Organisational Intersectionality: Do Gender and Migration Status Inequalities Reinforce or Offset each other in French Workplaces?[1]

Olivier Godechot

Sciences Po, CRIS, CNRS and AxPo.

Axa Chairholder in Economic Sociology

https://orcid.org/0000-0003-3199-8144

Mirna Safi

Sciences Po, CRIS, CNRS

https://orcid.org/0000-0002-6730-8320

Matthew Soener

University of Illinois – Urbana/Champaign

https://orcid.org/0000-0002-9261-9622

[1] The three co-authors contributed equally to this article. This research has been supported by the French Research Agency (Grant ANR-17-CE41-0009-01), MaxPo, Sciences Po, AxPo, Sciences Po and by Sciences Po (SAB 2023). We would like to thank Clem Aeppli, Théo Voldoire and the COIN team for useful suggestions on earlier versions of this paper.

Abstract

This study examines whether organisations with significant pay gaps along one dimension (gender, migration status, class, …) tend to exhibit similarly high inequalities along other dimensions, or whether there is a trade-off between inequality dimensions. Using French administrative data, it estimates correlations between class, gender, and migrant workplace earnings gaps, and studies how these gaps also relate to a fourth measure of intra-categorical inequality. To ensure robust results, this article introduces innovative methods to address measurement biases that may distort the relationship between earnings gaps. It establishes three key patterns. First, the gender gap is higher in more unequal workplaces. Second, the migrant gap is higher in more equal workplaces. Third, gender and migrant earnings gaps are negatively correlated within workplaces. These results suggest that workplace inequality regimes are shaped by both reinforcing and trade-off dynamics. Finally, this article explores factors influencing these patterns and highlights the role of industries.

Keywords: workplace inequality, intersectional inequality, organizational inequality, class, gender, migrant status

JEL : J310, J710

Introduction

Workplaces are considered as the locus of inequality in modern societies. The relational inequality approach developed by Tomaskovic-Devey and Avent-Holt (2019) builds on Acker's (2006) influential insight arguing that organisations are unique *inequality regimes*. This concept emphasizes the crucial role of workplaces not only in producing unequal outcomes but also in shaping the interrelations between various types of inequalities. Research has enriched the empirical content of the *inequality regime* concept by documenting how firms substantially vary in the way they produce categorical inequalities such as migrant status/ethnic/racial/gender wage gaps (Williams and O'Reilly III, 1998; Tomaskovic-Devey, Hällsten and Avent-Holt, 2015).

However, researchers too often address these forms of inequality separately. The interrelation of inequalities – the core idea of the inequality regime concept – has not been systematically investigated. One way of conceptualizing this is through intersectional approaches (McCall, 2001; Browne and Misra, 2003; Cho, Crenshaw and McCall, 2013). Quantitative studies in this vein generally implement intersectionality with interaction terms between individual variables such as respondents' gender, race or social class. This tends to overlook the workplace heterogeneity in the making of these different types of inequality (Baron and Bielby, 1980; Tomaskovic-Devey and Avent-Holt, 2019).

This article bridges this research gap by investigating *how different types of pay inequalities intersect within workplaces*. Specifically, it asks whether various dimensions of organisational inequality converge, resulting in workplaces that differ primarily in their overall level of earnings inequality consistently across all dimensions, or whether trade-offs emerge. For instance, do some workplaces exhibit significant gender pay gaps while maintaining relatively equal earnings across migrant status? By exploring the relations between gender, migration status and class dimensions of earning inequality, this article sheds light on whether inequality regimes operate in reinforcing or compensatory ways within workplaces.

This article makes three contributions to social inequality research. The first is theoretical: it combines two central sociological concepts – inequality regimes and intersectional inequalities (McCall, 2001; Tomaskovic-Devey and Avent-Holt, 2019). Building on previous literature, it specifically investigates two main patterns along which *inequality regimes* might vary: a *reinforced* or a *trade-off* inequality structure. This approach helps clarify a core concept in the organisational inequality literature.

The second contribution is methodological. There are several pitfalls to comparing workplace inequalities based on the same earnings variable. Correlating directly different measures of inequality contain artificial associations that bias results. Additionally, in small workplaces, measurement errors also affect the results. This article develops an original method to overcome this problem by netting out individual characteristics to compute unbiased correlations at the workplace level and propose two techniques for circumventing measurement errors.

The final contribution is empirical. Using linked employer-employee administrative population data on French workplaces between 1996 and 2021 allows us to analyse three dimensions of inequality in France: gender, migration status and class. In addition, this article proposes an overall workplace inequality measure called "intra-categorical inequality" (e.g., within the three-level cross-categorisation gender×migration_status×class). There are three major patterns. Gender earnings gap is positively correlated with the intra-categorical inequality measure (r=0.3) while the migrant earnings gap is higher in more equal workplaces (r=-0.1). Gender and migrant earnings gaps are negatively correlated within the workplace (r=-0.2). In contrast, class inequality exerts rather small and sometimes unstable correlation with gender, migrant status and intra-categorical inequality within workplaces. Finally, this study explores the factors driving the distribution of workplaces along these patterns and highlights the key role of the industrial sector, which helps us interpret the mechanisms at stake.

Bridging Multidimensional Forms of Inequality within Organisations

The class/gender/race trio is the basis for thinking about inequality in most contemporary societies. Intersectional scholars have called for studying these forms of inequality together (Crenshaw, 1989). By examining these processes simultaneously, intersectionality is geared towards understanding the constructed (and contested) nature of these categories and emphasizes the unique inequality "standpoints" – experiences based on particular social positionality and identity formation. While these are central insights, the term has been used in different ways (Cho, Crenshaw and McCall, 2013). Following McCall (2001, 2005), some social scientists have attempted to assess and quantify how different social positions shape socioeconomic outcomes. While the use of survey data (which implies some reification of categorical positions) may be in tension with the deconstructivist roots of intersectionality theory, McCall makes a strong case that intersecting positions have important labour market and distributional consequences. Along these lines, Greenman and Xie (2008) use interaction terms for overlapping inequalities and challenge the "double jeopardy" hypothesis for minority women. Instead, they show that the racial/ethnic wage penalty is smaller for women than for men.

This strand of research illuminates how categories combine in shaping individual outcomes and perceptions of inequality (Penner and Saperstein, 2013). Yet, an equally important element in this discussion is *where* inequality is generated. Recent scholarship emphasizes the role of organisations in generating inequality (Avent-Holt and Tomaskovic-Devey, 2012). Workplaces are sites where individual skills, agency and boundary-making processes interfere with organisational dynamics like discrimination, wage-setting, promotions, demotions, profit distribution and managerial power. These factors modify inequality intensity across categorical distinctions like race, migration status, gender and class in a number of different countries (Avent-Holt and Tomaskovic-Devey, 2012; Melzer *et al.*, 2018).

While relational inequality literature explains how and where categorical inequality is produced, it tends to only focus on one type of inequality (e.g., gender wage gap, migrant wage gap). Acker's concept of *inequality regimes* overcomes this limitation by combining insights from intersectional and organisational inequality. This is defined as "loosely interrelated practices, processes, actions and meanings that result in and maintain class, gender and racial inequalities within particular organizations" (2006, p. 443). As an example of inequality *interrelations*, Acker cites a study showing that, in Norway, large bureaucratic engineering firms produce both higher class hierarchies and larger gender wage gaps than small collegial ones (see Kvande and Rasmussen, 1994). This example thus suggests that gender and class inequality reinforce one another. However promising, Acker's seminal article was mostly a reflection on a lifetime of organisational-inequality research and an agenda-setting piece for sociology attempting to conceptually link access to power and resources within an organisation by race, class and gender. The precise interrelational dynamics within organisations were largely left to further research.

This article fills the research gap with a systematic study of the inequality structures along which inequality regimes are distributed. It specifically derives two competing hypotheses from the literature. The first is a *reinforced inequality structure* in which *different workplace inequalities are positively correlated with each other* (H1). The second type is a *trade-off inequality structure* in which different dimensions of inequality are *negatively correlated with each other* (H2).

A first argument supporting the *reinforced inequality* hypothesis is that workplaces exhibit various levels of tolerance of resistance to inequality, shaped by their organisational and/or cultural environments. These underlying patterns tend to affect all dimensions of inequality simultaneously, leading workplaces to cluster consistently around either high or low levels of inequality across multiple dimensions such as gender and racial wage gaps (Castilla, 2008). Organisational and management practices play a central role in these dynamics because of the cultural environments they create (Roscigno, Hodson and Lopez, 2009). For example, U.S. workplaces often foster cultures where high inequality is legitimized and internalized, while the Japanese workplace environment limits inequality (Milgrom and Roberts, 1988). Gender wage gaps also reflect workplace-specific cultural norms and associated wage setting practices (Avent-Holt and Tomaskovic-Devey, 2012). While these studies underscore the role of hierarchical workplace cultures in shaping inequalities, most of them focus on a single dimension. This raises the broader question of how such cultural environments affect multiple forms of inequality.

A second argument in favour of the *reinforced inequality* hypothesis is that categorical exogenous classifications, such as gender or migration status, naturalize and increase global wage inequalities (Hedström, 1991; Treiman, 2013). Both are the salient anchor for processes of exploitation and opportunity hoarding (Tilly, 1998). Relatedly, the occupational hierarchy is reinforced by the gender gap (Charles and Bradley, 2009) and the ethnic/racial gap (Grodsky and Pager, 2001). Tomaskovic-Devey and Avent-Holt summarize this reinforcing dynamic, noting that "when social distinctions of class and citizenship reinforce each other, they exaggerate categorical distinctions, which in turn legitimates exploitation" (2019, p. 121). Much of this literature focuses on the relationship between gender inequality and overall inequality in the workplace. This research tends to show that gender inequality is larger in environments with high competition and unequal rewards (Eagly and Karau, 2002; Roth, 2006). The gender gap was also shown to be pronounced in German firms with higher levels of overall inequality (Tomaskovic-Devey and Avent-Holt, 2019, p. 79-81). The persistence of traditional gender roles in families explain why women are channelled (or self-select) into jobs, functions and hierarchical positions that are less exposed to competition or that involve more risk-taking behaviours and require more assertiveness. In addition, the "glass ceiling" keeps both the gender and overall wage inequality high (Baxter and Wright, 2000).

This reinforcing dynamic can be contrasted with the *inequality trade-off* hypothesis. The latter is less widely adopted in the literature and rarely formulated as such. However, some lines of research provide evidence that such a dynamic is possible, mostly as a result of the opposition between productivity-related wage scales and categorical inequalities. Indeed, the efficient market perspective recalls that organisations have a clear incentive to base the wage scale on productivity when the latter is easy to observe (Holmström, 1979). As suggested by tournament theory, when employees' efforts and contributions are difficult to assess accurately, promotion decisions may become disproportionately rewarding in order to incentivize those employees that have not yet been promoted (Lazear and Rosen, 1981; Connelly *et al.*, 2014). This policy is likely to result in large class-based inequalities (managers versus workers) and small within-class wage inequalities. Similarly, statistical discrimination mechanisms suggest that, in the absence of accurate measures of employee productivity, employers may base hiring, promotion and pay on their prior beliefs about the average productivity of salient categories (Aigner and Cain, 1977). Moreover, greater transparency and accountability also limit the use of gender and racial categories as a basis for wage inequality (Dobbin, Schrage and Kalev, 2015). These perspectives thus suggest that there may be a trade-off between organisational inequalities, particularly between categorical and overall inequalities.

This article systematically examines whether the distribution of inequality regimes in France follows a *reinforced inequality* or a *trade-off inequality* pattern in France. This is based on four types of organisational inequalities: gender, migration status, class and intra-categorical inequality. Migration status is here a proxy for ethno-racial origin. In line with recent literature on the professional-managerial class, class position is based on the distinction between professional and managers on the one hand and other employees on the other hand. Intra-categorical inequality is a global and gradient measure of inequality computed at the workplace level. As explained in the methodological section, this measure cannot be a simple wage variance indicator because it would be trivially correlated with categorical wage gaps. Instead, the intra-categorical inequality measure estimates the level of workplace inequality net of gender, migration status and class attributes. It is related to how workplaces reward unobservable characteristics such as creativity, personal involvement, availability and other types of soft skills.

Inequality in France

This study focuses on France. In France, wage inequalities by gender, migration status and class are well established in empirical research. French labour markets are marked by strong labour market protections but by low union density and increasing labour market flexibility (Amable, 2017). In addition, unlike many other OECD countries, wage inequality is relatively stable and within-workplace inequality has been declining (Tomaskovic-Devey *et al.*, 2020). This is important because France wage inequality is comparatively less extreme, and the results may be more conservative in comparison to other contexts.

The gender wage gap is slowly declining (Meurs and Pora, 2019). Although lower than in many high-income countries, the gender gap remains pronounced. In 2015, it amounts to 12 percent controlling for age and number of hours and drops to 6.5 percent when measuring the wage gaps within the same occupations and workplaces (Penner *et al.*, 2023). Researchers also find forms of labour market disadvantages for women similar to other high-income countries such as occupational segregation, motherhood penalty and involuntary part-time (Meurs and Pora, 2019). In addition, sorting into firms is a key explanation to the French gender gap (Coudin, Maillard and Tô, 2018).

France has a long history of immigration, marked by successive waves since the late 19th century, mostly from European countries. During the 1950s and 1960s, the government encouraged immigration from former colonies due to labour shortages, leading to significant arrivals from North Africa, as well as countries like Portugal. Since the 1970s, stricter policies have curtailed labour migration. Immigration steadily diversified with growing numbers coming from Southeast Asia, Sub-Saharan Africa and more recently from South Asia and China. In 2017, 33 percent of migrant workers came from Europe, 27 percent from North Africa, 18 percent from Sub-Saharan Africa, 16 percent from Asia and seven percent from America (Desjonquères *et al.*, 2021).

In 2021, 70 percent of immigrants aged 15–64 in France participated in the labour market. Immigrants hold about one in ten jobs in France, and their labour force is highly diverse in origin and skill levels (Desjonquères *et al.*, 2021; Insee, 2023). While 40 percent of immigrants lack formal educational credentials compared to 16 percent of non-immigrants, 25 percent of immigrants hold advanced degrees, a proportion similar to the native population. While present in all sectors, immigrants tend to be disproportionately concentrated in construction, catering, accommodation and cleaning sectors.

