate the model given by equation (3.1) on the original dataset for five-year time intervals. Results of regressions for five-year intervals with contemporary covariates are reported in Table 3.2. The effect of income inequality appears to be weakly significant only in the first specification when top 1% is the only regressor. When GDP is included, income inequality loses significance. When the covariates are lagged, no variable is significant at 5% significance level. A more detailed analysis of the results of Bordo and Meissner (2011) reveals some inconsistencies between the model and their estimates, but contributes little to the current discussion; so, it is relegated to the Appendix.

Table 3.2: Debt growth and contemporaneous covariates for a five-year time period on sample 1920-2005.

VARIABLES	(1)	(2)	(3)	(4)
$\Delta\ln(\text{Debt}_{t-1})$	0.156	0.0968	0.164	0.138
	(0.129)	(0.102)	(0.0989)	(0.105)
$\Delta\text{Top }1\%_t$	0.0486*	0.00891	0.0523	0.0404
	(0.0259)	(0.0267)	(0.0348)	(0.0302)
$\Delta\ln(\text{Real GDP}_t)$		1.590***	1.605***	1.188***
		(0.339)	(0.311)	(0.405)
$\Delta\text{Top }1\%_t \times \ln(\text{Real GDP}_t)$			-0.472*	-0.354
			(0.240)	(0.216)
$\Delta\text{Interest rate}_t$				0.316
				(0.862)
$\Delta\text{I/GDP}_t$				0.140
				(0.582)
$\Delta\ln(\text{Real Money}_t)$				0.512***
				(0.173)
Observations	105	105	105	105
R-squared	0.465	0.569	0.609	0.664
Number of id	14	14	14	14

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Next, I move to the error-correction model defined by equation (3.2). As a first step, it is important to establish that the variables of interest are I(1). To do that, I perform two unit root tests for panel data: Maddala and Wu (1999)[7] and Pesaran (2007).[8] Both tests estimate the following model:

$$\Delta y_{i,t} = \alpha_i + \rho_i y_{i,t-1} + \sum_{j=1}^{q} \theta_{ij} \Delta y_{i,t-j} + \varepsilon_{i,t} \tag{3.3}$$

[7]Implemented in Stata program xtfisher.
[8]Implemented in Stata program pescadf.

where $y_{i,t}$ is the variable being tested for stationarity. The null-hypothesis for both tests is that the time series is non-stationary, $H_0 : \rho_i = 0$ for all i. The tests estimate (3.3) for each i, then combine the test statistics. The main advantage of Pesaran (2007) is that it allows for cross-sectional interdependence, whereas Maddala and Wu (1999) assume independent individual time series.

Table 3.3: P-values for unit root tests, annual data, 1955-2008.

Test	q	Top 1%	Debt/GDP	Inv/GDP
Maddala, Wu (1999)	0	0.989	1.000	0.412
	1	0.953	1.000	0.000
	2	0.975	1.000	0.011
Pesaran (2007)	0	0.472	1.000	0.028
	1	0.433	0.998	0.000
	2	0.872	1.000	0.002

P-values for both tests for top 1% income share, debt-to-GDP ratio and investment-to-GDP ratio are shown in Table 3.3 for $q = 0, 1, 2$. The null of unit root cannot be rejected for top 1% income share and debt-to-GDP ratio at any common significance level for any number of lags. For investment-to-GDP ratio, however, the null can be rejected at 5% significance level in all cases except for Maddala and Wu test with zero lags. Such a strong rejection of unit root for investment is surprising considering that a whole strand of literature on the Feldstein-Horioka puzzle (e.g. Coakley, Kulasi, and Smith (1996), Coakley and Kulasi (1997), Ho (2002), Pelgrin and Schich (2008)) finds the investment rate being I(1) and cointegrated with saving.

To test for cointegration, I use three different methods: the Westerlund (2007) and Pedroni (1999, 2004) cointegration tests for panel data, and the Johansen (1988) cointegration test for individual countries.

