How Could Wage Inequality Within and Across Enterprises be Reduced?

Moser, Christian

Columbia University

2016
How Could Wage Inequality Within and Across Enterprises be Reduced?*

Christian Moser†

December 15, 2016

1 Between- and within-employer inequality

There has been a recent resurgence in interest among academic and policy circles in income inequality trends around the world. Studies of the differences in income between and within population subgroups such as education groups, races, or genders continues to be a hot topic in current debates. Yet the study of between- and within-group inequality is not a new phenomenon. Fortin et al. (2011) give a comprehensive overview of decomposition methods in economics, with their roots reaching back to at least Oaxaca (1973) and Blinder (1973) seminal work on wage differentials, and later Bourguignon (1979) contribution on inequality decompositions.

A more recent literature has emphasized the importance of an individual’s workplace in determining wage income. The interest of the workplace as a determinant of income inequality stems from the fact that classical wage regressions fail to explain a substantial part of the observed overall variation in earnings (Mincer, 1974; Heckman et al., 2003). Challenging the competitive wage setting paradigm, Krueger and Summers (1988) show that there are large pay differences between seemingly identical workers employed in different sectors of the U.S. economy, with other studies arriving at similar conclusions across many countries\(^1\). Yet substantial wage differentials between workers within industries remained unexplained.

\(^*\)This piece was used as background material for the International Labour Organization (ILO) Global Wage Report 2016/17 available at this link. The ILO is a specialized agency of the United Nations.

\(^†\)Graduate School of Business, Columbia University, 3022 Broadway, New York, NY 10027 (e-mail: c.moser@columbia.edu). Email: c.moser@columbia.edu

\(^1\)Other studies for the U.S. include Katz and Revenga (1989), Bound and Johnson (1992), and Murphy and Welch (1992).
To address this gap, a recent strand of work has examined the patterns of income inequality between and within employers, i.e. either firms or establishments. This analysis is based on a decomposition of the variance of overall (log) earnings into two components: first, the between-firm dispersion in workers’ average earnings; and second, the within-firm dispersion of workers’ earnings conditional on between-firm level differences. Formally, we can write the (log) earnings of worker \( i \) at employer \( j \) at time \( t \), denoted \( y_{ijt} \), using the identity:

\[
y_{ijt} = \bar{y}_t + (\bar{y}_j - \bar{y}_t) + (y_{ijt} - \bar{y}_j)
\]

where \( \bar{y}_t \) stands for economy-wide average earnings and \( \bar{y}_j \) denotes average earnings within employer \( j \), both at time \( t \). Taking variances on both sides and simplifying we get that the overall variance of (log) earnings of worker at employer \( j \) at time \( t \) can be written as:

\[
Var(y_{ijt}) = Var(\bar{y}_t) + Var(y_{ijt}|i \in j)
\]

where the first term on the right hand side denotes the variance of average earnings at the employer level, and the second term denotes the variance of earnings within employers conditional on between-employer differences in average pay.

In recent work, Alvarez et al. (2016) apply this decomposition into between- versus within-firm earnings inequality to Brazil’s formal labor market and find that almost two thirds of the overall earnings dispersion in Brazil’s formal sector in 1996 comes from between-firm differences in average earnings, i.e. the first component in the above decomposition. Conversely, one third of the overall dispersion comes from within-firm differences in pay, i.e. the second component above. Moreover, the authors show that between 1996 and 2012, most of Brazil’s decline in earnings inequality is explained by falling between-firm pay heterogeneity, while a fall in the within-firm pay distribution contributed less.

For the pre-period from 1986-1995 in Brazil, Helpman et al. (2016) show that most of the rise in wage dispersion during those years is also explained by rising between-firm pay heterogeneity. Akerman et al. (2013) confirm that also in Sweden from 2001-2007 a large share of overall wage inequality is observed within sectors and occupations for seemingly identical workers, but less of it is explained by between-firm pay differences. In recent work, Barth et al. (2016) show that in the United States a large share of the increase in earnings inequality between the 1970s and 2010s
is explained by rising dispersion of earnings among establishments. Song et al. (2016) show that in the U.S., contrary to the case of Brazil, the within-firm variance is larger than the between-firm variance for the period from 1978 to 2013. But looking at changes over time, they find, that two-thirds of the increase in wage inequality can be explained by growing wage inequality between firms, and one-third by the increase in wage inequality within firms.

Together, these findings suggest that the workplace may be a quantitatively important component in wage determination and that changes in the distribution of firm-level pay can explain a substantial share of overall earnings inequality movements over the last few decades—the rise and subsequent fall in earnings inequality in Brazil between 1986 and 2012, as well as the secular rise in earnings inequality in and the United States between the 1970s and 2010s, and more recently in Sweden.