Unemployment is notably higher among immigrants. While European immigrants generally have better employment outcomes, immigrants from

North and Sub-Saharan Africa experience the highest unemployment rates. Research indicates part of this disparity stems from hiring discrimination (Safi and Simon, 2013; Quillian *et al.*, 2019). Once employed, the native/migrant hourly wage gap is close to that of other EU and OECD countries (Athari, Brinbaum and Lê, 2019), amounting to -4 percent in 2018 for employees in the same occupation and workplace (Hermansen *et al.*, 2023).

Data

We use administrative data from the *Base Tous Salariés* (BTS)[2] – a database administered by the French National Statistical Institute (INSEE) based on the *Déclarations Annuelles de Données Sociales* (DADS), which firms submit to pay social security contributions. The BTS covers all private sector firms as well as local public services and hospitals.[3] Every firm is legally required by law to submit it every year. BTS provides an exhaustive series of detailed earnings and occupation information for each employee in each establishment of each firm in France. As the establishment level is the best proxy for organisational activity, it is used here as the elementary unit.[4] Since the annual number of hours worked is less reliable in the years 1994 and 1995, analysis begins in 1996. Our sample includes all prime-age (25-55) French employees in the private sector, hospitals and local public services of metropolitan France who earn at least half a yearly minimum wage. Thus, the sample ranges from 13.4 million (in 1996) and 16.3 million (in 2021) employees (Table A1).

We construct a consistent sample to compute meaningful categorical wage gaps at the workplace level between 1996 and 2021. Sampled establishments had to have more than fifty employees and with at least two members of each key category used in this study. That is, establishments with at least two female manager/professionals (and workers), two male manager/professionals (and workers), two migrant manager/professionals (and workers), two non-migrant

---

[2] Access to the BTS data was obtained through the CASD (e.g., *Centre d'accès sécurisé aux données* - https://www.casd.eu/) dedicated to researchers authorized by the French *Comité du secret statistique*.

[3] State public servants are included only after 2009. For reasons of time consistency, we did not use them for this analysis.

[4] In this study, the notion of workplace (or, alternatively, establishment used here as a synonym) is based on the French SIRET, which is the legal identifier of an establishment.

manager/professionals (and workers), two female migrants (and non-migrants) and two male migrants (and non-migrants). The online Appendix includes comparisons with the sample before and after making these changes (Table A1). These selection rules, in particular the exclusion of workplaces below fifty employees, divide the number of observations by four and result in a sample with a larger share of managers (28 versus 13 percent) and with 14 percentage points higher earnings. The gender and migration status balance remains the same as in the original sample. While the main goal is to achieve internal validity within the sample definition, the results are globally representative of the earnings structure in large French establishments.

In the following analyses, the logarithm of gross annual wage (e.g., earnings) serves as the first stage dependent variable. Three categorical inequalities can be measured in the data: gender, migration status and class.

Class positions in this article follow classical approaches distinguishing ownership and workplace authority (Wright, 1997) as well as social prestige conferred through education and related forms of closure (Erikson and Goldthorpe, 1992; Weeden and Grusky, 2005). Aggregate occupational groups distinguish between those with workplace authority (managers) and social prestige (professionals) and everyone else (workers). Therefore, while occupational categories are embedded in the class positions, the class positions are not reducible to occupation (Zhou and Wodtke, 2019). The upper-class position, moreover, conforms to a standard French class designation known as the *cadres* – an institutionalized category of both managers and professionals (senior executives, commercial managers, administrators and professional positions such as scientists, lawyers, journalists) (Boltanski, 1982).[5]

Gender is coded with the binary variable provided in BTS based on the first digit of the social security number. It thus corresponds to the gender status assigned at birth. Finally, migration status is proxied by using information on the French

[5] We also add the small category of salaried CEOs with more than 10 employees (PCS *23. Chefs d'entreprise de 10 salariés et plus*) to the French PCS group 3 (*Cadres et professions intellectuelles supérieures).*

département of birth also based on the social security number.[6] While the data lacks information on nationality at birth to properly apply the French definition of a migrant (e.g., foreign born with a foreign nationality at birth), using foreign-born employees (coded *département* of birth= "99") is a reasonable proxy for identifying migrant employees. Indeed, the 2016 French Labour Force Survey shows that 85 percent of foreign-born employees aged 25-55 are migrants. Similarly, information on migrant countries of origin is lacking. However, we do have information after 2005 whether respondents are EU or non-EU citizens. In the Appendix, we use this information to re-estimate the main analyses as a robustness check (Figure A4).

Methods: measuring the intersection of inequalities at the workplace level

Studying the intersection of inequality at the workplace level poses several methodological challenges. These challenges are discussed in the first subsection. In the second subsection, we estimate the contribution of socioeconomic factors to the structuring of workplace inequality.

A two-stage design for estimating workplace categorical and intra-categorical inequalities

A "naive" approach for studying the workplace correlation between different earnings inequality measures would be to estimate separately first-order measures of these inequalities in each workplace[7] and then compute their correlation. This strategy would be inherently biased because it would capture an artefactual relationship between measures derived from the same earnings variable. For example, if gender and migration status are positively correlated, the first-order gender gap will be overestimated because it will also capture part of the first-order migrant gap (Woolridge, 2013). As a consequence, the correlation between the first-order gender and migrant gaps would be

[6] Information on the *département* of birth is incomplete between 2002 and 2004 and missing in 2011. Following Babet, Godechot, Palladino (2023), we construct a pseudo-panel which allows us to complete information.

[7] First order measures of inequalities refer here to simple gross measures of workplace inequalities such as differences in average log earnings for gender, class and migration status.

overestimated. Similarly, workplaces with large gender gaps mechanically have higher log earnings variance, because the gender gap is one component of earnings inequality. In Appendix 3, further details about these biases with a mathematical decomposition and simulations are provided.

To circumvent these artefacts, we run earnings regressions in each workplace in which we simultaneously estimate gender, migration status and class fixed effects. This strategy estimates gender, migration status and class gaps netted out from one another. Moreover, we also compute the variance of the residuals from these earnings regressions. This variance can be interpreted as an indicator of workplace inequality net of categorical inequality. We will refer to it below as the intra-categorical inequality measure.

This strategy is implemented in a two-stage design. The goal of the first-stage model is to generate a within-workplace residual earnings variable w_{ij} furhter used in the second-stage to estimate earnings gaps at the workplace level net of differences in individual characteristics. This first-stage (equation 1) is a classical log earnings OLS regression on the full population.

$$\begin{aligned} log(earnings_{ij}) = {} & X_{ij}.\alpha + a^{gdr}.\, gender_{ij} + a^{mig}.\, mig\_stat_j + a^{gdr.cla}.\, gender_{ij}\times class_{ij} \\ & + a^{mig.cla}.mig\_stat_{ij}\times class_{ij} + a^{gdr.mig}.\, gender_{ij}\times mig\_stat_{ij} \\ & + a^{gdr.mig.cla}.\, gender_{ij}\times mig\_stat_{ij}\times class_{ij} + a_j + w_{ij} \end{aligned} \tag{1}$$

It accounts for the earnings impact of workplace composition related to activity, technology, sector and local labour market composition. This first-stage filter is important because it accounts for trivial within-workplace wage differences due to selection. Thus, netting out these factors allows us in our second-stage to estimate specific ways in which workplaces organize inequalities.

The term X_{ij} captures individual drivers of individual wages, especially experience, work involvement and skills.[8] Individual variables include age (as

[8] We recognize that the composition of workplaces is not determined solely by technology and economic conditions, and that workplaces have leeway to adapt the composition to the type of workplace inequality they wish to create. Thus, the first stage could be too constraining and lead to overlooking key organizational inequalities in the second stage. On the other hand, dropping control variables in the first stage could also lead to capturing the exogenous compositional effect in the second stage, which is precisely what we want to avoid. Therefore, we have opted for a conservative estimation of workplace specific inequalities, by introducing strong controls in the first-stage. In table A7, we compare our results with two alternative specifications. In the first specification, we do not compute a first-stage and go directly to the second stage. In the

well as its squared term), number of hours worked (simple and squared), number of days worked (simple and squared) and part-time status. Unfortunately, due to its administrative nature, the data does not include information on the educational level of employees. However, we control for detailed occupation (2-digit codes), which proxies for differences in education and accounts for occupation-specific human capital.[9] Additionally, interaction terms between occupation and age as well as age squared account for strong occupational heterogeneity in earnings trajectories. We include a three-term interaction between gender, class and migration status at the individual level. This nets out the second-stage correlation of workplace wage gaps from the individual-level intersection of inequalities. Finally, a workplace fixed effect a_j is included to account for the positive (or negative) workplace pay premium.

In the second-stage, the parameter w_{ij} (estimated in equation 1) serves as the dependent variable to perform one regression per establishment j with gender, migration status and class dummies as three independent variables (equation 2).

$$w_{ij} = b_j^0 + b_j^{gdr} \times gender_{ij} + b_j^{mig} \times mig\_stat_{ij} + b_j^{cla} \times class_{ij} + e_{ij} \qquad for\ each\ j \quad (2)$$

Each fixed effect b_j^{gdr}, b_j^{mig}, b_j^{cla} describes the positive/negative deviation of the workplace j's gender, migrant and class gaps to the mean gaps calculated in equation 1 (estimated with a^{gdr}, a^{mig}, a^{cla} and their interactions). Building on equation 2's estimates, we compute the covariances and correlations between the estimates of interest (weighted by the number of workers per establishment):

$$cov(b_j^{gdr}, b_j^{mig}) \qquad (3)$$

Additionally, we are also interested in the correlation between each of these categorical earnings gaps and the intra-categorical workplace inequality. We measure the latter in each workplace with the standard deviation of errors (root

second, we implement a minimal first stage, which controls age, number of hours and days worked and part-time status. The latter yields very similar substantive results than the full first-stage.

[9] Measured on the French Labor Force Survey 2016, the table linking occupations (25 categories) and education (16 categories) yields a Cramer's V of 29 for younger wage-earners (e.g., 25-35) and of 27 for elder ones (45-55).

mean square error $RMSE_j$ – with $RMSE_j=(\Sigma_i e_{ij}^2/n_j)^{1/2}$ where e_{ij} are the residuals estimated in equation 2).

While this approach allows us to estimate more consistent workplace correlations of inequality measures, it still faces a second methodological challenge. When workplaces are large (e.g., more than two hundred workers), fixed effects and residual variance are measured accurately. For smaller workplaces, estimation errors are more likely. Moreover, the measurement errors for different estimates within the same workplace are not independent. This problem can bias the variance, covariance and correlation of the inequality indicators. This bias is similar to the one found in the "AKM" literature (initiated by the Abowd, Kramarz, Margolis (1999) article) when estimating the covariance between workers and firm fixed effects (Andrews *et al.*, 2008).

Two methods can overcome this measurement error bias which is discussed in extensive detail in the Appendix. The first strategy (hereafter called *vcov*) is inspired by Andrews et al. (2008). It takes advantage of the fact that an OLS regression comes with an estimate of the variance-covariance matrix (M_j) of the errors on the parameter estimates. The solution for correcting equation 3 is to subtract from the covariance the mean of the $m_j^{gdr\times mig}$ components of the (M_j) matrices estimated in each workplace regression (cf. Equation 4).

$$cov^{vcov} = cov_j(b_j^{gdr}, b_j^{mig}) - (\Sigma_j m_j^{gdr\times mig})/J \quad (4)$$

In the second solution (hereafter called *halves*), we follow recent literature correcting for biases in the estimation of fixed effects (Babet, Godechot and Palladino, 2023) and adopt a split sample approach. Employees are randomly assigned in each workplace to two subsamples: "half 1" or "half 2." We then estimate separate regressions on the two randomized halves and calculate cross-halves covariance:

$$cov^{halves} = cov_j\,[(b_{j1}^{gdr}, b_{j2}^{gdr}),(b_{j2}^{mig}, b_{j1}^{mig})] \quad (5)$$

Results in the Appendix confirm these two solutions estimate unbiased covariances. They yield nearly identical results on the present dataset and on simulated data. Moreover, these two methods can also be used to correct variances of these fixed effects and therefore correctly estimates correlation coefficients.

Modelling heterogeneity across workplaces

The last part of this article examines the sources of variation in inequality and their links. To efficiently summarize these variations, we use gender, migrant and class gaps fixed effects and the RMSE (all estimated in equation 2) as dependent variables and perform seemingly unrelated regressions with generalized structural equations.[10]

$$\left| \begin{array}{l} b_j^{gdr} = X_j.c^{gdr} + u_j^{gdr} \\ b_j^{mig} = X_j.c^{mig} + u_j^{mig} \\ b_j^{cla} = X_j.c^{cla} + u_j^{cla} \\ RMSE_j = X_j.c^{RMSE} + u_j^{RMSE} \end{array} \right. \tag{6}$$

Simultaneous equation models are useful when there are multiple interdependent relationships among dependent variables as they account for the fact that the errors from one equation are correlated with the errors from another. The interpretation of the model is close to interpreting four separated OLS regressions with the advantage of measuring autocorrelations between the error terms of the four equations (Table A6). Comparing the full model (Equation 6) with a null model (with no control variables) enables us to measure the contribution of the independent variables to the correlation structure found in the second-stage (Table 2).

Control variables account for the diversity of *inequality regimes*. This includes industry, length of service, urbanity, wage levels and workplace composition and nine industry dummies based on INSEE's groupings – *Nomenclature d'Activités Française* (NAF).[11] The reference category for industry is transportation. A

---

[10] For these regressions, each establishment is weighted by the number of workers.

[11] NAF is the French version of the NACE rev. 2, the Statistical classification of economic activities in the European Community. The two-digit codes are as follows: Manufacturing & Mining (01-35), Construction & Utilities (36-44), Transportation (45-53), Retail & Accommodation (45-56), Information & Communications (58-63), Finance & Insurance (64-66), Real estate and technical services (68-75), Administrative Support (77-82), Social Services (84-88), and Other (90-99).

dummy for establishments located in the five most populated urban areas in France is used to account for urbanisation.[12] In addition, we control for the establishment length of service with two variables: the number of years of presence in the panel and a dummy variable coding the existence at the beginning of the panel in 1996 (to account for left censoring). There are also controls for establishment-specific wage premium (e.g., the establishment fixed effect estimated in equation 1), the log number of employees in an establishment and the proportion of "younger" (25-34) and "older" (45-55) employees, as well as the proportion of managers, migrants or women.[13] Finally, models include year fixed effects. All the continuous variables are standardized.