The test proposed by Westerlund (2007)[9] relies on using the mean group estimator to es-

[9] Persyn and Westerlund (2008) implement the test as a Stata program xtwest.

timate equation (3.2) under the assumption of heterogenous coefficients, and testing whether the coefficient for the error-correction term, ϕ, equals zero. Unlike Pedroni and other residual-based tests that perform a unit root test on the residual of a regression on the cointegration term and, in doing so, impose the common factor restriction (see Kremers, Ericsson and Dolado, 1992), error-correction based tests do not impose such a restriction. This feature gives additional power to error-correction based tests. An additional advantage of the Westerlund test is that it can accommodate for cross-sectional dependence in the data by using bootstrap.

However, the Westerlund (2007) test is more restrictive with respect to the nature of the error correcting relationship than the one by Pedroni (1999). The key assumption that has to be satisfied is that regressors in $\mathbf{X}$ are weakly exogenous.[10] Only the dependent variable should be adjusting. If regressors adjust too, the test loses its power. It is difficult to say whether this assumption is too restrictive. According to the theoretical framework presented in the previous chapter, it is not. The income distribution changes exogenously and credit responds to it. However, it is difficult to completely eliminate possibility of causality in the opposite direction. If it exists, Pedroni test would be more appropriate.

Another important restriction that applies to most cointegration tests for panel data is the assumption of cross-sectional independence, or that $\nu_t = 0$ in model (3.2). If this assumption does not hold, the omitted common factor affects the estimate of the error-correction relationship. For the panel used in the present paper, this assumption is almost certainly too strong. The interconnectedness of economies through trade and capital flows makes the presence of a common factor highly probable. A statistical test for cross-sectional independence based on Pesaran (2004)[11] confirms that to be the case. To deal with cross-sectional interdependence, I demean the time series variables. Variables are expressed as deviations from their cross-sectional means in each period: $\tilde{x}_{i,t} = x_{i,t} - \bar{x}_t$, where $\tilde{x}_{i,t}$ is a deviation of

[10] In fact, strict exogeneity is assumed, but Westerlund (2007) finds that weak exogeneity can be accommodated by including leads of the regressors, in addition to lags.

[11] The test is implemented in Stata program xtcd.

variable $x_{i,t}$ from its cross-sectional mean $\bar{x}_t$. An alternative way to deal with cross-sectional interdependence that is available in the Westerlund (2007) test is bootstrapping.

The Westerlund (2007) test constructs four test statistics – two *panel* and two *group mean* statistics – to test if the cointegration parameter ϕ_i is zero. All four statistics, under the null hypothesis of no cointegration ($\phi_i = 0$), converge to normal distributions. The difference between the panel and the group mean statistics is in the alternative hypothesis. The alternative hypothesis for the panel statistics assumes that the error correction coefficient is not zero and that it is the same for all countries: $H_1 : \phi_i = \phi < 0$. The alternative hypothesis for the group mean statistics is less restrictive: $H_1 : \phi_i < 0$. A rejection of the null in this case implies cointegration for at least one country. While the latter hypothesis is less restrictive than the former, it does not follow that the test based on the group statistics has more power than the one based on the panel statistics. Westerlund (2007) states that the power of the two tests cannot be compared because of inability to obtain the joint rate of divergence under the two alternative hypotheses.

I perform the Westerlund (2007) test on two specifications. One tests cointegration between debt-to-GDP ratio and top 1% income share (model (1) in Table 3.4); the other tests cointegration between debt-to-GDP ratio, top 1% income share, and investment-to-GDP ratio (model (2) in Table 3.4). To deal with cross-country dependency, I use bootstrap and demeaned variables. In each test, the number of lags and leads[12] of the differences of the regressors for each country is selected by the AIC criterion to be either one or two. Where bootstrapping is used, the number of replications is 500. The top panel in Table 3.4 shows p-values for one-sided tests based on the four test statistics. The power of these tests as reported by Westerlund (2007) for the true value of $\phi = -0.03$, $N = 10$ and $T = 100$ at the 5% significance is fairly high for P_τ (79-98%), P_α (57-86%), and G_τ (61-88%), but low for G_α (12-56%).