2 Two-way fixed effects decomposition

The central challenge in interpreting pay differences between firms is to distinguish between true firm-specific pay premia versus underlying worker heterogeneity. The observation that workers at some firms are paid more highly than other workers at other firms could derive from two fundamental sources of heterogeneity: One the one hand, the identical worker could face large pay gaps between firms, referred to as between-firm pay differences or firm pay heterogeneity. On the other hand, workers may differ in earnings even conditional on their employer identity, which refer to as between-worker pay differences or worker pay heterogeneity.

Both types of heterogeneity—between firms and between workers—will generally show up in the observed earnings distribution. Indeed, the two are indistinguishable in cross-sectional data. To tell apart underlying firm pay differences form worker pay differences, Abowd, Kramarz and Margolis (1999a, henceforth AKM) suggested following workers across different employers in longitudinal data. Under standard econometric assumptions, the pay differences within an individual across different employers is informative of the relative pay component across firms. Furthermore, the average earnings net of the firm pay differences for a given worker of fixed characteristics is then informative of the underlying worker pay heterogeneity. They show that a model with worker- and firm-fixed effects in addition to worker and firm characteristics can be

---

2See Abowd and Kramarz (1999) and Card et al. (2013) for details.
estimated by the method of Ordinary Least Squares using linked employer-employee data.

Starting with AKM’s seminal contribution, a growing literature has employed the two-way fixed effects framework to investigate the sources of wage inequality. While AKM’s original study was on French labor markets, similar decompositions of overall wage inequality into firm- and worker-components have been applied to labor markets around the world. Later work includes Andrews et al. (2008) and Card et al. (2013) for Germany, Iranzo et al. (2008) for Italy, Card et al. (2016) for Portugal, Lopes de Melo (2016) and Alvarez et al. (2016) for Brazil, Bonhomme et al. (2016) for Sweden, as well as Abowd et al. (1999b), Woodcock (2015), Sorkin (2016) and Song et al. (2016) for the United States.

Although these studies differ in important methodological aspects, their results are approximately in line with a decomposition of overall cross-sectional earnings inequality into 50% due to worker pay heterogeneity, 20% due to firm pay heterogeneity, and an overall explanatory power (R²) in the range of 85%. Furthermore, Card et al. (2013), Alvarez et al. (2016), and Song et al. (2016) attribute a substantial share of the shifts in earnings inequality over time to changes in the distribution of firm pay heterogeneity in the AKM framework. Therefore, these findings broadly underline the levels and time trends in between- and within-firm inequality highlighted by the previously mentioned literature.

An important follow-up investigation to the AKM decomposition is: what underlying factors give rise to firm-pay differences on the one hand and worker-pay differences on the other hand? To address this question, Alvarez et al. (2016) find that a close to 60% of the pay heterogeneity across employers is explained by differences in labor productivity, measured by value added per worker at the firm-level. Barth et al. (2016) also find that revenues per worker at the establishment-level is a significant predictor of employer pay differences in the United States, although they find weaker explanatory power. On the worker side, Alvarez et al. (2016) find a moderate correlation between proxies for workers’ human capital and worker pay. Yet a weakening link between productivity and pay accounts for the largest share of the decline in dispersion of both worker pay and firm pay over time.

---

3 A potential explanation for why Barth et al. (2016) find a weaker correlation between establishment-level revenue per worker and employee pay could be that their is not a direct measure of productivity. Many models would predict a direct relationship between labor productivity and worker pay, but not necessarily between revenues per employee and worker pay.
3 Some implications for the human capital model of wages

The new literature surrounding firm-level pay differences challenges the classical view of competitive labor markets, in which workers of different skill groups are each paid their marginal product\textsuperscript{4}. This classical view is based on the notion that workers can move freely between employers, leading the ladder to bid wage rates up to the competitive level at which the marginal cost of a worker, i.e. their wage, is equated to their marginal product in production. Through this lens, differences in pay across employers—while controlling for time-varying human capital proxies such as education and labor market experience—would be interpreted as differences in marginal products for the same worker at different firms. The looming question is then: why do workers not relocate to the most productive, hence highest-paying firms in the economy?

Two explanations for this conundrum have been put forth: The first explanation reconciles observed firm pay differences with the presence of unobserved but heterogeneous non-pecuniary (or non-wage) benefits across employers. These amenities may include qualities of the workplace such as geographic location, office culture and environment, and other unmeasured employee benefits such as employer health care and pension benefit contributions. In this framework, wages act as compensating differentials between employers of different non-wage qualities\textsuperscript{5}. Sorkin (2016) applies such a framework to linked employer-employee data from a subset of the United States labor force and concludes that compensating differentials are at least as important as wage differentials in explaining worker flows between employers. More broadly, an open question remains to what extent non-wage amenities can explain wage dispersion within and across firms in other contexts.