Results

The intersection of categorical inequalities at the individual level

We start by briefly discussing the standard intersectional inequality approach based on the individual-level results from equation 1. Figure 1 displays the full interactions between gender, migration status and class effects. It also compares the earnings of different groups relative to the most disadvantaged group – female migrant workers – the reference category. In addition, Figure A1 plots the main and marginal interactions effects from Equation 1 and Table A2 provides descriptives.

---

[12] These are the Île-de-France region (Paris), and the following *départements*, Rhône (Lyon), Bouches-du-Rhône (Marseille), Haute-Garonne (Toulouse), and Nord (Lille).

[13] We checked that our results were not significantly different when we introduced sales and profit control variables in a subsample for which we could use this information. Estimates are available upon request.

Figure 1. Wages effects for full set of interactions for each social position

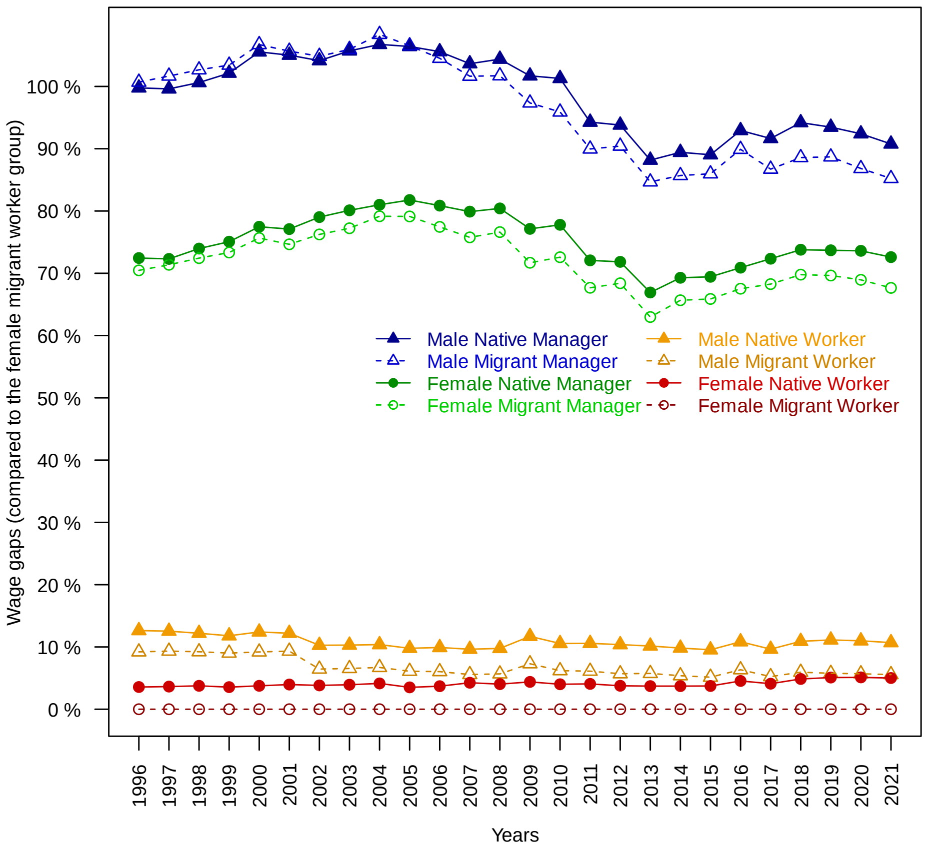


Figure 1 and the first row of Table 1 show that the class gap is the largest earnings gap. For example, managers earn on average 74 percent more than workers during the period – an effect that has been decreasing over time. This is followed by gender (with men earning 8.5 percent more than women) and migration status (with natives earnings 2.7 percent more than migrants).[14] There is evidently a cumulative disadvantage effect in line with previous research (Greenman and Xie, 2008; Woodhams, Lupton and Cowling, 2015). Female migrant workers are worse off relative to female migrant managers (a 72 percent gap), male migrant workers (a seven percent gap) and female native workers (a four percent gap) (Table A3). However, the combination of the class and gender

[14] From the first-stage equation, we post-estimate the average gender gap (and similarly average migrant and class gaps) as the average of the gender gaps measured among respectively migrant workers, native workers, migrant managers and native managers (weighted by the size of these respective four groups).

disadvantages is less than additive. Both Figures 1, A1 and Table A3 show that the gender gap for managers (+14 percent) is double that for workers (+7.0 percent). Conversely the migrant gap is stronger among workers (+4.1 percent) relative to managers (+2.0 percent on average).

Table 1. Earnings gaps variation in French establishments

	Gender	Migration status	Class	Intra-categorical inequality
First step gaps (log earnings) (percentage)	0.081 8.5%	0.026 2.7%	0.554 74.1%	/
Second step estimates Mean	-0.0050	-0.0035	0.0095	0.2273
Std Dev.	0.0700	0.0661	0.1471	0.0771
Std Dev. Corrected *"vcov"*	0.0629	0.0490	0.1425	/
Std Dev. Corrected *"halves"*	0.0626	0.0445	0.1403	0.0754
95% Gap Variation	[-5%, 22%]	[-7%, 13%]	[33%, 132%]	[0.08, 0.38]
% of employees in reverted gap establishment	11%	32%	0%	/

Note: In the first line, we derive from the first-stage equation the average gender gap (and similarly average migrant and class gaps) as the average of the gender gaps measured among respectively migrant workers, native workers, migrant managers and native managers (weighted by the size of these respective four groups).
All measures are based on log earnings except line 2 "(percentage)", which represents gaps in proportion to the earnings of the disadvantaged group.

There are some evolutions worth mentioning during the 26-year period. First, the gender gap declined and the migrant gap increased (Figure A1). The average male effect moved from +10.3 percent to +7.3 percent, the average native effect from +2.5 to +3.2. The average manager effect also declined from +73 percent to +69 percent yet with more volatile trends.

Gender, migrant and class gaps at the establishment-level

Moving to the organisational level, how do gender, migrant and class gaps shape inequality within the workplace, once individual determinants and workplace characteristics are taken into account?

Table 1 displays the results for within workplace gender, migrant and class gaps (equation 2). Since the main individual level determinants are accounted for in equation 1, the three average residual earnings gaps estimated in equation 2 are consequently close to 0. Equation 2 enables us to estimate the organisational heterogeneity of wage gaps. Table 1 shows considerable variation in the distribution of these inequalities across workplaces, measured with the corrected standard deviation values: 0.06 for gender, 0.05 for migration status, 0.14 for class and 0.08 for intra-categorical inequality. This means that for workers in workplaces one standard deviation below the average gender gap, the gender gap is reduced by 0.063 log points – almost closing the gap (0.081-0.005-0.063=0.018). To put this into perspective, 95 percent of employees work in establishments with gender gaps ranging from -5 percent to 22 percent, migrant gaps ranging from -7 percent to 13 percent, class gaps from 33 percent to 132 percent and intra-categorical inequality from 0.08 to 0.38 log points.

In 11 percent of establishments, there is a "reverse gender gap", meaning that in these workplaces women earn more than men, net of the other covariates included in Equation 1. Similarly, 32 percent of establishments have a "reverse migrant gap." Conversely, the class gap is never reversed: managers/professionals consistently earn more than workers in all establishments even after controlling for Equation 1 covariates. Thus, the class gap varies only in its positive magnitude across establishments, while some of the variation in the gender and migrant gaps relates to "reverse gaps."

These results confirm the importance of the workplace level in analysing inequality and highlight differences in the degree of between-establishment variations across single inequality factors. We now turn to the correlations between inequalities within workplaces.

The Intersection of Inequalities at the Establishment-Level: patterns of correlation

Table 2 shows the correlation matrix between the three earnings gaps with both correction methods (we focus here on the "*vcov"* method but the results across the two correction methods are qualitatively the same). There is a relatively strong negative correlation coefficient between the gender and migration status inequalities (-0.216), which remains consistent over the period (Figure A2). The gender×class gap and migration_status×class gap correlations are non-significant and much smaller in magnitude (-0.004 and +0.006 respectively).

Thus, comparing the relationship between the categorical inequalities at the workplace level, these results point to an inequality trade-off that is particularly pronounced between gender and migration status. Interestingly, these findings contrast with the individual level interaction effects. At the individual level, there was no indication the gender gap differed between migrants and natives. At the establishment level, more gender-equal workplaces tend to be more migrant-unequal. There are also diverging results between the establishment and individual level by class as well. The interaction terms in Equation 1 show that gender×class and migration status×class are strongly and negatively associated but this negative association vanishes at the establishment level. These discrepancies between the individual and establishment level findings confirms the value of a workplace-level analysis of intersectional inequality.

Table 2. Corrected correlation of inequality measures

	Gender	Migration status	Class	Intra-categorical inequality
Gender	Method 2 halves→	-0.235	-0.007	0.327
	Method 1 vcov ↓	[-0.247; -0.224]	[-0.016; 0.002]	[0.318; 0.336]
Migration status	-0.216		0.004	-0.109
	[-0.226; -0.206]		[-0.006; 0.014]	[-0.122; -0.096]
Class	-0.004	0.006		0.086
	[-0.012; 0.005]	[-0.003; 0.015]		[0.076; 0.095]
Intra-categorical inequality	/	/	/	

Note: 95% confidence intervals in square brackets were calculated with bootstrapping (4000 samples).

Next, we ask if some categorical inequalities are more intense in more unequal workplaces. Intra-categorical inequality are strongly and positively correlated with the gender gap (0.327) and negatively, although less intensively, correlated with the migrant gap (-0.109). Intra-categorical inequality is only mildly (but significantly) correlated with the class gap (0.086) and this correlation is less consistent throughout the period (Figure A3).

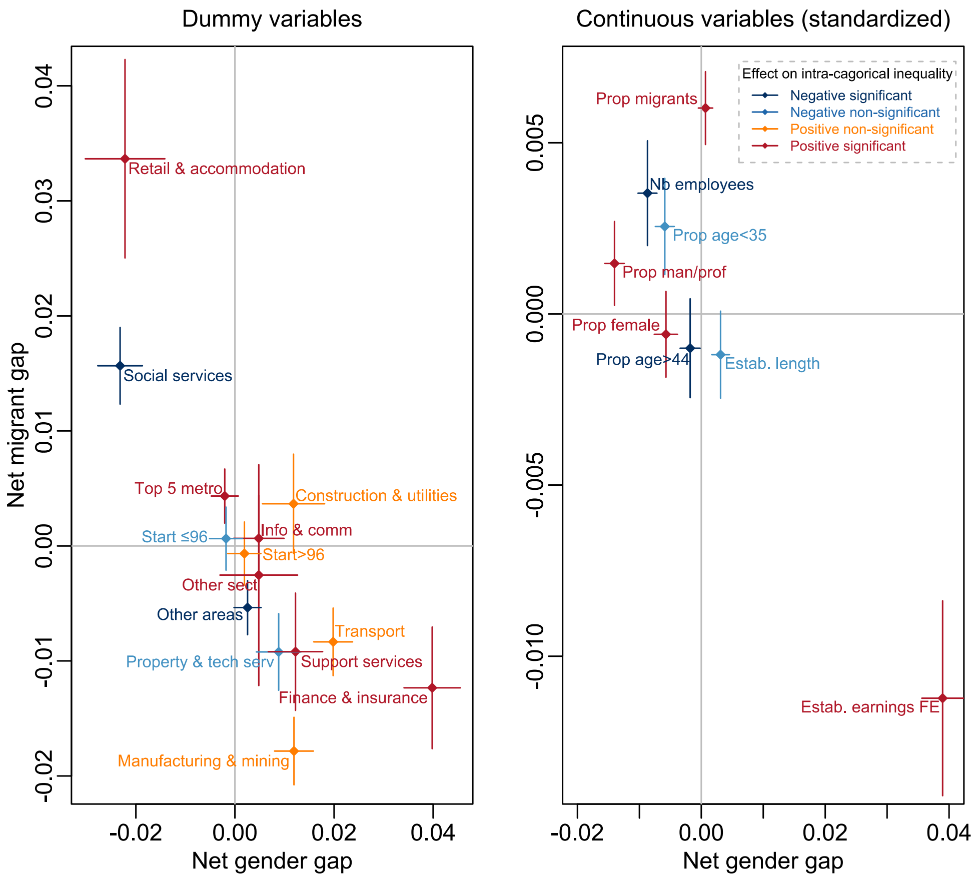
Taken together, the inquiry into whether within workplace inequality regimes follow a reinforcing or a trade-off pattern shows evidence for both. In line with H1, there is reinforcing inequality between the gender gap and the intra-categorical earnings inequality. Class and intra-categorical inequalities also tend to reinforce each other although their correlation magnitude is small. In line with H2, there is evidence of inequality trade-offs between the gender and migrant gaps. Moreover, there are lower migrant earnings gaps in more egalitarian workplaces overall. This partial evidence for both H1 and H2 does not follow a general rule of association along the categorical/non categorical or exogenous/endogenous analytical lines.

Heterogeneity across workplaces

This section discusses the contexts driving the most salient findings: the negative correlation between gender and migrant gaps, the negative correlation between intra-categorical inequality and migrant gap and the positive correlation between gender gap and intra-categorical inequality. To do so, we estimate the third-stage model (equation 6), described above. In Table A6, by comparing the residual correlation in the null model (e.g., -0.24 for gender×migration) with the full model (e.g., -0.06 for gender×migration), this model model accounts for 76 percent of the negative correlation between gender and migrant gaps, 38 percent of the negative correlation of migrant gap and intra-categorical inequality, but only of 13 percent of the positive correlation of gender gap and intra-categorical inequality.

In Figure 2, we visually display the migration×gender×intra-categorical inequalities (full results with the coefficients of the four equations are shown in Table A5). This figure shows the heterogeneity of inequality regimes. The y axis reports the coefficients for the migrant gap while the x axis reports the coefficients for the gender gap (from Table A5). In Figure 2, we separate the coefficients based on dummy variables (left panel) from the standardized continuous variables (right panel). For the sake of readability, the dummy variables are rescaled to contrast them to the mean inequality rather than a reference category. Variables on the upper-left hand quadrant are positively associated with the migrant gap and negatively associated with the gender gap while those in the lower right-hand quadrant show the reverse pattern. In addition, the colour scale represents sign and significance on a virtual third axis representing intra-categorical inequalities.