The Pedroni (1999, 2004) test estimates equation (3.2) and computes regression residuals.

[12]Leads are used when the regressors are weakly but not strictly exogenous.

Based on the residuals, it calculates seven different statistics that are used to test the residuals for unit root. The four panel statistics are based on pooling along the within dimension, and the three group statistics are based on pooling along the between dimension. One from each category is a panel analogue of the augmented Dickey-Fuller t-statistics; others are non-parametric statistics. P-values based on these seven statistics for tests on demeaned variables, with and without investment-to-GDP ratio are shown in the bottom panel of Table 3.4. The power of the tests is highly dependent on the value of T and proximity of the residual time series to unit root. For $N = 20$, $T = 50$, and the significance level of 5%, Pedroni (2004) reports power between 90 and 100% for all but the G_ρ test, if the coefficient of autoregression is $\rho = 0.9$. However, the power drops dramatically for $\rho = 0.95$ to less than 30% for G_ρ and 50-60% for other tests.

Table 3.4: P-values for panel cointegration tests.

		Demeaned		Bootstrap	
		(1)	(2)	(1)	(2)
Westerlund (2007)	G_τ	0.440	0.976	0.998	1.000
	G_α	0.854	0.995	1.000	1.000
	P_τ	0.283	0.999	0.954	1.000
	P_α	0.474	0.990	0.958	1.000
Pedroni (1999,2004)	G_ρ	0.840	0.987		
	G_t	0.502	0.933		
	G_{ADF}	0.014	0.455		
	P_v	0.337	0.591		
	P_ρ	0.463	0.856		
	P_t	0.321	0.784		
	P_{ADF}	0.101	0.462		

(1) Credit/GDP, Top 1% Income Share.
(2) Credit/GDP, Top 1% Income Share, Investment/GDP.

It can be seen from Table 3.4 that Westerlund (2007) and Pedroni (1999, 2004) tests overwhelmingly indicate no cointegration between the variables being analyzed. In only one case – the Pedroni group ADF test statistic – the null hypothesis of no cointegration between debt-to-GDP ratio and top 1% income share is rejected at 5% significance level. When investment-to-GDP ratio is part of the model, all test statistics fail to reject the null of no cointegration.

To get a more detailed picture of the relationship between debt and income distribution for individual countries, I use the Johansen (1988) cointegration test for each country separately. Individual tests allow use of the level of debt, rather than debt-to-GDP ratio, which removes the restriction on the long-run relationship between debt and GDP. Table 3.5 shows P-values for Johansen cointegration tests based on maximum eigenvalue of the parameter matrix of the underlying VAR model. The test here is for cointegration between real debt, real GDP, and top 1% income share, where real debt and real GDP are entered in logarithmic form.

Table 3.5: P-values for the Johansen eigenvalue cointegration test between logarithm of real credit, logarithm of real GDP and top 1% income share.

Country	No CI	At most 1 CI vector	At most 2 CI vectors
Australia	0.0801	0.1719	0.2463
Canada	0.1886	0.4567	0.8788
Denmark	0.0023	0.4209	0.1680
France	0.1751	0.4948	0.5072
Germany	0.0061	0.0301	0.9024
Italy	0.0001	0.1252	0.0605
Japan	0.2112	0.1463	0.7753
Netherlands	0.0007	0.0413	0.4507
Norway	0.7646	0.5811	0.1485
Spain	0.4417	0.5175	0.1456
Sweden	0.1045	0.5146	0.3380
Switzerland	0.0042	0.2449	0.0106
United Kingdom	0.4881	0.2568	0.6097
United States	0.3935	0.5987	0.3111

The null-hypothesis of no cointegration is rejected at 5% significance level for five countries: Denmark, Germany, Italy, the Netherlands, and Switzerland. For three of them (Denmark, Italy and Switzerland), the test detects at least one cointegrating vector, and for the other two, it detects two cointegrating vectors. To verify that the cointegrating relationship between the three variables detected in the five countries indeed includes all three variables, I run the Johansen cointegration test between logarithms of real debt and real GDP, and between logarithm of real credit and top 1% income share (these tests are not reported here). The former indicates cointegration only for Italy; the latter finds no cointegration for any country.