The second explanation deviates from the competitive view of labor markets and instead focuses on determinants of rent sharing between firms and their employees in frictional labor markets. In this view, worker relocation is costly, in terms of resources or time, and potentially has a stochastic (unexplained) component. For example, a currently unemployed individual may search for jobs but could not know the entire list of job postings potentially available to them. Such an individual may at first become employed at a relatively unproductive, low-paying, local firm that had a salient job vacancy posted. But throughout their career, the individual may find better and better job opportunities, being poached by firms with higher productivity or a better personal job

\textsuperscript{4}See also Krueger and Summers (1988) for a similar argument at the industry-level.

\textsuperscript{5}See Rosen (1974), (1987) and Lazear and Shaw (2007) for explorations of such theories.
match. In such a world, pay differences between employers may persist and show up in estimated firm pay differences in the AKM framework.

A myriad of frameworks, commonly summarized under the heading of search models, incorporate such labor market imperfections into an economic environment of wage setting with worker and firm heterogeneity\(^6\). What these frameworks have in common is that a firm-worker relationship generates rents that in the presence of imperfect labor market competition are split, usually non-trivially, between employers and employees. This deviation from the competitive framework has starkly different implications for labor market efficiency and total factor productive relative to the classical framework of labor demand being equated to labor supply for all workers in the economy.

Across firms, the key distinguishing factor between these theories is the presence or absence of abnormal firm rents. In classical theories, firm pay heterogeneity arises only as a result of differences in labor productivities across firms. But such productivities are hard to reconcile with widely documented regularities in labor market dynamics, including the systematic movement of workers towards higher-paying firms over their life cycle\(^7\). In theory, both theories are empirically testable if individual worker productivity and pay is observed; yet requirements on data availability of this kind make this hard to test directly and more work is needed to distinguish between the two.

Within firms, what distinguishes the classical from the frictional view of labor markets is again the presence of rents, but in this context across worker types. Differences in pay across workers within a firm may derive to some extent from observable and unobservable worker productivity differences. Yet a second component of worker pay, which may be heterogeneous across different worker types such as occupations or skill groups, may be related to the extent to which workers capture rents in a given employment relationship. For example, Cahuc et al. (2006) find differences in bargaining power across different worker skill groups, with intermediate- and low-skilled workers having substantially lower bargaining power than high-skilled worker in the context of France. Similarly, workers of different occupations or at different levels of the firm hierarchy may be able to capture different amounts of rents, and this heterogeneity may be influenced by institu-


\(^7\)See Bontemps et al. (1999), Postel-Vinay and Robin (2002), and Engbom and Moser (2016) for evidence in support of this job ladder view of labor markets in different contexts.
tional details such as collective bargaining agreements and compensation policies. Stories of pure rent capturing have gained much attention in the recent debate around the Top 1% and bonus pay policies, particularly in the highest paid sectors and among chief executive officers (CEOs). Yet there exists little concrete evidence on the nature of wage setting for top earners, including CEOs, and hence statements about rent sharing and rent appropriation must be considered with caution.

4 Policy implications

4.1 Minimum wage

A minimum wage acts directly as an institutional wage floor on the pay for workers in the formal sector, as well as indirectly as an outside option for workers in or out of employment. An online database published by the International Labour Organization shows that national nominal minimum wages have followed diverging trends across countries. While many Latin American countries have increased the minimum wage in real value and some European countries like Germany have recently introduced a minimum wage for the first time, the real minimum wage has been declining in many other contexts. How do these trends affect the income distribution in general, and the decomposition into between- and within-firm pay inequality in particular?

A binding minimum wage generically affects both within- and between-firm earnings inequality components. As the minimum rises from non-binding to binding levels, it will by definition push up first and foremost the lowest wages in the economy. This will—all else equal—compress wages within the lowest-paying firms, hence reduce within-firm inequality. But at the same time, by raising the lowest-paying firms’ average pay, it will also lead to convergence in between-firm pay. Quantitatively, the extent to which each of these components is affected by the minimum wage will depend on the initial distribution of pay within and between firms. In the Brazilian case, where between-firm inequality dominated in 1996 and explains the larger share of the inequality decline until 2012, one study found that, accounting for worker and firm pay heterogeneity in an AKM framework, the minimum wage resulted in an approximately equal decline in both com-

8A recent example of direct evidence on the incidence of firm rents is the work of Furman and Orszag (2015) who show find increased dispersion in profit rates among publicly listed firms in the United States.