Figure 2. Parameter estimates for gender, migration and intra-categorical inequalities



Note: Effect of contextual variables on the net gender gap (horizontal scale), the net migrant gap (vertical scale) and on intra-categorical inequality (colour scale). In the left panel, the estimates for dummy variables are rescaled so that the weighted parameters of the same categorical variable sum to zero. This transformation allows us to display the reference categories (e.g. "Transportation" for the sector variable, "Other areas" for the top five metropolitan area variable and "Start ≤96" for the "Start>96" variable), and to make the centre of the figure correspond to the mean gender and migrant gaps. In the right panel, standardized coefficients of continuous variables are shown to facilitate comparisons. Estimates taken from Table A4.

Figure 2 shows a diagonal trend moving from the upper-left to lower-right hand sides. Sector and establishment wage premium strongly contribute to workplace inequalities. Based on them, we classify three important inequality regimes in France: 1a) low gender gap, high migrant gap and low intra-categorical inequality, 1b) low gender gap, high migrant gap and high intra-categorical inequality and 2) high gender gap, low migrant gap and high intra-categorical inequality.[15]

[15] If we quantify the importance of these inequality regimes based on positive/negative male and non-migrant fixed effects and above/below median intracategorical inequality, we find the

Retail, accommodation, social services and large workplaces are characterized by a relatively high migrant gap and a relatively low gender gap. However, workplaces in these categories differ in their level of intra-categorical inequality; lower in large workplaces and social service sectors, while higher in retail and accommodation. These results relate to previous findings. The lower level of workplace inequality in the public sector is a feature common to many OECD countries (Tomaskovic-Devey *et al.*, 2020). France does have a relatively lower gender wage gap within the accommodation and retail sector (Coudin, Maillard and Tô, 2018) and social services are less gender unequal (Allen and Sanders, 2002). The relatively high migrant penalty in a sector like social services is in line with evidence of immigrant disadvantage in the French public sector (Fougère and Pouget, 2003). Moreover, since many public service positions require European citizenship, migrants tend to be hired in these sectors under conditions that are more precarious and with temporary contracts. This is namely the case in hospitals and local municipal services, which are included in the data (OECD, 2008).

In contrast, the finance and insurance sector (and to a lesser extent manufacturing, transport and support services) as well as higher paying establishments tend to have strongly and positively associated with high gender gaps, high intra-categorical inequality and low migrant gaps. The financial sector is well documented as high paying, highly masculine and gender biased (Roth, 2006; Lin and Neely, 2017). The fact this sector attracts highly qualified migrants could also explain why this sector stands low on the migrant gap scale. However, other sectors like manufacturing, recruiting less qualified migrants, score even lower on the migrant gap scale.

---

following distribution. 1a. 19.1 percent of workers work in an inequality regime with low gender gap, high migrant gap, and low intracategorical inequality; 1b. 13.8 percent in an inequality regime with low gender gap, high migrant gap and high intra-categorical inequality; 2. 14.5 percent in an inequality regime with high gender gap, low migrant gap and high intra-categorical inequality. An inequality regime with an alignment of these three forms of inequality (e.g., high gender gap, high migrant gap, and high intra-categorical inequality) does exist, but is less common, comprising 10.4 percent of the workforce.

Discussion and Conclusion

This article combines recent interests in intersectional and organisational mechanisms to ask how different types of inequalities relate to each other within French workplaces. Building specifically on Acker's (2006) idea of *inequality regimes,* it investigates whether French workplaces are shaped by *reinforcing inequality* or *inequality trade-off* structures. It does so with innovative methodological techniques to net out measurement biases that would ordinarily arise from correlating different forms of earnings gaps. This is a key contribution of this article.

This framework allows us to uncover striking patterns of correlation in workplace inequalities that have been largely overlooked in the previous literature and that differ substantially from the usual individual-level intersection of inequalities. More specifically, this articles finds a positive correlation between intra-categorical inequality and the gender gap (r=0.3), negative between the former and the migrant gap (r=-0.1) and negative between the gender and migrant gaps (r=-0.2). Industrial sectors and workplace wage premiums contribute to these patterns and cluster inequality regimes in three dominant groups: low gender gap, high migrant gap and low intra-categorical inequality for example in the social services sector; low gender gap, high migrant gap and high intra-categorical inequality for example in the retail trade and accommodation sector; and high gender gap, low migrant gap and high intra-categorical inequality for example in the finance and insurance sector.

These results show that there is no "one story" when it comes to the structure of inequality regimes. Indeed, there is evidence for both *reinforcing* and *trade-off inequality* structures. Thus, these findings first challenges the dominant approach which often implicitly and sometimes explicitly assumes that organisational inequalities are mutually reinforcing. The common assumption that the organisational level of tolerance to inequality produces a positive correlation structure between different types of inequality is contradicted. Moreover, exogenous categorical inequalities do not operate in the same way and serve as salient anchors for durable inequalities. Unlike gender, migration-based inequality does not go hand in hand with intra-categorical inequality. This discrepancy is also at odds with the main argument underlying the *inequality trade-*

off perspective according to which categorical wage gaps should be reduced in the context of high inequality if the latter results from high productivity-related wage dispersion. While this type of argument is supported for migration status, it is not supported for gender.

The positive association between gender and intra-categorical inequality is in line with the literature on the congruence between gender stereotypes and dominant perceptions of labour market productivity. A close look at the heterogeneity of this pattern across workspaces highlights the role of sectors such as the finance industry that is a striking illustration of the coexistence of high gender gap and high wage inequality (Roth, 2006; Lin and Neely, 2017).

By contrast, the trade-off structure documented between the migrant gap and intra-categorical inequality could be a byproduct of industrial relations and the "dualisation" of the labour market. Historically, French trade unions tried to include migrants, especially in the 1930s and in the post-WWII period (Noiriel, 1988). However, with the rise of French colourblind narrative in the 1980s, addressing social inequalities through the lens of ethnic minorities or migrants came to be seen as contrary to the principles of French universalism. Even left-leaning unions embraced this "French republicanism" in a way that failed to address specific problems faced by migrant workers (Safi, 2017; Bataille, 2018). In some sectors like public utilities and social services, unions are even eager to defend insider worker status. In such dualistic contexts, migrant worker's legal vulnerability translates into lower wages. This may explain why larger workplaces and social service sectors, where unions are involved in defending insider workers status, combine narrower pay dispersion with larger migrant gaps.

The negative correlation between the gender and migrant gaps is among the most novel findings of this study, remaining robust across alternative first-stage specifications (see Table A7). This result challenges the implicit assumption in organisational inequality research that gender and migration status (or ethno-racial) inequalities go hand in hand (McGuire, 2002). Indeed, these two types of inequality are often assumed to share underlying mechanisms (e.g., discrimination) and to be similarly affected by the same organisational processes, leading to expectations of, if not a direct, then at least an indirect positive correlation. This article provides strong empirical evidence to the contrary. Workplaces with large gender gaps generally tend to have smaller migrant gaps and vice versa. While the data do not allow for a precise interpretation of this inequality trade-off, it partly reflects the divergent temporal evolution of these two types of inequality over the period of observation. Indeed, the average gender gap has decreased by 3 percentage points, from 10.3 to 7.3 percent, while the average migrant gap has increased by 0.7 percentage points, from 2.5 to 3.2 percent (Figure A1). While this study is the first to document this pattern of a negative correlation between gender and migration status wage gaps within workplaces, further research is needed to investigate whether similar patterns exist in other contexts and to uncover the underlying mechanisms driving these relationships.

Despite these contributions, this study contains several important limitations that warrant discussion.

First, the magnitude and in some cases the signs of the inequality correlations depend on the variables used in the first stage (equation 1) (Table A7). However, alternative specifications such as no first stage or a first stage with minimum control variables provide a global qualitative confirmation of the key correlation structure.[16] This allows us to be confident in the robustness of these results. Ideally, educational level should be controlled for the first stage equation, as it may be considered as more exogenous than occupation. Unfortunately, the BTS data does not contain any education variable and we use the French 2-digit occupational classification as an alternative socio-economic variable. This variable is strongly correlated with diploma and constitutes a good proxy of the latter. However, it is also partially endogenous to workplace policy and shaped by occupational segregation and glass ceiling practices. It would be also useful to control for parental status (interacted with gender) as it drives family arrangement and involvement at work as a consequence. Unfortunately, this variable is also absent from the French data. However, future research on

[16] Interestingly, in the no first-stage alternative specification, the class gap tends to replace the intra-categorical inequality in the correlation structure. Moreover, the negative correlation between migrant gap and intra-categorical inequality disappears. However, controlling for obvious variables (minimal first stage) such as working time suffice to yield similar results to the full first-stage.

countries with detailed register data containing both education and family composition could help overcome this double limitation.

Second, the measure of migration based on those born outside of France overlooks the heterogeneity of the migrant groups and imperfectly mirrors ethno-racial origin. While this is widely used measure of immigrant status, it cannot account for important labour-market factors such as age at arrival, fluency in French, abroad accreditation and other aspects that affect immigrant socioeconomic attainment in host countries (Tomaskovic-Devey, Hällsten and Avent-Holt, 2015; Melzer *et al.*, 2018). Perhaps most importantly, it does not capture the racialized dimensions of French immigration from North and Sub-Saharan Africa. While important, the broad-based migration estimate should therefore be regarded as a conservative case for the study of intersectional inequality.

Finally, this empirical investigation of intersectional inequality only focuses on wages while research shows that other labour market measures of inequality (such as inactivity, unemployment or occupational segregation) are strongly racialized and gendered (Altonji and Blank, 1999). Moreover, the intersectional literature also calls attention to other experiences of inequality in workplaces such as promotions and demotions, harassment and microaggressions (Roscigno, 2019). Our focus on wage inequality makes the results more conservative here as well.

This research proposed to map the structure of workplace inequalities but remains largely silent on the causal relations between these inequalities. Future research needs therefore to address this causality issue directly. Going forward, researchers should consider specifying inequality regimes and, especially, to study the effects inequality regimes have on the workplace. Different regimes can lead to different organisational environments and interpersonal relations. Future research should consider how overall levels of inequality and the multiple forms in which they relate affect outcomes like workplace safety, anti-discrimination policies, collective bargaining capacity, etc.. If inequalities move together, for example, it compounds these effects within that space. If there are inequality trade-offs, this can alter the balance of power and claims making strategies available to actors. Addressing these questions can account for the multifaceted character of inequality and its various socio-economic consequences.

References

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

Acker, J. (2006) 'Inequality regimes: Gender, class, and race in organizations', *Gender & society*, 20(4), pp. 441–464.

Aigner, D.J. and Cain, G.G. (1977) 'Statistical theories of discrimination in labor markets', *ILR Review*, 30(2), pp. 175–187.

Allen, J. and Sanders, K. (2002) 'Gender gap in earnings at the industry level', *European Journal of Women's Studies*, 9(2), pp. 163–180.

Altonji, J.G. and Blank, R.M. (1999) 'Race and gender in the labor market', *Handbook of labor economics*, 3, pp. 3143–3259.

Amable, B. (2017) *Structural crisis and institutional change in modern capitalism: French capitalism in transition*. Oxford University Press.

Andrews, M.J. *et al.* (2008) 'High wage workers and low wage firms: negative assortative matching or limited mobility bias?', *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 171(3), pp. 673–697.

Athari, E., Brinbaum, Y. and Lê, J. (2019) 'Le rôle des origines dans la persistance des inégalités d'emploi et de salaire', in *Insee Références, Emploi, chômage, Revenus du travail*. Insee.

Avent-Holt, D. and Tomaskovic-Devey, D. (2012) 'Relational inequality: Gender earnings inequality in US and Japanese manufacturing plants in the early 1980s', *Social Forces*, 91(1), pp. 157–180.

Babet, D., Godechot, O. and Palladino, M.G. (2023) 'In the Land of AKM: Explaining the Dynamics of Wage Inequality in France'. Insee (Documents de Travail, 2023/20).

Baron, J.N. and Bielby, W.T. (1980) 'Bringing the firms back in: Stratification, segmentation, and the organization of work', *American sociological review*, 45(5), pp. 737–765.

Bataille, P. (2018) 'The French trade union movement and the struggle against racism in the workplace', in *Cultural Diversity in Trade Unions: A Challenge to Class Identity?* Routledge, pp. 227–238.

Baxter, J. and Wright, E.O. (2000) 'The glass ceiling hypothesis: A comparative study of the United States, Sweden, and Australia', *Gender & society*, 14(2), pp. 275–294.

Boltanski, L. (1982) *Les cadres. La formation d'un groupe social*. Paris: Éditions de Minuit.

Browne, I. and Misra, J. (2003) 'The intersection of gender and race in the labor market', *Annual review of sociology*, 29(1), pp. 487–513.

Castilla, E.J. (2008) 'Gender, race, and meritocracy in organizational careers', *American journal of sociology*, 113(6), pp. 1479–1526.

Charles, M. and Bradley, K. (2009) 'Indulging our gendered selves? Sex segregation by field of study in 44 countries', *American journal of sociology*, 114(4), pp. 924–976.

Cho, S., Crenshaw, K.W. and McCall, L. (2013) 'Toward a field of intersectionality studies: Theory, applications, and praxis', *Signs: Journal of women in culture and society*, 38(4), pp. 785–810.

Connelly, B.L. *et al.* (2014) 'Tournament theory: Thirty years of contests and competitions', *Journal of management*, 40(1), pp. 16–47.

Coudin, É., Maillard, S. and Tô, M. (2018) 'Family, firms and the gender wage gap in France'. Insee (Documents de Travail de la DSDS, F1805).

Crenshaw, K. (1989) 'Demarginalizing the intersection of race and sex: A black feminist critique of antidiscrimination doctrine, feminist theory and antiracist politics', *University of Chicago Legal Forum*, 1989(1), pp. 139–167.

Desjonquères, A. *et al.* (2021) 'Quels sont les métiers des immigrés?', *Dares Analyses*, (36), pp. 1–8.

Dobbin, F., Schrage, D. and Kalev, A. (2015) 'Rage against the iron cage: The varied effects of bureaucratic personnel reforms on diversity', *American Sociological Review*, 80(5), pp. 1014–1044.

Eagly, A.H. and Karau, S.J. (2002) 'Role congruity theory of prejudice toward female leaders.', *Psychological review*, 109(3), p. 573.

Erikson, R. and Goldthorpe, J.H. (1992) *The constant flux: A study of class mobility in industrial societies*. Oxford: Oxford University Press.

Fougère, D. and Pouget, J. (2003) 'Les déterminants économiques de l'entrée dans la fonction publique', *Économie et statistique*, 369(1), pp. 15–48.

Greenman, E. and Xie, Y. (2008) 'Double jeopardy? The interaction of gender and race on earnings in the United States', *Social Forces*, 86(3), pp. 1217–1244.