This selection of countries is somewhat surprising. The theoretical case for a link between income inequality and debt is much stronger for the case of increasing, rather than decreasing, income inequality. However, cointegration is detected in countries where income inequality falls (Germany, the Netherlands and Switzerland) or is flat (Denmark). In only one of these countries income inequality moderately increases (Italy). On the other hand, the U.S., the U.K, and Australia, countries where debt and income inequality grew together for more than two decades, are left out (Australia comes relatively close; the Johansen test indicates cointegration at the 10% significance level). Furthermore, with the exception of Italy, debt and inequality in the other four countries move in opposite directions for most of the observed period.

These results point to a functional misspecification. The theoretical case for the relationship between income inequality and debt is strong only in the case when income inequality is increasing. The case for deleveraging due to lower inequality is much weaker. But this asymmetry is not accounted for by the econometric model. Furthermore, the measurement error in the debt variable is likely to play a significant role.

3.6 Conclusion

This study uses vector cointegration techniques to look for evidence of long-run relationship between increasing income inequality and credit growth. It finds no such evidence. However, this result is far from being conclusive. Limitations of the methodology and available data leave a number of options for further exploration.

Methodologically, perhaps the most serious limitation is asymmetry of the proposed relationship between income inequality and credit. The theoretical model predicts that household credit rises when income inequality increases, but it gives no prediction for a falling income inequality. This asymmetry in the relationship poses an econometric challenge, particularly in the case of cointegration models. Finding a more suitable econometric model is one direction of future research.

Another limitation is related to data. The theory establishes relationship between income inequality and household debt, but the debt variable used in this paper does not include credit card debt, which has been a growing component of overall household debt. Furthermore, the variable includes loans to firms. If the prediction of the theory holds, these loans, used mostly for investment, are likely to be negatively correlated with consumer credit. As Kumhof and Ranciére (2010) shows, inequality can make lending to consumers more attractive than lending to firms, so that total credit can remain the same even if consumer credit increases. Adding investment to the model only partially addressed the problem. For further improvement, higher quality of data is necessary.

Appendix

3.A Replication of Bordo and Meissner's Results

Bordo and Meissner (2011) provide results based on the annual and five-year period data. While trying to reproduce their results, I encountered several inconsistencies in the five-year period case. First, variables for the investment-to-GDP ratio and the nominal interest rate are misspecified. Rather than taking a five-period seasonal difference of the level-variables, the authors use a five-period seasonal difference of the differenced variables. Second, because of this misspecification, the largest common sample available for all specifications appears to have more observations than it actually does. Third, the selection of five-year periods is arbitrary. While first two issues are easily resolved mistakes, the third one represents a more serious problem. I test four regressions for robustness to choice of time periods.

There are five possible choices of five-year time periods, depending on the choice of the first year: $\{1920, 1925, \dots\}$, $\{1921, 1926, \dots\}$, $\{1922, 1927, \dots\}$, etc. Regressions in Table 3.2 use the one where five-year periods are delimited by years $\{1920, 1925, 1930, \dots, 2005\}$. While this choice is arbitrary, it might affect estimation results. I estimate the specification for all five choices and compare the results. The test shows that results in some cases indeed depend on choice of intervals. Even though there are no big changes related to the variable of interest, top 1% income share, the extent to which results depend on the choice of intervals is surprising and it may be useful to report. Tables 3.A.1–3.A.4 correspond to specifications (1)–(4) in Table 3.2 for the five choices of time intervals. Columns (1) through (5) correspond to the samples that start with years 1920, 1921, 1922, 1923, and 1924, respectively.