9This is immediately obvious from the decomposition of the overall variance of (log) earnings into between-firm and with-firm components, see above.
ments\textsuperscript{10}. Overall the study found that the rise in the real national minimum wage by 119 percent explains 70 percent of the decline in wage inequality in Brazil over the period.

4.2 Collective bargaining

The minimum wage can be viewed as a special case of collective bargaining, in which the wage floor is set uniformly across constituencies\textsuperscript{11}. To see this, note that for every minimum wage—be it national, regional, industry, or occupational—one can replicate its effects by an appropriately designed system of union bargaining institutions at the respective level of implementation. Hence, following the same logic as above, collective bargaining may affect the income distribution both between and within firms.

But different types of collective bargaining may have vastly different effects on the earnings distribution. Its coverage, which is affected by automatic extension mechanisms and opt-out clauses, will determine which workers and firms are (not) affected by centralized bargaining outcomes. For example, if the bargaining power of worker associations in some of the highest-paid industries is exceptionally strong then the effect of collective bargaining may actually be to exacerbate between-firm pay inequality. For example, Alder et al. (2014) study the Rust Belt industrial region of the United States, in which strong unions used their bargaining power to obtain higher worker pay. A similar context is that of Brazil’s ABC manufacturing region around São Paulo, which historically has been the focus of many sectoral and occupational worker unions’ bargaining efforts. Hence, the effects of decentralized bargaining outcomes on the overall earnings distribution is ambiguous.

On the other hand, coordinated wage bargaining at a centralized level can affect the earnings of the employed much like a national minimum wage would. Such a system naturally has other drawbacks, potentially including adverse effects on economic growth and the business cycle, as has been argued in the case of Sweden and Germany. At the same time, it can also effect lower dispersion in average firm-level wages by leading to higher wages in some enterprises and wage restraints in others; and simultaneously help to close the within-firm pay gap between unionized and non-unionized worker groups as exemplified by the cases of Norway and Sweden (Waller-\textsuperscript{10}See Engbom and Moser (2016).
\textsuperscript{11}If course collective bargaining may involve other employment terms such as hours worked, holidays, equal pay policies, and other worker rights. In the context of pay inequality, we will focus only on the pay component of collective bargaining, although the other components may also indirectly affect the wage setting protocol.
stein, 1999; Barth et al., 2014) and the United States (Card, 1996).

Furthermore, enacting a minimum wage or bargaining agreements tailored to one part of the earnings distribution potentially can have spillover effects in other parts of the distribution as well. For example, Lee (1999) studies the extent to which a decline in the minimum wage from 1979 to 1989 in the United States can explain widening dispersion in the entire earnings distribution. He study finds that spill-over effects of the decline in the real minimum wage have had sizable effects on the lower half of the earnings distribution. A follow-up study by Autor et al. (2016) casts doubt on the robustness of these findings over a longer time period, concluding that spill-over effects of movements in the real minimum wage in the United States between 1979 and 2012 are indistinguishable from measurement error. Engbom and Moser (2016) find that spillover effects of the minimum wage in the Brazilian context have been sizable, accounting for approximately one half of the decline in earnings inequality due to the minimum wage. In light of these findings, a re-examination of spillover effects due to the minimum wage and other wage setting institutions, both in the United States and in other contexts, seems warranted.

4.3 Limits to the minimum wage, collective bargaining, and other wage setting institutions

While much of the above discussion of minimum wages and collective bargaining agreements has focused on the effects on the earnings distribution conditional on employment, other potentially adverse effects of such wage setting institutions should not be neglected. There are at least three reasons why these institutions’ effectiveness may be limited.

First, the debate about the displacement effects of a binding minimum wage floor\(^\text{12}\) has not reached a consensus on the elasticities of employment with respect to the minimum wage, and the employment effects of collective bargaining remain an elusive object in economic research. In this case, potential surges in non-employment or job downgrading for parts of the labor force have to be weighed against wage gains for those who remain in full employment.

Second, any binding factor market price floors, which include minimum wages and collective bargaining agreements, may to some degree be passed down into final goods prices, as Harasztosi and Lindner (2015) find for the case of a minimum wage introduction in Hungary. To the extent that consumer prices increase along with workers’ earnings, this channel will counteract some of

\(^{12}\)See Card and Krueger (1994) and a rich set of follow-up work.
the welfare gains for the affected worker groups, thereby limiting the effectiveness of minimum wage and collective bargaining institutions.