Grodsky, E. and Pager, D. (2001) 'The structure of disadvantage: Individual and occupational determinants of the black-white wage gap', *American Sociological Review*, 66(4), pp. 542–567.

Hedström, P. (1991) 'Organizational differentiation and earnings dispersion', *American Journal of Sociology*, 97(1), pp. 96–113.

Hermansen, A.S. *et al.* (2023) 'Immigrant–native pay gap driven by lack of access to high-paying jobs'. SocArXiv. Available at: https://doi.org/10.31235/osf.io/2p4vw.

Holmström, B. (1979) 'Moral Hazard and Observability', *The Bell Journal of Economics*, 10(1), pp. 74–91.

Insee (2023) *Immigrés et descendants d'immigrés en France: édition 2023*. INSEE.

Kvande, E. and Rasmussen, B. (1994) 'Men in male-dominated organizations and their encounter with women intruders', *Scandinavian Journal of Management*, 10(2), pp. 163–173.

Lazear, E.P. and Rosen, S. (1981) 'Rank-order tournaments as optimum labor contracts', *Journal of political Economy*, 89(5), pp. 841–864.

Lin, K.-H. and Neely, M.T. (2017) 'Gender, parental status, and the wage premium in finance', *Social Currents*, 4(6), pp. 535–555.

McCall, L. (2001) *Complex inequality: Gender, class and race in the new economy*. Routledge.

McCall, L. (2005) 'The complexity of intersectionality', *Signs: Journal of women in culture and society*, 30(3), pp. 1771–1800.

McGuire, G.M. (2002) 'Gender, race, and the shadow structure: A study of informal networks and inequality in a work organization', *Gender & society*, 16(3), pp. 303–322.

Melzer, S.M. *et al.* (2018) 'A relational inequality approach to first-and second-generation immigrant earnings in German workplaces', *Social Forces*, 97(1), pp. 91–128.

Meurs, D. and Pora, P. (2019) 'Gender equality on the labour market in France: A slow convergence hampered by motherhood', *Economie et Statistique*, 510(1), pp. 109–130.

Milgrom, P. and Roberts, J. (1988) 'An economic approach to influence activities in organizations', *American Journal of sociology*, 94, pp. S154–S179.

Noiriel, G. (1988) *Le creuset français*. Paris: Seuil.

OECD (2008) *Jobs for Immigrants (Vol. 2): Labour Market Integration in Belgium, France, the Netherlands and Portugal*. OECD Publishing.

Penner, A.M. *et al.* (2023) 'Within-job gender pay inequality in 15 countries', *Nature Human Behaviour*, 7(2), pp. 184–189.

Penner, A.M. and Saperstein, A. (2013) 'Engendering racial perceptions: An intersectional analysis of how social status shapes race', *Gender & Society*, 27(3), pp. 319–344.

Quillian, L. *et al.* (2019) 'Do some countries discriminate more than others? Evidence from 97 field experiments of racial discrimination in hiring', *Sociological Science*, 6, pp. 467–496.

Roscigno, V.J. (2019) 'Discrimination, sexual harassment, and the impact of workplace power', *Socius*, 5, p. 2378023119853894.

Roscigno, V.J., Hodson, R. and Lopez, S.H. (2009) 'Workplace incivilities: the role of interest conflicts, social closure and organizational chaos', *Work, Employment and Society*, 23(4), pp. 747–773.

Roth, L.M. (2006) *Selling women short: Gender and money on Wall Street*. Princeton University Press.

Safi, M. (2017) 'Promoting diversity in French workplaces: Targeting and signaling ethnoracial origin in a colorblind context', *Socius*, 3, p. 2378023117728834.

Safi, M. and Simon, P. (2013) 'Les discriminations ethniques et raciales dans l'enquête "Trajectoires et Origines"', *Economie et Statistique/Economics and Statistics*, 464(1), pp. 245–275.

Tilly, C. (1998) *Durable inequality*. University of California Press.

Tomaskovic-Devey, D. *et al.* (2020) 'Rising between-workplace inequalities in high-income countries', *Proceedings of the National Academy of Sciences*, 117(17), pp. 9277–9283.

Tomaskovic-Devey, D. and Avent-Holt, D. (2019) *Relational inequalities: An organizational approach*. New York: Oxford University Press.

Tomaskovic-Devey, D., Hällsten, M. and Avent-Holt, D. (2015) 'Where do immigrants fare worse? Modeling workplace wage gap variation with longitudinal employer-employee data', *American Journal of Sociology*, 120(4), pp. 1095–1143.

Treiman, D.J. (2013) *Occupational prestige in comparative perspective*. Elsevier.

Weeden, K.A. and Grusky, D.B. (2005) 'The case for a new class map', *American Journal of Sociology*, 111(1), pp. 141–212.

Williams, K.Y. and O'Reilly III, C.A. (1998) 'Demography and diversity in organizations: A review of 40 years of research', *Research in organizational behavior*, 20, pp. 77–140.

Woodhams, C., Lupton, B. and Cowling, M. (2015) 'The snowballing penalty effect: multiple disadvantage and pay', *British Journal of Management*, 26(1), pp. 63–77.

Woolridge, J.M. (2013) *Introductory econometrics: a modern approach*. Mason: South Western, Cengange Learning.

Wright, E.O. (1997) *Class counts: Comparative studies in class analysis*. Cambridge, UK: Cambridge University Press.

Zhou, X. and Wodtke, G.T. (2019) 'Income stratification among occupational classes in the United States', *Social Forces*, 97(3), pp. 945–972.

Organisational Intersectionality: Do Gender and Migration Status Inequalities Reinforce or Offset each other in French Workplaces?

Olivier Godechot

Sciences Po, CRIS, CNRS and AxPo.

Axa Chairholder in Economic Sociology

Mirna Safi

Sciences Po, CRIS, CNRS

Matthew Soener

University of Illinois – Urbana/Champaign

Online Appendix

Table of content

1. Supplementary tables

Table A1. Description of the initial and estimation sample.

Year	Workers	Workplaces	Male	Natives	Managers	Log earnings (std)
Initial sample						
1996	13,376,312	1,223,949	59%	89%	13%	10.06 (0.57)
2008	15,973,476	1,459,538	56%	88%	17%	10.17 (0.56)
2021	17,069,986	1,607,469	54%	85%	23%	10.20 (0.58)
Average	15,464,721	1,425,227	56%	88%	17%	10.16 (0.57)
Estimation sample						
1996	3,486,918	7,990	52%	89%	21%	10.31 (0.53)
2008	4,069,411	10,082	50%	89%	29%	10.40 (0.57)
2021	4,329,112	11,770	48%	86%	37%	10.44 (0.59)
Average	3,906,832	9,805	50%	88%	29%	10.40 (0.57)

Note: Standard deviation of log earnings in parentheses. Log earnings corrected for inflation and expressed in log euros 2021. We display estimates for the first, middle and last year and the average for the full period. Estimates for other years between are available upon request.

Table A2. Description of the average variables used for the first-stage regressions.

	Mean	Std Dev	Min	Max
Log earnings	10.396	0.565	8.94	15.89
Male	0.498	0.500	0	1
Native	0.884	0.320	0	1
Male × Manager	0.183	0.386	0	1
Native × Manager	0.252	0.432	0	1
Male × Native	0.435	0.496	0	1
Male × Native × Manager	0.157	0.363	0	1
Part-time	0.121	0.325	0	1
Age	40.000	8.885	25	55
$(\text{Age})^2$	1,688.373	698.384	625	3,025
Number of hours	1,652.808	435.731	1	5,436
$(\text{Number of hours})^2$	2,925,728.765	1,192,669.550	1	44,540,207
Number of days	331.231	69.769	1	401
$(\text{Number of days})^2$	114,638.010	34,424.636	1	165,488
Occupation: Artisans and retailers	4.6E-5	0.006	0	1
CEOs (10+ firms)	0.001	0.035	0	1
Professionals	0.003	0.044	0	1
Civil servant managers	0.020	0.140	0	1
Professors and scientific professions	0.038	0.192	0	1
Artists and media professionals	0.011	0.102	0	1
Managers	0.110	0.311	0	1
Engineers	0.111	0.312	0	1
Primary school teachers	0.006	0.080	0	1
Health and social workers	0.100	0.300	0	1
Clergy	1.1E-4	0.011	0	1
Public administration intermediates	0.019	0.137	0	1
Business administration intermediates	0.080	0.271	0	1
Technicians	0.065	0.246	0	1
Intermediate supervisors	0.018	0.131	0	1
Public administration clerks	0.155	0.360	0	1
Security agents	0.009	0.093	0	1
Business administration clerks	0.076	0.264	0	1
Retail salespersons	0.015	0.122	0	1
Personal service employees	0.009	0.093	0	1
Skilled manufacturing workers	0.079	0.269	0	1
Skilled artisans	0.020	0.140	0	1
Drivers	0.005	0.073	0	1
Handling, transport skilled workers	0.015	0.122	0	1
Unskilled manufacturing workers	0.027	0.160	0	1
Unskilled artisans	0.007	0.083	0	1
Farm workers	1.5E-4	0.012	0	1

Note: We show here the 26-year average of the annual descriptive statistics. The average number of observations is 3,906,832. In the regressions, we also interact detailed occupations with age and age squared, for which we do not show descriptives due to space limitations.

Table A3. First-stage estimates for first, middle, last year, and average

	1996			2008			2021			96-21#
	Beta		Std. Err.	Beta		Std. Err.	Beta		Std. Err.	Mean
Male	0.088	***	(0.004)	0.055	***	(0.005)	0.054	***	(0.006)	0.065
Native	0.035	***	(0.002)	0.040	***	(0.002)	0.049	***	(0.004)	0.040
Male × Manager	0.075	***	(0.006)	0.078	***	(0.008)	0.046	***	(0.008)	0.069
Native × Manager	-0.024	***	(0.005)	-0.018	***	(0.003)	-0.020	*	(0.008)	-0.021
Male × Native	-0.004		(0.004)	-0.002		(0.003)	-0.001		(0.005)	-0.002
Male × Native × Manager	-0.012		(0.006)	-0.007		(0.004)	0.002		(0.007)	-0.007
Part time	0.012		(0.009)	0.076	***	(0.004)	-0.028	**	(0.010)	0.028
Age	4.7E-2	***	(1.4E-3)	1.1E-2	***	(7.5E-4)	2.2E-2	***	(7.8E-4)	2.2E-2
(Age)²	-4.9E-4	***	(1.7E-5)	-5.0E-5	***	(1.0E-5)	-2.0E-4	***	(9.0E-6)	-1.9E-4
Number of hours	1.4E-3	***	(2.4E-5)	1.2E-3	***	(5.5E-5)	1.1E-3	***	(3.1E-5)	1.2E-3
(Number of hours)²	-2.0E-7	***	(1.0E-8)	-1.0E-7	***	(2.0E-8)	-2.0E-7	***	(1.0E-8)	-1.7E-7
Number of days	-1.5E-4		(3.1E-4)	5.7E-4		(5.4E-4)	1.6E-3	***	(3.7E-4)	9.6E-4
(Number of days)²	8.0E-7		(6.8E-7)	-1.1E-6		(1.1E-6)	-1.1E-6		(6.8E-7)	-1.2E-6
Occupation (27 dummies)	Yes			Yes			Yes			Yes
Occupation × Age	Yes			Yes			Yes			Yes
Occupation × (Age)²	Yes			Yes			Yes			Yes
Establishment fixed effect	Yes			Yes			Yes			Yes
Number of observations	3,486,918			4,069,411			4,329,112			3,906,832
Manager main effect*	0.533			0.569			0.516			0.541

Note: OLS estimates with robust standard errors clustered at the French *département* level in parentheses. *** p<0.001, ** p<0.01, * p<0.05
We only show the first-stage estimates here for the first, middle, and last years. Other years and detailed occupation estimates are available upon request.
We compute the simple average of annual parameters in the last column.
* We post estimate the manager main effect from the detailed occupation parameters. Cf. Appendix 3.2.

Table A4. Uncorrected covariance and correlation coefficient, and correlated coefficient

		Gender FE	Migration status FE	Class FE	Intra-categorical inequality
Uncorrected Covariance					
	Gender FE		-0.00067	-0.00019	0.00158
	Migration status FE			0.00008	-0.00038
	Class FE				0.00094
	Intra-categorical inequality				
Uncorrected Correlation coefficient					
	Gender FE		-0.144	-0.019	0.294
			[-0.148; -0.14]	[-0.023; -0.015]	[0.29; 0.297]
	Migration status FE			0.008	-0.074
				[0.004; 0.012]	[-0.078; -0.071]
	Class FE				0.083
					[0.079; 0.087]
	Intra-categorical inequality				
Corrected covariance					
	Gender FE	*Method 1 vcov* ↓	-0.00066	-0.00006	0.00154
	Migration status FE	-0.00067	*Method 2 halves* →	0.00002	-0.00037
	Class FE	-0.00003	0.00004		0.00091
	Intra-categorical inequality				

Note: We weight workplaces by the number of workers and, we additionally reweight the sample so that all years have equal weights.