Table 3.A.1: Interval robustness test for specification (1) in Table 3.2

VARIABLES	(1)	(2)	(3)	(4)	(5)
$\Delta\ln(\text{Debt})_{t-1}$	0.156	0.183	0.296**	0.218**	0.183
	(0.129)	(0.137)	(0.131)	(0.109)	(0.121)
$\Delta\text{Top } 1\%_t$	0.0486*	0.0319	0.0137	0.0329	0.0484*
	(0.0259)	(0.0265)	(0.0261)	(0.0277)	(0.0285)
Observations	105	101	103	106	102
R-squared	0.465	0.544	0.613	0.589	0.543
Number of id	14	14	14	14	14

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 3.A.2: Interval robustness test for specification (2) in Table 3.2

VARIABLES	(1)	(2)	(3)	(4)	(5)
$\Delta\ln(\text{Debt})_{t-1}$	0.0968	0.158	0.242*	0.170*	0.116
	(0.102)	(0.124)	(0.124)	(0.102)	(0.105)
$\Delta\text{Top } 1\%_t$	0.00891	0.00115	-0.0157	0.00832	0.0156
	(0.0267)	(0.0273)	(0.0265)	(0.0302)	(0.0295)
$\Delta\ln(\text{Real GDP})_t$	1.590***	1.334***	1.229***	1.010***	1.258***
	(0.339)	(0.352)	(0.343)	(0.345)	(0.371)
Observations	105	101	103	106	102
R-squared	0.569	0.607	0.670	0.639	0.613
Number of id	14	14	14	14	14

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 3.A.3: Interval robustness test for specification (3) in Table 3.2

	(1)	(2)	(3)	(4)	(5)
VARIABLES					
$\Delta\ln(\text{Debt})_{t-1}$	0.164	0.187	0.241*	0.171*	0.156
	(0.0989)	(0.126)	(0.125)	(0.102)	(0.0943)
$\Delta\text{Top }1\%_t$	0.0523	0.0167	-0.0156	0.00926	0.0337
	(0.0348)	(0.0279)	(0.0278)	(0.0315)	(0.0305)
$\Delta\ln(\text{Real GDP})_t$	1.605***	1.486***	1.230***	1.010***	1.125***
	(0.311)	(0.356)	(0.352)	(0.346)	(0.362)
$\Delta\text{Top }1\%_t \times \ln(\text{Real GDP}_t)$	-0.472*	-0.257	-0.00139	-0.0152	-0.255
	(0.240)	(0.187)	(0.114)	(0.0942)	(0.170)
Observations	105	101	103	106	102
R-squared	0.609	0.618	0.670	0.639	0.629
Number of id	14	14	14	14	14

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 3.A.4: Interval robustness test for specification (4) in Table 3.2

VARIABLES	(1)	(2)	(3)	(4)	(5)
$\Delta \ln(\text{Debt})_{t-1}$	0.138	0.136	0.228*	0.162	0.161*
	(0.105)	(0.133)	(0.126)	(0.0987)	(0.0941)
ΔTop 1%$_t$	0.0404	0.0125	-0.0114	0.00319	0.0254
	(0.0302)	(0.0256)	(0.0266)	(0.0285)	(0.0305)
$\Delta \ln(\text{Real GDP})_t$	1.188***	1.360***	1.036**	0.868**	0.948**
	(0.405)	(0.477)	(0.492)	(0.381)	(0.392)
ΔTop 1%$_t \times \ln(\text{Real GDP}_t)$	-0.354	-0.303	0.00753	0.00155	-0.231
	(0.216)	(0.191)	(0.105)	(0.0883)	(0.172)
ΔInterest rate$_t$	0.316	-0.365	1.103	1.042	0.617
	(0.862)	(0.896)	(0.960)	(0.925)	(0.893)
ΔI/GDP$_t$	0.140	0.179	-0.0165	-0.0894	-0.301
	(0.582)	(0.684)	(0.609)	(0.544)	(0.673)
$\Delta \ln(\text{Real Money})_t$	0.512***	0.360	0.255	0.345	0.396*
	(0.173)	(0.278)	(0.267)	(0.222)	(0.202)
Observations	105	101	103	106	102
R-squared	0.664	0.645	0.685	0.661	0.656
Number of id	14	14	14	14	14

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Bibliography

Abowd, John M., Francis Kramarz, and David N. Margolis. 1999. "High Wage Workers and High Wage Firms." *Econometrica* 67, no. 2 (March 1): 251–333.