Third and finally, in many lower income countries the informal sector may absorb some of the previously formal jobs. To the extent that labor market policies and wage agreements are less enforced in the informal economy relative to the formal sector, the presence of an informal economy may hence offset any changes in wage institutions that are formally instituted.

4.4 Outsourcing and the boundaries of the firm

One explanation that has been proposed for rising between-firm inequality in the United States and other countries is an increased assortativeness in the assignment of workers across forms. Such increases in assortativeness between workers and firms may arise for example from subcontracting and re-defining the boundaries of the firm. Concretely, if low-skill tasks are increasingly concentrated in specialized enterprise vehicles and “core value” tasks become the focus of the original businesses, one would expect increasing dispersion in firm-level pay consistent with the observations of the between- and within-firm inequality literature. Song et al. (2016) argue precisely for this phenomenon in the context of the United States, where they find that increased worker segregation explains a substantial share of the rise in between-firm pay inequality in the United States. This explanation is also broadly consistent with a related literature contrasting the rising labor productivity with falling labor share in earnings in the United States since the 1980s (Elsby et al., 2013; Karabarbounis and Neiman, 2014).

There are reasons to be critical of this trend towards workers’ segregation across specialized firms by nature of their tasks. For instance, large firms of mostly low-paid workers may derive large bargaining power, particularly in frictional and local labor markets. With the acquisition of large market power, these firms may appropriate most of the rents on the table and also provide lower levels of employer benefits, including health care and pension contributions. Such a shift may plausibly lead to an erosion of general worker bargaining power, which may squeeze large parts of the labor force.

Keeping these critical points in mind, one should not throw the baby out with the bathwater. Boundaries of firms are picked by optimizing stakeholders and reflect available gains from trade

---

13See Song et al. (2016) for a further discussion of reasons for why increasing worker segregation across firms can be problematic from a distributional standpoint.
and organizational restructuring. Through this lens, the trend towards outsourcing may well be viewed as efficient. The crux of the matter is whether or not the re-organization is leading to additional efficiency gains beyond facilitating the redistribution of rents away from workers and towards capital owners. In this context, it seems reasonable to ensure that workers at all stages of the supply chain are given adequate rights and social provisions.

How can this be achieved? Reforms to help guarantee workers’ livelihood under changing organizational structures could involve re-defining legal firm boundaries and making it harder for large corporations to avoid paying benefits simply by re-defining boundaries of the firm. Ensuring equal inclusion of all parts of the supply chain in worker unions and collective bargaining agreements, i.e. avoiding conditionalities and opt-out agreements, could be another remedy to this trend. Alternative measures that work through incentives, rather than restricting firms’ choice sets, may be equally worth exploring.

4.5 Promoting the right sort of enterprises

Given that measures of firm performance are strongly related to firm-level worker pay (AKM; Barth et al., 2016; Alvarez et al., 2016), there need not be a trade-off between growth and inequality. Instead, promoting productivity growth among firms may simultaneously reduce income inequality on the worker side. For instance, policies that lead to convergence in the firm productivity distribution can be expected to also close the pay gap among workers. Examples of such policies are discussed in Moser (2016) and include industrial policies promoting employment and productivity growth of small and medium-sized enterprises, entry of new competitors, and investment in product quality-improving innovation. Yet the proposition that policies leading to productivity gains among the lowest-paying firms also induce wage increases at the bottom of the distribution, thereby closing the income gap, remains a hypothesis with little empirical evidence to cite.

4.6 How to deal with inequality within enterprises, and specifically the Top 1%?

Should we target policies directly at inequality within enterprises, and specifically the Top 1%? Given the earlier discussion of market versus frictional views of wage setting, it is not clear what constitutes “too much” inequality, especially at the top of the income distribution. Efficient market theories and in particular theories of organization of the firm suggest that high paid managers and
CEOs may be paid their marginal product, but that vertical branching out of firm structures leads to large marginal benefits from small differences in ability at the top of the enterprise management.

Two shared goals seem uncontroversial, at least in theory: First, like for other parts of workforce it should be guaranteed that managers are paid at or below their marginal product within the enterprise. Second, compensation contracts should be designed to incentivize among managers to engage in long-term value creation rather than short-term rent extraction. But of course another question altogether is course how such policies can be efficiently implemented in different business contexts.

Putting policies into place to promote these goals may require innovative incentive schemes. For example, public salary disclosures as in Card et al. (2012) and a recently implemented SEC rule on CEO pay ratio disclosure may make it harder for top managers to extract exorbitant rents without uproar among both stakeholders and the public. Such transparency could ultimately feed back into well-informed actions of managers and workers themselves, enhancing long-term economic growth and stability.

References