Table A5. Generalized Structural Equation Estimates of Workplace Inequalities

	Descriptives Mean (Std. dev)	Gender FE	Migration status FE	Class FE	Intra-categorical inequality (RMSE)
Industries: Manufacturing and mining	0.2113	-0.0079***	-0.0095***	0.0560***	0.0005
	(0.4082)	(0.0020)	(0.0015)	(0.0082)	(0.0021)
Construction & utilities	0.0183	-0.0080*	0.0120***	0.0640***	0.0009
	(0.1341)	(0.0032)	(0.0022)	(0.0081)	(0.0025)
Transportation	0.1224	*Ref*	*Ref*	*Ref*	*Ref*
	(0.3278)				
Retail & accommodation	0.0080	-0.0420***	0.0420***	-0.0099	0.0140*
	(0.0890)	(0.0041)	(0.0044)	(0.0140)	(0.0071)
Information and communication	0.0539	-0.0150***	0.0090***	-0.0097	0.0071*
	(0.2257)	(0.0026)	(0.0019)	(0.0081)	(0.0031)
Finance and insurance	0.0773	0.0200***	-0.0040	0.0083	0.0260***
	(0.2671)	(0.0029)	(0.0027)	(0.0110)	(0.0043)
Property & technical services	0.0730	-0.0110***	-0.0009	0.0070	-0.0037
	(0.2602)	(0.0023)	(0.0017)	(0.0073)	(0.0034)
Administrative support services	0.0215	-0.0076**	-0.0009	-0.0750***	0.0240***
	(0.1449)	(0.0028)	(0.0026)	(0.0088)	(0.0024)
Social services	0.4008	-0.0430***	0.0240***	0.0046	-0.0120***
	(0.4901)	(0.0023)	(0.0017)	(0.0100)	(0.0023)
Other industry	0.0136	-0.0150***	0.0058	-0.0410***	0.0330***
	(0.1157)	(0.0040)	(0.0049)	(0.0120)	(0.0047)
Top 5 metropolitan areas (dummy)	0.5529	-0.0046**	0.0097***	-0.0380***	0.0094***
	(0.4972)	(0.0014)	(0.0012)	(0.0044)***	(0.0013)
Establishment starts after 1996 (dummy)	0.4875	0.0037*	-0.0013	0.0120*	0.0024
	(0.4998)	(0.0017)	(0.0014)	(0.0054)	(0.0019)
Establishment's length of service*	8.4348	0.0031***	-0.0012	0.0039	-0.0009
	(6.6190)	(0.0007)	(0.0006)	(0.0028)	(0.0009)
Establishment's earnings fixed effect*#	10.5320	0.0390***	-0.0112***	0.0211***	0.0522***
	(0.6606)	(0.0017)	(0.0015)	(0.0054)	(0.0022)*
Number of employees (log)*	6.7122	-0.0087***	0.0035***	0.0012	-0.0067***
	(1.2601)	(0.0008)	(0.0008)	(0.0037)	(0.0012)**
Proportion of younger employees*	0.1745	-0.0059***	0.0026***	0.0034	-0.0004
	(0.0983)	(0.0008)	(0.0007)	(0.0029)	(0.0010)
Proportion of older employees*	0.4590	-0.0018*	-0.0010	-0.0167***	-0.0061***
	(0.1392)	(0.0008)	(0.0007)	(0.0029)	(0.0009)
Proportion of managers/professionals*	0.2904	-0.0140***	0.0015*	-0.0073**	0.0220***
	(0.2503)	(0.0008)	(0.0006)	(0.0028)	(0.0012)
Proportion of migrants*	0.1154	0.0007	0.0060***	0.0194***	0.0073***
	(0.0970)	(0.0006)	(0.0005)	(0.0017)	(0.0006)
Proportion of women*	0.5042	-0.0057***	-0.0006	0.0274***	0.0139***
	(0.2284)	(0.0010)	(0.0006)	(0.0027)	(0.0008)
Constant		-0.6000***	0.1600***	-0.4000***	-0.7100***
		(0.0310)	(0.0260)	(0.1100)	(0.0390)
Year FE		Yes	Yes	Yes	Yes
Number of obs. (e.g., establishments)		243,646	243,646	243,646	243,646

Note: In the first column, we show descriptive statistics for all independent variables (e.g., the mean, followed by the standard deviation in parentheses). In columns, 2 to 5, we show the parameters of the GSEM regressions with the robust standard errors clustered by firm in parentheses. *** p<0.001, ** p<0.01, * p<0.05

We weight establishments by the number of employees and, we additionally reweight the sample so that all years have equal weights.

* In the regressions, we show the standardized coefficients.

We use the first-stage establishment fixed effects.

Table A6. Estimates of variance and covariance of residuals

	GSEM estimates		"Halves" correction		"Halves" corrected correlations		Con-tribu-tion
Estimates	Null	Full	Null	Full	Null	Full	
Var(gender FE)	0.0048***	0.0037***	0.0039***	0.0028***			
	(0.00008)	(0.00006)	(0.00007)	(0.00006)			
Var(migration status FE)	0.0043***	0.0040***	0.0020***	0.0016***			
	(0.00008)	(0.00008)	(0.00006)	(0.00006)			
Var(class FE)	0.021***	0.020***	0.020***	0.018***			
	(0.00049)	(0.00051)	(0.00048)	(0.00050)			
Var(RMSE)	0.0059***	0.0032***	0.0056***	0.0029***			
	(0.00016)	(0.00007)	(0.00016)	(0.00007)			
cov(gender FE, mig. st. FE)	-0.00067***	-0.00013***	-0.00066***	-0.00012***	-0.236	-0.057	76%
	(0.00005)	(0.00003)	(0.00005)	(0.00003)			
cov(gender FE, class FE)	-0.00019	-0.000053	-0.000062	0.000072	-0.007	0.010	n.s.
	(0.00011)	(0.00008)	(0.00001)	(0.00007)			
cov(gender FE, RMSE)	0.0016***	0.00084***	0.0015***	0.00080***	0.321	0.281	13%
	(0.00008)	(0.00004)	(0.00008)	(0.00004)			
cov(mig. st. FE, class FE)	0.000069	0.000049	0.000018	-0.00000091	0.003	0.000	n.s.
	(0.00008)	(0.00007)	(0.00008)	(0.00007)			
cov(mig. st. FE, RMSE)	-0.00039***	-0.00016***	-0.00038***	-0.00015***	-0.114	-0.070	39%
	(0.00006)	(0.00004)	(0.00006)	(0.00004)			
cov(class FE, RMSE)	0.00094***	0.00096***	0.00091***	0.00093***	0.086	0.129	-50%
	(0.00016)	(0.00010)	(0.00016)	(0.00010)			

Note: We show General Structural Equation Models estimation of the residual variance/covariance of our inequality measures. The first column provides estimates for a null model (with no control variables), which are thus similar to the uncorrected variances and covariances shown in Table 1 and A4. The second column provides estimates for the full model found in Table A5. In columns 3 and 4, we replicate the exercise with the *"halves"* correction technique. As the halves correction does not enable us to estimate the four equations at the same time, we implement several separate models combining each two equations. Columns 5 and 6 display the corrected correlation coefficients. The last column displays the contribution of the model to the reduction in the correlation of inequality measures.
We weight establishments by the number of employees and, we additionally reweight the sample so that all years have equal weights.
RMSE: Root Mean Square of Errors, used as an indicator of intra-categorical inequality.
Cluster robust standard errors in parenthesis. *** p<0.001, ** p<0.01, * p<0.05

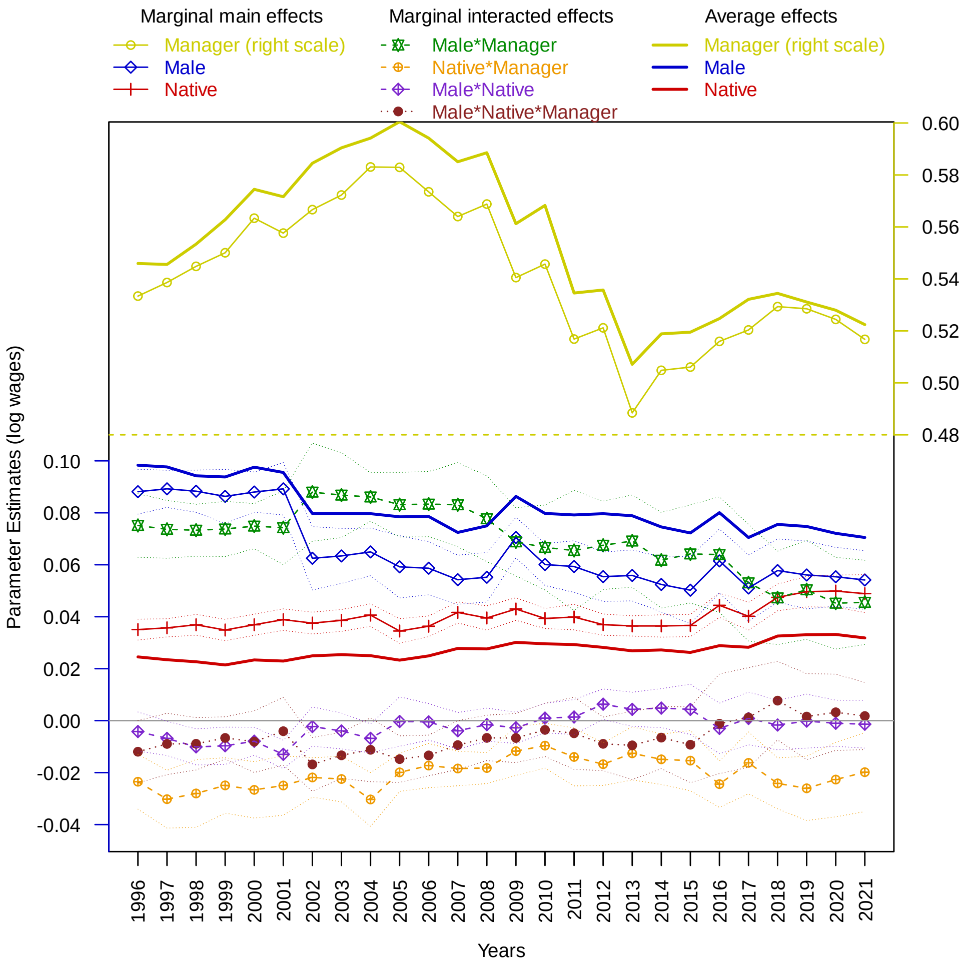
Table A7. Alternative estimates of Table 2 correlations following different first-stages

	No first-stage	Minimal first-stage	Full first-stage
vcov correction method:			
cor(gender FE, mig_stat FE)	-0.060	-0.113	-0.216
	[-0.069, -0.051]	[-0.122, -0.104]	[-0.226, -0.206]
cor(gender FE, class FE)	0.306	0.050	-0.004
	[0.298, 0.314]	[0.042, 0.058]	[-0.012, 0.005]
cor(mig_stat FE, class FE)	0.026	-0.008	0.006
	[0.017, 0.034]	[-0.017, -0.000]	[-0.003, 0.015]
halves correction method:			
cor(gender FE, RMSE)	0.073	0.274	0.327
	[0.064, 0.082]	[0.266, 0.283]	[0.318, 0.336]
cor(mig_stat FE, RMSE)	0.012	-0.121	-0.109
	[0.002, 0.023]	[-0.131, -0.110]	[-0.122, -0.096]
cor(class FE, RMSE)	-0.031	0.055	0.086
	[-0.040, -0.023]	[0.046, 0.063]	[0.076, 0.095]

Note: We re-estimate the correlation structure shown in Table 2 with alternative specifications of the first-stage regression. In column 1, we skip the first stage and estimate the second stage directly, using the log gross earnings as the dependent variable for workplace regressions. In column 2, we use a minimal first-stage specification that controls for quadratic functions of age, number of hours and days worked and a dummy variable for part-time. In the last column, we replicate the estimates from Table 2. No first-stage controls inflate the correlation of gender and migrant gaps with class and reduce that with intra-categorical inequality. However, correlations based on very minimal and obvious first-stage controls for work involvement yield similar qualitative results to correlations based on the full first-stage.
95% confidence intervals in square brackets were calculated with bootstrapping (4000 samples).
RMSE: Root Mean Square of Errors, used as an indicator of intra-categorical inequality.