Alvaredo, Facundo, Anthony B. Atkinson, Thomas Piketty, and Emmanuel Saez. 2012. "The World Top Incomes Database." April 26. `http://g-mond.parisschoolofeconomics.eu/topincomes`.

Autor, D. H., L. F. Katz, and M. S. Kearney. 2008. "Trends in US wage inequality: Revising the revisionists." *The Review of Economics and Statistics* 90 (2): 300–323.

Axtell, Robert L. 2001. "Zipf Distribution of U.S. Firm Sizes." *Science* 293, no. 5536 (September 7): 1818–1820.

Baker, George, Michael Gibbs, and Bengt Holmstrom. 1994a. "The Internal Economics of the Firm: Evidence from Personnel Data." *The Quarterly Journal of Economics* 109, no. 4 (November 1): 881–919.

———. 1994b. "The Wage Policy of a Firm." *The Quarterly Journal of Economics* 109, no. 4 (November 1): 921–955.

Banerjee, Abhijit V., and Benjamin Moll. 2010. "Why Does Misallocation Persist?" *American Economic Journal: Macroeconomics* 2, no. 1 (January 1): 189–206.

Bernanke, B., M. Gertler, and S. Gilchrist. 1999. "The Financial Accelerator in a Quantitative Business Cycle Framework." *Handbook of Macroeconomics* 1:1341–1393.

Birley, Sue, and David Norburn. 1987. "Owners and managers: The venture 100 vs the fortune 500." *Journal of Business Venturing* 2 (4): 351–363.

Blinder, Alan S. 1973. "Wage Discrimination: Reduced Form and Structural Estimates." *The Journal of Human Resources* 8, no. 4 (October 1): 436–455.

Bordo, Michael, and Christopher Meissner. 2011. "Does Inequality Lead to a Financial Crisis?" *NBER* (December).

Borjas, George J., Richard B. Freeman, Lawrence F. Katz, John DiNardo, and John M. Abowd. 1997. "How much do immigration and trade affect labor market outcomes?" *Brookings papers on economic activity* 1997 (1): 1–90.

Brown, Charles, James Hamilton, and James L. Medoff. 1990. *Employers large and small.* Harvard University Press.

Campbell, J. Y, and N. G Mankiw. 1989. "Consumption, income, and interest rates: Reinterpreting the time series evidence." *NBER macroeconomics annual* 4:185–216.

Clauset, Aaron, Cosma Rohilla Shalizi, and Mark E. J. Newman. 2009. "Power-Law Distributions in Empirical Data." *SIAM Review* 51, no. 4 (November 4): 661–703.

Coakley, J., and F. Kulasi. 1997. "Cointegration of long span saving and investment." *Economics Letters* 54 (1): 1–6.

DeLong, J. B., and L. H. Summers. 1986. "The changing cyclical variability of economic activity in the United States." In *The American Business Cycle: Continuity and Change,* 679–734. University of Chicago Press.

DiNardo, John, Nicole M. Fortin, and Thomas Lemieux. 1996. "Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach." *Econometrica* 64, no. 5 (September 1): 1001–1044.

Dinardo, John, and Thomas Lemieux. 1997. "Diverging Male Wage Inequality in the United States and Canada, 1981-1988: Do Institutions Explain the Difference?" *Industrial and Labor Relations Review* 50, no. 4 (July 1): 629–651.

Duesenberry, J. S. 1949. *Income, Saving and the Theory of Consumption Behavior.* Cambridge, Mass.: Harvard University Press.

Dynan, Karen E., Jonathan Skinner, and Stephen P. Zeldes. 2004. "Do the Rich Save More?" *Journal of Political Economy* 112, no. 2 (April): 397–444.

Fortin, Nicole, Thomas Lemieux, and Sergio Firpo. 2011. "Decomposition Methods in Economics." In *Handbook of Labor Economics,* edited by Orley Ashenfelter and David Card, vol. 4, Part A, 1–102. Elsevier.

Gabaix, Xavier, and Augustin Landier. 2008. "Why has CEO pay increased so much?" *The Quarterly Journal of Economics* 123 (1): 49–100.