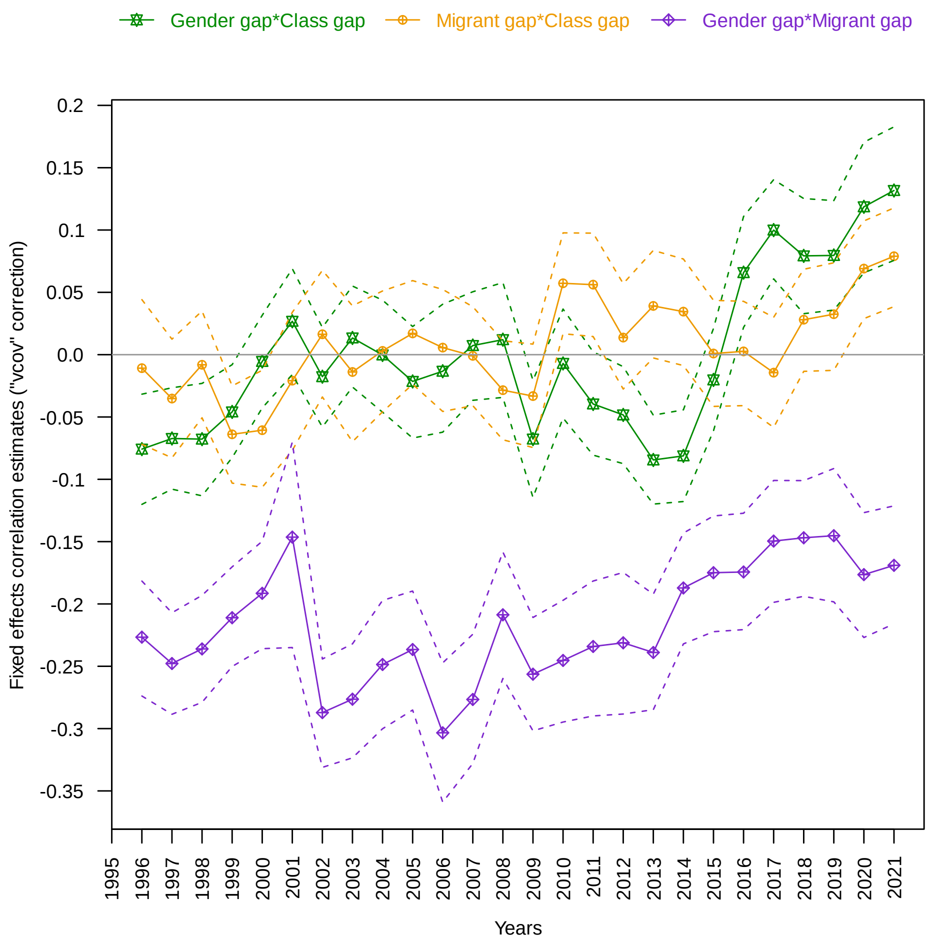
2. Supplementary figures

Figure A1. Marginal and average effects



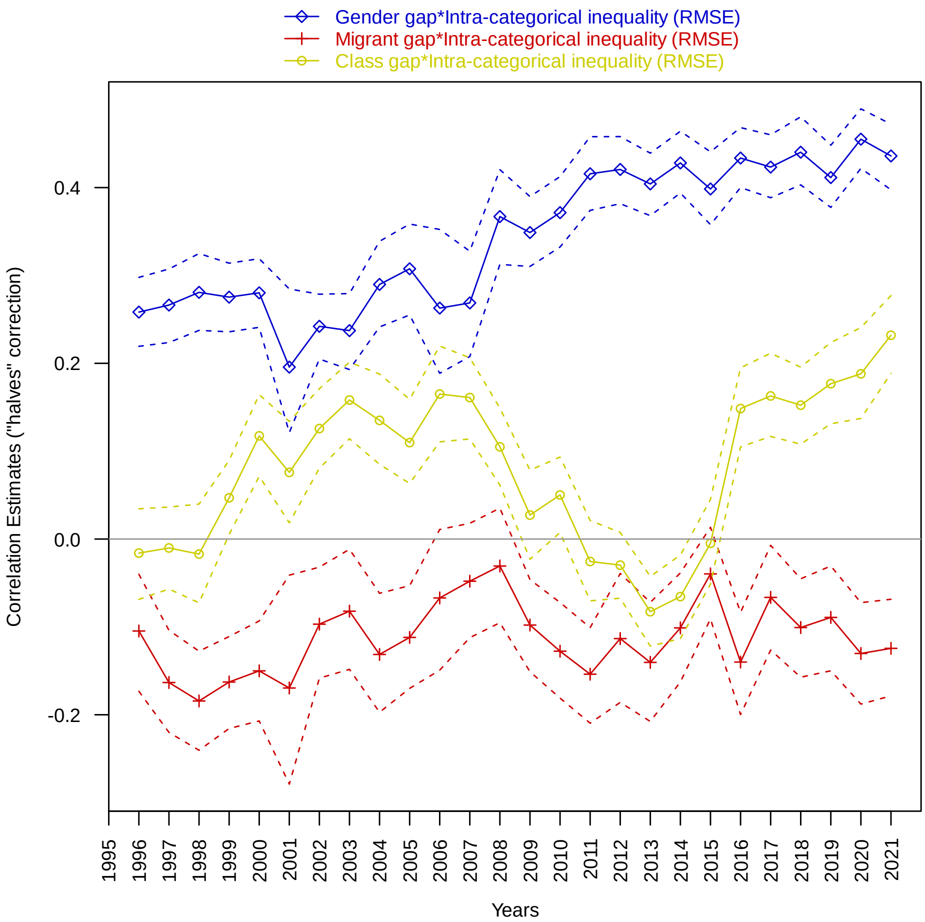
Note: Gender, migration status and class marginal main effects and their interactions on log earnings are calculated according to equation 1 using a yearly log earnings dependent variable controlling for age, number of hours worked, and 2-digit occupation categories (interacted with age) and introducing a workplace fixed effect. The 95% confidence intervals are based on cluster-robust standard errors, clustered at the regional level (i.e., French *départements*). Because we control for detailed occupation, class marginal main effects cannot be estimated directly and are post estimated (cf. Appendix 3.2). Therefore, we could not compute confidence intervals. Nevertheless, given the size of the effect, we have no doubt that it is highly significant. We also added the average effects (thick lines) which are similarly post-estimated.

Figure A2. Yearly correlation of categorical wage gaps



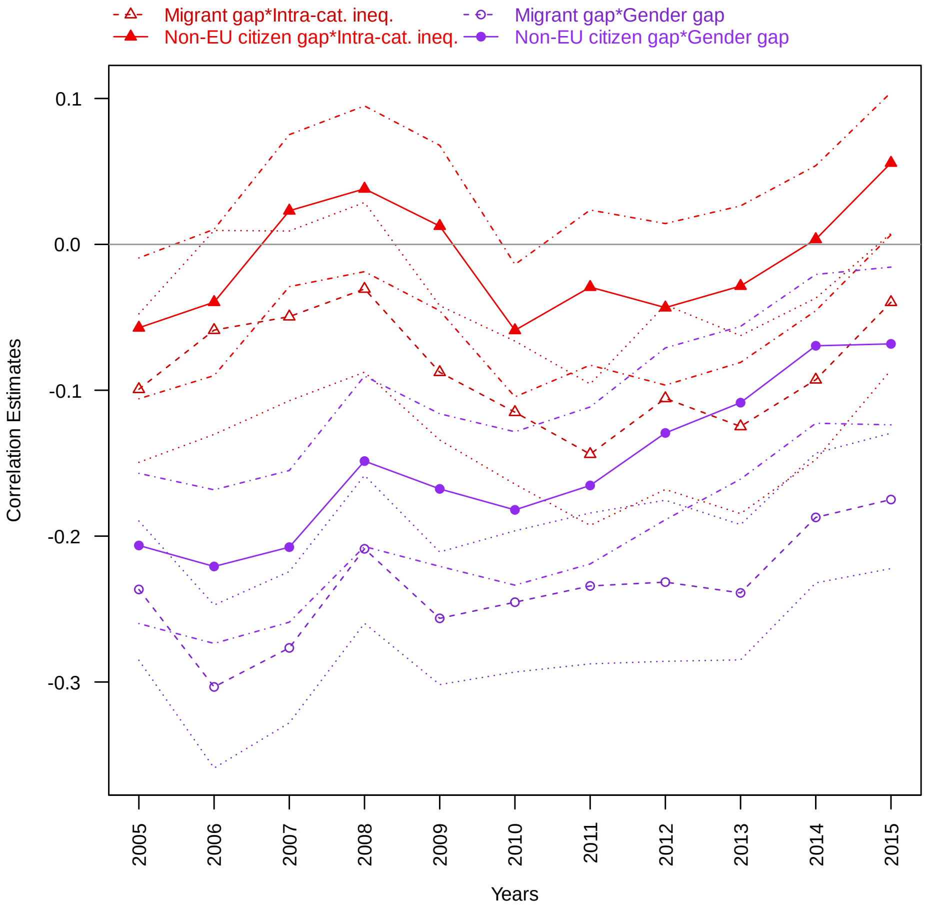
Note: The estimated values are shown in a solid line and the 95% confidence interval (estimated with a bootstrap) in dashed lines.

Figure A3. Yearly correlation of categorical wage gaps and intra-categorical inequality



Note: The estimated values are shown in a solid line and the 95% confidence interval (estimated with a bootstrap) in dashed lines.

Figure A4. Correlations comparing all migrant gaps and non-EU citizen gaps.



3. Mathematical appendix

3.1 Covariance of simple measures of workplace wage inequality leads to artefactual biases

Covariance of two categorical wage gaps

Suppose we estimate the first-order gender and migrant wage gaps in each workplace j (e.g., *mean male log wages – mean female log wages* and *mean native log wages – mean migrant log wages*) and that we compute their covariance.

$$C = cov_j(gender\_gap_j, migrant\_gap_j) \qquad (1)$$

The workplace j gender and migrant gaps correspond respectively to the parameters g_j^{gdr} and g_j^{mig} estimated in two separate workplace j regressions: $log(w_{ij})=a_{j1} + g_j^{gdr}.gdr_{ij} + u_{ij1}$, and $log(w_{ij})=a_{j2}+ g_j^{mig}.mig_{ij} + u_{ij2}$.

$$C = cov_j(g_j^{gdr}, g_j^{mig}) \qquad (2)$$

The classic analysis of the omitted variable bias in regressions shows that if the gender and migration status dummy variables are correlated, the first-order gender (respectively migration status) gap g_j^{gdr} is biased and captures part of the migration status (respectively gender) gap g_j^{mig} (Woolridge, 2013). This leads to an artefactual correlation between g_j^{gdr} and g_j^{mig}, the gender and migration status parameters.

Let us explore this bias by considering that the "true" gender and migrant gaps can be estimated with the following "true" workplace j regression:

$$log(wage_{ij})=a_j + \beta_j^{gdr}.gdr_{ij} + \beta_j^{mig}.mig_{ij} + u_{ij} \qquad (3)$$

We know from econometric textbooks (Woolridge, 2013, p. 90) that in this structure:

$$g_j^{gdr} = \beta_j^{gdr}+ \beta_j^{mig}\, cov_i(gdr_{ij}, mig_{ij})/V_i(gdr_{ij}) \qquad (4)$$

and

$$g_j^{mig} = \beta_j^{mig} + \beta_j^{gdr}\, cov_i(gdr_{ij}, mig_{ij})/V_i(mig_{ij}) \qquad (5)$$

Hence,

$$C = cov_j(\ [\beta_j^{gdr}+ \beta_j^{mig}\, cov_i(gdr_{ij}, mig_{ij})/V_i(gdr_{ij})],\ [\beta_j^{mig} + \beta_j^{gdr}\, cov_i(gdr_j, mig_{ij})/V_i(mig_{ij})]\ ) \qquad (6)$$

$$C = cov_j(\beta_j^{gdr}, \beta_j^{mig}) + D \qquad (7)$$

where $D= cov_j(\beta_j^{mig}, \beta_j^{mig}. [cov_i(gdr_{ij}, mig_{ij})/V_i(gdr_{ij})]_j) +$

$$cov_j(\beta_j^{gdr}, \beta_j^{gdr} . [cov_i(gdr_{ij}, mig_{ij})/V_i(mig_{ij})]_j) +$$

$$cov_j(\beta_j^{mig}.[cov_i(gdr_{ij}, mig_{ij})/V_i(gdr_{ij})]_j, \beta_j^{gdr} .[cov_i(gdr_{ij}, mig_{ij})/V_i(mig_{ij})]_j) \qquad (8)$$

C estimates our estimand, the empirical covariance of true parameters $cov_j(\beta_j^{gdr}, \beta_j^{mig})$, with a bias D. The structure of this bias can be quite complex. To get a sense of its structure, consider that $cov_i(gdr_{ij}, mig_{ij})$ is independent from β_j^{gdr} and β_j^{mig}. We can rewrite D as follows[1]:

$$D= V_j(\beta_j^{mig}) .E_j(cov_i(gdr_{ij}, mig_{ij})/V_i(gdr_{ij})) +$$

$$V(\beta_j^{gdr}) . E_j(cov_i(gdr_{ij}, mig_{ij})/V_i(mig_{ij})) +$$

$$cov_j(\beta_j^{mig}.[cov_i(gdr_{ij}, mig_{ij})/V_i(gdr_{ij})]_j, \beta_j^{gdr} .[cov_i(gdr_{ij}, mig_{ij})/V_i(mig_{ij})]_j) \qquad (9)$$

In this case, we see that the bias will disappear only if the covariance of gender and migration status in each workplaces $cov_i(gdr_{ij}, mig_{ij})$ equals 0. If $cov_i(gdr_{ij}, mig_{ij}) \neq 0$, then C will estimate our estimand $cov_j(\beta_j^{gdr}, \beta_j^{mig})$ with a positive bias when gender and migration status are positively correlated and with a negative bias when gender and migration status are negatively correlated.

To get a better sense of this bias in this last case, we will explore different configurations with simulations.

1 As a result of properties of covariance, cov(x.y, z)= E(x.y.z)-E(x.y).E(z). If x is independent from y and z, we can rewrite:
cov(x.y, z)=E(x)E(y.z)-E(x).E(y).E(z)=E(x).cov(y,z).

Table A8. The bias in the correlation first-order categorical gaps. A simulation exercise

Simulation parameters		Estimated parameters			
		Naive estimates		Two-stage estimates	
Cor(gdr,mig)	*Cov(β^{gdr}, β^{mig}) & Cor(β^{gdr}, β^{mig})*	*Cov(g^{gdr}, g^{mig})*	*Cor(g^{gdr}, g^{mig})*	*Cov(b^{gdr}, b^{mig})*	*Cor(b^{gdr}, b^{mig})*
0.2	0	0.41	0.38	-0.01	-0.01
		(0.02)	(0.01)	(0.01)	(0.01)
0	0	0	0	0	0
		(0.01)	(0.01)	(0.01)	(0.01)
-0.2	0	-0.4	-0.37	0.01	0.01
		(0.01)	(0.01)	(0.01)	(0.01)
0.2	0.5	0.93	0.73	0.5	0.48
		(0.02)	(0.01)	(0.01)	(0.01)
0	0.5	0.5	0.48	0.5	0.48
		(0.02)	(0.02)	(0.01)	(0.01)
-0.2	0.5	0.12	0.14	0.5	0.49
		(0.01)	(0.01)	(0.01)	(0.01)
0.2	-0.5	-0.11	-0.12	-0.5	-0.48
		(0.01)	(0.02)	(0.02)	(0.01)
0	-0.5	-0.5	-0.49	-0.5	-0.49
		(0.01)	(0.01)	(0.01)	(0.01)
-0.2	-0.5	-0.93	-0.73	-0.49	-0.48
		(0.02)	(0.01)	(0.01)	(0.01)

Note: For each simulation, we randomly generate 20 databases of 200,000 observations composed of workplaces with on average 500 employees (sd=22). We create workplace wage variable according to the equation $log(wage_{ij}) = \beta_j^{gdr}.gdr_{ij} + \beta_j^{mig}.mig_{ij} + a_j + u_{ij}$ where parameters and variables follow the correlation structure listed in the two first columns. We then estimate the "naive" covariance and correlation of gaps as in equation 2. We further estimate these moments with the (uncorrected) two-stage approach adopted in this article. We display the average of the 20 simulations and the standard deviation in parentheses.

The simulation exercise in Table S1.1 shows that the correlation of first order categorical gaps is highly positively (respectively negatively) biased when gender and migration status are positively (respectively negatively) correlated. Conversely, when we use the two-stage approach, our estimates are very close to the theoretical parameters used to generate the wage structure.

Covariance of a categorical wage gap with a variance-base inequality measure

Suppose that in each workplace *j* we estimate the covariance between the first-order workplace gender wage gap and the workplace variance of log wages (as a simple indicator of dispersion which will ease calculations below).

$$C = cov_j(\, gender\_gap_j \,,\, V_i(log(w_{ij}))\,) \qquad (10)$$

The *gender gap$_j$* can be rewritten as the parameter g_j^{gdr} of a simple OLS regression within workplace *j* where the gender variable is a male dummy variable:

$$log(wage_{ij}) = a_j + g_j^{gdr}.\, gdr_{ij} + u_{ij} \qquad (11)$$

We can therefore rewrite C as follows:

$$C = cov_i(\, g_j^{gdr},\ V_i(a_j + g_j^{gdr}.\ gdr_{ij} + u_{ij})\,) \tag{12}$$

The workplace variance can be decomposed into the following elements:

$$V_i(a_j + g_j^{gdr}.gdr_{ij} + u_{ij}) = (g_j^{gdr})^2 V_i(gdr_{ij}) + V_i(u_{ij}) + 2\, g_j^{gdr}.\ cov_i(gdr_{ij},\, u_{ij}) \tag{13}$$

We know from the OLS regression properties that $2\, g_j^{gdr}.\ cov_i(gdr_{ij},\, u_{ij})=0$. $V_i(gdr_{ij})$ is the variance of the proportion of males in the workplace j. Thus, $V_i(gdr_{ij}) = p_j^{gdr}(1- p_j^{gdr})$.