Gourinchas, P. O., R. Valdes, and O. Landerretche. 2001. *Lending booms: Latin America and the world.* National Bureau of Economic Research.

Hall, R. E., and F. S. Mishkin. 1982. *The sensitivity of consumption to transitory income: estimates from panel data on households.* National Bureau of Economic Research Cambridge, Mass., USA.

Heathcote, J., F. Perri, and G. L. Violante. 2010. "Unequal we stand: An empirical analysis of economic inequality in the United States, 1967–2006." *Review of Economic Dynamics* 13 (1): 15–51.

Hill, Bruce M. 1975. "A Simple General Approach to Inference About the Tail of a Distribution." *The Annals of Statistics* 3, no. 5 (September 1): 1163–1174.

Ho, T. 2002. "The Feldstein–Horioka puzzle revisited." *Journal of International Money and Finance* 21 (4): 555–564.

Iacoviello, M. 2008. "Household debt and income inequality, 1963–2003." *Journal of Money, Credit and Banking* 40 (5): 929–965.

Idson, Todd L., and Walter Y. Oi. 1999. "Workers Are More Productive in Large Firms." *The American Economic Review* 89, no. 2 (May 1): 104–108.

Johansen, S. 1988. "Statistical analysis of cointegration vectors." *Journal of economic dynamics and control* 12 (2): 231–254.

Juhn, Chinhui, Kevin M. Murphy, and Brooks Pierce. 1993. "Wage Inequality and the Rise in Returns to Skill." *Journal of Political Economy* 101, no. 3 (June 1): 410–442.

Katz, Lawrence F., and David H. Autor. 1999. "Changes in the wage structure and earnings inequality." In *Handbook of Labor Economics,* edited by Orley C. Ashenfelter and David Card, vol. 3, Part A, 1463–1555. Elsevier.

Ke, Bin, Kathy Petroni, and Assem Safieddine. 1999. "Ownership concentration and sensitivity of executive pay to accounting performance measures: Evidence from publicly and privately-held insurance companies." *Journal of Accounting and Economics* 28 (2): 185–209.

Kennickell, Arthur B., R. Louise Woodburn, and Douglas A. McManus. 1996. *Weighting design for the 1992 Survey of Consumer Finances.* SCF Working Papers. Federal Reserve Board.

Keynes, J. M. 1936. *The general theory of employment, interest and money.* Harcourt, Brace & World, New York.

King, Miriam, Steven Ruggles, J. Trent Alexander, Sarah Flood, Katie Genadek, Matthew B. Schroeder, Brandon Trampe, and Rebecca Vick. 2010. "Integrated Public Use Microdata Series, Current Population Survey: Version 3.0. [Machine-readable database]." `https://cps.ipums.org/cps/`.

Koenker, Roger, and Gilbert Bassett Jr. 1978. "Regression Quantiles." *Econometrica* 46, no. 1 (January 1): 33–50.

Kremers, J. J. M., N. R. Ericsson, and J. J. Dolado. 1992. "The power of cointegration tests." *Oxford bulletin of economics and statistics* 54 (3): 325–348.

Krusell, P., and Jr Smith. 1998. "Income and wealth heterogeneity in the macroeconomy." *Journal of Political Economy* 106 (5): 867–896.

Krusell, Per, Lee E. Ohanian, José-Víctor Ríos-Rull, and Giovanni L. Violante. 2000. "Capital-skill Complementarity and Inequality: A Macroeconomic Analysis." *Econometrica* 68 (5): 1029–1053.

Kumhof, Michael, and Romain Ranciére. 2010. "Inequality, Leverage and Crises." *IMF* (November).

Lambert, Peter J., and J. Richard Aronson. 1993. "Inequality Decomposition Analysis and the Gini Coefficient Revisited." *The Economic Journal* 103, no. 420 (September 1): 1221–1227.

Lemieux, Thomas. 2006. "Increasing Residual Wage Inequality: Composition Effects, Noisy Data, or Rising Demand for Skill?" *The American Economic Review* 96, no. 3 (June 1): 461–498.

Lucas, Robert E. 1978. "On the Size Distribution of Business Firms." *The Bell Journal of Economics* 9, no. 2 (October 1): 508–523.

Machado, José A. F., and José Mata. 2005. "Counterfactual decomposition of changes in wage distributions using quantile regression." *Journal of Applied Econometrics* 20 (4): 445–465.

Maddala, G. S., and S. Wu. 1999. "A comparative study of unit root tests with panel data and a new simple test." *Oxford Bulletin of Economics and statistics* 61 (S1): 631–652.

Maliar, Lilia, Fernando Valli, and Seguei Maliar. 2009. "Matlab code for solving the incomplete markets model with aggregate uncertainty using the Krusell-Smith algorithm." `http://ideas.repec.org/c/dge/qmrbcd/180.html`.

Melly, Blaise. 2005. "Decomposition of differences in distribution using quantile regression." *Labour Economics* 12, no. 4 (August): 577–590.

Mendoza, Enrique G., and Marco E. Terrones. 2008. *An Anatomy Of Credit Booms: Evidence From Macro Aggregates And Micro Data.* NBER Working Paper 14049. National Bureau of Economic Research, Inc.

Moore, Henry Ludwell. 1911. *Laws of wages.* Macmillan Company.

Oaxaca, Ronald. 1973. "Male-Female Wage Differentials in Urban Labor Markets." *International Economic Review* 14, no. 3 (October 1): 693–709.

Oi, Walter Y., and Todd L. Idson. 1999. "Firm size and wages." In *Handbook of Labor Economics,* edited by Orley C. Ashenfelter and David Card, vol. 3, Part B, 2165–2214. Elsevier.

Pedroni, P. 1999. "Critical values for cointegration tests in heterogeneous panels with multiple regressors." *Oxford Bulletin of Economics and statistics* 61 (S1): 653–670.

———. 2004. "Panel cointegration: asymptotic and finite sample properties of pooled time series tests with an application to the PPP hypothesis." *Econometric theory:* 597–625.

Pelgrin, F., and S. Schich. 2008. "International capital mobility: What do national saving–investment dynamics tell us?" *Journal of International Money and Finance* 27 (3): 331–344.

Persyn, D., and J. Westerlund. 2008. "Error-correction-based cointegration tests for panel data." *Stata Journal* 8 (2): 232.

Pesaran, M. H. 2007. "A simple panel unit root test in the presence of cross-section dependence." *Journal of Applied Econometrics* 22 (2): 265–312.

Pesaran, M. H., Y. Shin, and R. P. Smith. 1999. "Pooled mean group estimation of dynamic heterogeneous panels." *Journal of the American Statistical Association* 94 (446): 621–634.

Pesaran, M. Hashem. 2004. "General Diagnostic Tests for Cross Section Dependence in Panels." *SSRN eLibrary* (August).

Piketty, T., and E. Saez. 2003. "Income Inequality in The United States, 1913-1998." *Quarterly Journal of Economics* 118 (1): 1–39.

Rajan, Raghuram. 2010. *Fault lines : how hidden fractures still threaten the world economy.* Princeton: Princeton University Press.

Rosen, Sherwin. 1982. "Authority, Control, and the Distribution of Earnings." *The Bell Journal of Economics* 13, no. 2 (October 1): 311–323.

Stiglitz, J. E., and A. Weiss. 1981. "Credit rationing in markets with imperfect information." *The American economic review:* 393–410.

Westerlund, Joakim. 2007. "Testing for Error Correction in Panel Data." *Oxford Bulletin of Economics and Statistics* 69 (6): 709–748.

Williamson, Oliver E. 1967. "Hierarchical Control and Optimum Firm Size." *Journal of Political Economy* 75, no. 2 (April 1): 123–138.

Wolff, Edward N. 2010. *Recent trends in household wealth in the United States: Rising debt and the middle-class squeeze - An update to 2007* 589. Working paper, Levy Economics Institute.

Zeldes, S. P. 1989. "Consumption and liquidity constraints: An empirical investigation." *The Journal of Political Economy:* 305–346.