Based on these properties, we can rewrite C:

$$C = cov_j(\, g_j^{gdr},\ V_i(u_{ij})\,) + cov_j(\, g_j^{gdr},\ [(g_j^{gdr})^2 \times p_j^{gdr}(1- p_j^{gdr})]\,) \tag{14}$$

The second component of this simple decomposition clearly shows that C captures an artifactual correlation of the gender gap g_j^{gdr} with itself. The nature of the bias depends on the distribution of gender shares and gender gaps and their covariance.

In the case where gender shares and gender gaps are independent, we can rewrite C as follows:

$$C = cov_j(g_j^{gdr},\ V_i(u_{ij})) + cov_j(g_j^{gdr},\ (g_j^{gdr})^2) \times E_j(p_j^{gdr}(1- p_j^{gdr})) \tag{15}$$

Let us examine further the case of independence between gender gaps and gender shares. If the average gender gap is also zero-centred (e.g., $E(g_j^{gdr})=0$), which is extremely unlikely, the covariance of gender gaps with squared gender gaps would be zero ($cov_j(g_j^{gdr},\ (g_j^{gdr})^2)=0$), and there would be no artefactual bias. However, it is very likely that gender gaps are mostly positive and only reverse in a few workplaces. Gender gaps are therefore positively correlated with their square and this can only artifactually inflate C.

To further give a sense of the artifact, we will explore different configurations with simulations, where we modify the gender gap mean (from -2 to 2) and its correlation with residual log wage.

Table A9. The bias in the correlation of first-order categorical gap with the standard deviation of workplaces' log wages. A simulation exercise

Simulation parameters			Estimated parameters	
			Naïve	Two-stage
Mean(β_j^{gdr})	*Mean(β_j^{gdr}<0)*	*Cor(β_j^{gdr} , $sd_i(u)_j$)*	*Cor(g^{gdr}, $sd_i(ln(w))$)*	*Cor(b_j^{gdr} , $RMSE_j$)*
2	2%	0	0.25	0.01
			(0.04)	(0.04)
0.5	31%	0	0.07	-0.01
			(0.03)	(0.03)
0	50%	0	-0.01	-0.01
			(0.03)	(0.02)
-2	98%	0	-0.24	-0.01
			(0.04)	(0.04)
2	2%	0.5	0.65	0.49
			(0.02)	(0.02)
0	52%	0.5	0.46	0.49
			(0.02)	(0.02)
-2	97%	0.5	0.29	0.5
			(0.04)	(0.02)
2	3%	-0.5	-0.28	-0.49
			(0.02)	(0.02)
0	48%	-0.5	-0.46	-0.49
			(0.02)	(0.02)
-2	98%	-0.5	-0.66	-0.49
			(0.02)	(0.02)

Note: For each simulation, we randomly generate 20 databases of 200,000 observations composed of workplaces with on average 500 employees (sd=22). We create workplace wage variable according to the equation $log(wage_{ij}) = \beta_j^{gdr}.gdr_{ij} + a_j + u_{ij}$ where parameters and variables follow the correlation structure listed in the three first columns. In all simulations, the variance of the gender gap β_j^{gdr} is set to 1 and the average proportion of gdr_{ij} to 0.5. We then estimate the "naïve" correlation between the workplace gender gap and the standard deviation of log wages. We further estimate this link with the (uncorrected) two-stage approach adopted in this article. We display the average of the 20 simulations and the standard deviation in parentheses.

The simulation exercise in Table S1.2 shows that the correlation between the first order gender gap and the standard deviation of wage is highly positively (respectively negatively) biased when the expectation of gender gap is positive (respectively negative). To put it differently, when the gender gap only reverts in 2% of workplaces, the naive correlation generates a strong positive bias. In the hypothetical situation where there's an equal share of workplaces with a positive and negative gender gap, this bias tends to disappear. Conversely, when we use the two-stage approach, our estimates are very close to the theoretical parameters used to generate the wage structure.

3.2. First-stage regression and indirect estimation of the class effect

In the first stage, we estimate a workplace fixed effect regression on annual log-earnings:

$$\begin{aligned} log(earnings_{ij}) = X_{ij}.a + a^{gdr}.\ gender_{ij} + a^{mig}.\ mig\_stat_{ij} + a^{gdr.cla}.\ gender_{ij} \times class_{ij} \\ + a^{mig.cla}.\ mig\_stat_{ij} \times class_{ij} + a^{gdr.mig}.\ gender_{ij} \times mig\_stat_{ij} \\ + a^{gdr.mig.cla}.\ gender_{ij} \times mig\_stat_{ij} \times class_{ij} + a_j + w_{ij} \end{aligned} \quad (16)$$

Estimation of the individual class effect

Using this equation, we first want to estimate individual-level categorical inequality through class, gender and migration main effects within workplaces. While this is straightforward for the gender a^{gdr} and migration status gaps a^{mig}, we cannot estimate the class main effect directly because, as shown in equation (15), we control for 2-digit occupational codes in X_{ij} and these occupational codes are nested within our class measure of inequality. We therefore estimate a manager main effect in a post-estimation step by computing a weighted difference between all the occupational estimates nested in the manager/professional category and those nested in the worker category:

$$a^{cla} = \frac{\sum_k \left[p^{mgr_k} \left(\alpha^{mgr_k} + \alpha^{age.mgr_k}.\overline{age} + \alpha^{age^2.mgr_k}.\overline{age}^2 \right) \right]}{\sum_k p^{mgr_k}} - \frac{\sum_k \left[p^{wkr_k} \left(\alpha^{wkr_k} + \alpha^{age.wkr_k}.\overline{age} + \alpha^{age^2.wkr_k}.\overline{age}^2 \right) \right]}{\sum_k p^{wkr_k}} \quad (17)$$

where α^{mgr_k}, α^{wkr_k}, p^{mgr_k} and p^{wkr_k} are the estimated parameters (in the α vector of parameters of equation 16) and the weight of occupation k within the "managers and professionals" and "workers" categories respectively, and $\overline{\text{age}}$ is the average age in the full sample.

3.3. Biases in the variance and covariance of fixed effects

In a second stage, we estimate J regressions in each workplaces j on the residual log earnings w_{ij} coming from equation (16):

$$w_{ij} = b_j^0 + b_j^{gdr} \times gender_{ij} + b_j^{mig} \times mig\_stat_{ij} + b_j^{cla} \times class_{ij} + e_{ij} \qquad \textit{for each } j \quad (18)$$

with w_{ij}: residual log earnings; $gender_{ij}$: male dummy; mig_stat_{ij}: native dummy; $class_{ij}$: upper class dummy; e_{ij}: error; i: individual i; j: workplace j.

Our estimands are respectively the empirical covariances and variances of the true workplace categorical fixed effects. Using gender and migration status as example, we aim to estimate $cov_j(\beta_j^{gdr}, \beta_j^{mig})$, $V_j(\beta_j^{gdr})$ and $V_j(\beta_j^{mig})$. However, in finite samples, the empirical estimators b_j^{gdr} and b_j^{mig} obtained in equation 18 estimate the true β_j^{gdr} gender and β_j^{mig} migrant gaps with errors ε_j^{gdr} and ε_j^{mig} (thus, $b_j^{gdr} = \beta_j^{gdr} + \varepsilon_j^{gdr}$ and $b_j^{mig} = \beta_j^{mig} + \varepsilon_j^{mig}$). We analyse below the biases when computing the covariance and variances of empirical categorical fixed effects.

Covariance bias

We first analyse the covariance[2] of the gender and migration status empirical fixed effects:

$$cov_j(b_j^{gdr}, b_j^{mig}) = (\Sigma_j(b_j^{gdr} - b_.^{gdr})(b_j^{mig} - b_.^{mig}))/J \qquad (19)$$

Since w_{ij} is the residual earnings already conditional on gender and migration status, we can simplify and assume that the average earnings gaps are null ($b_.^{gdr} \approx 0$ and $b_.^{mig} \approx 0$).

We can thus rewrite:

$$\begin{aligned} cov_j(b_j^{gdr}, b_j^{mig}) &= (\Sigma_j(\beta_j^{gdr} + \varepsilon_j^{gdr})(\beta_j^{mig} + \varepsilon_j^{mig}))/J \\ &= cov_j(\beta_j^{gdr}, \beta_j^{mig}) + (\Sigma_j(\beta_j^{gdr}.\varepsilon_j^{mig} + \beta_j^{mig}.\varepsilon_j^{gdr} + \varepsilon_j^{gdr} . \varepsilon_j^{mig}))/J \\ &= cov_j(\beta_j^{gdr}, \beta_j^{mig}) + (\Sigma_j(S_j))/J \qquad (20) \end{aligned}$$

Let us examine the expectation of the element S_j for workplace j of this covariance:

$$\begin{aligned} E_j[S_j] &= E_j[\beta_j^{gdr}. \varepsilon_j^{mig} + \beta_j^{mig}.\varepsilon_j^{gdr} + \varepsilon_j^{gdr}.\varepsilon_j^{mig}] \\ &= E_j[\beta_j^{gdr}.\varepsilon_j^{mig}] + E_j[\beta_j^{mig}.\varepsilon_j^{gdr}] + E_j[\varepsilon_j^{gdr} . \varepsilon_j^{mig}] \qquad (21) \end{aligned}$$

In the classical OLS regression, the vector ε_j^k of errors in parameters k estimations follows the following law:

$$\varepsilon_j^k \sim N(0, \sigma_j^2(X_j'X_j)^{-1}), \qquad (22)$$

with σ_j^2 the variance of the residual e_j.

[2] To simplify the writing of variance, covariance and correlation equations, we neglect the fact that each workplace is weighted by the number of workers. Introducing these weights do not change the result of the decomposition.

$\sigma_j^2(X_j'X_j)^{-1}$ is the variance-covariance matrix of the parameters estimated in the earnings regression for workplace *j*.

By definition parameter errors ε_j^{mig} and ε_j^{gdr} are independent from β_j^{gdr} and β_j^{mig}. Thus,

$$E_j[\beta_j^{gdr}\ \varepsilon_j^{mig}] = E_j[\beta_j^{gdr}].E_j[\varepsilon_j^{mig}] = 0 \tag{23}$$

and similarly,

$$E_j[\beta_j^{mig}\ \varepsilon_j^{gdr}\ ] = 0 \tag{24}$$

As a result of equation 22, $\varepsilon_j^{gdr}\ \varepsilon_j^{mig} = m_j^{gdr\times mig} + \eta_j$ where $m_j^{gdr\times mig}$ is the $gdr \times mig$ element of the matrix $\sigma_j^2(X_j'X_j)^{-1}$ and η_j is random 0-centred error. Thus:

$$E_j[\varepsilon_j^{gdr}\ \varepsilon_j^{mig}] = E_j[m_j^{gdr\times mig} + \eta_j\ ]= E_j[m_j^{gdr\times mig}] \tag{25}$$

Hence,

$$cov_j(b_j^{gdr}, b_j^{mig}) \rightarrow cov_j(\beta_j^{gdr}, \beta_j^{mig}) + \Sigma_j(m_j^{gdr\times mig})/J \tag{26}$$

The covariance of empirical fixed effects is therefore a biased estimator of our estimand, the covariance of the true fixed effects $cov_j(\beta_j^{gdr}, \beta_j^{mig})$. Note that $m_j^{gdr\times mig}$ depends on $X'X$. When there are only two independent variables such as gender and migration status, the inverted covariance will depend approximately on $-a\sigma^2.cor(gender, mig\_stat)$ (with a>0)). The more these independent variables are tied positively (resp. negatively) the more $m_j^{gdr\times mig}$ will be biased negatively (resp. positively). Similarly, the smaller the variance of the residual, the smaller $m_j^{gdr\times mig}$ and therefore the bias. With three variables, the structure of the bias can become more complex.

Variance bias

For reasons similar to those discussed above, the variance of categorical fixed effects (for example gender) is biased.

$$V_j(b_j^{gdr}) = (\Sigma_j(b_j^{gdr} - b_.^{gdr})^2)/J \tag{27}$$

And with ($b_.^{gdr} \approx 0$), we can rewrite:

$$\begin{aligned} V_j\ (b_j^{gdr}) &= (\Sigma_j(\beta_j^{gdr} + \varepsilon_j^{gdr})^2)/J \\ &= V_j\ (\beta_j^{gdr}) + V_j\ (\varepsilon_j^{gdr}) + 2.cov_j(\varepsilon_j^{gdr}, \beta_j^{gdr}) \end{aligned} \tag{28}$$

Since the parameter measurement error in ε_j^{gdr} is independent of β_j^{gdr}, then $cov_j(\varepsilon_j^{gdr}, \beta_j^{gdr}) \rightarrow 0$.

The variance of empirical fixed effects is therefore a biased estimator of our estimand, the variance of the true fixed effects $V_j(\beta_j^{gdr})$:

$$V(b_j^{gdr}) \rightarrow V_j(\beta_j^{gdr}) + \Sigma_j(m_j^{gdr\times gdr})/J \qquad (29)$$

3.4. Correction of the biases

Correction of the covariance bias

We thus have two strategies for estimating $cov_j(\beta_j^{gdr}, \beta_j^{mig})$

1/ The first one (hereafter called *"vcov")* is to estimate

$$cov_j^{vcov}(\beta_j^{gdr}, \beta_j^{mig}) \approx cov_j(b_j^{gdr}, b_j^{mig}) - (\Sigma_j m_j^{gdr\times mig})/J \qquad (30)$$

2/ The second (hereafter called *"halves")* consists of randomly dividing each workplace j in two halves, estimating separate regressions on the two randomized halves, and calculating cross-halves covariance:

$$cov_j(b_{j1}^{gdr}, b_{j2}^{mig}) = cov_j(\beta_j^{gdr}, \beta_j^{mig}) + \Sigma_j(\beta_{j1}^{gdr}.\varepsilon_{j2}^{mig} + \beta_{j2}^{mig}.\varepsilon_{j1}^{gdr} + \varepsilon_{j1}^{gdr} . \varepsilon_{j2}^{mig})/J$$

$$cov_j(b_{j1}^{gdr}, b_{j2}^{mig}) \rightarrow cov_j(\beta_j^{gdr}, \beta_j^{mig}) + \Sigma_j(\varepsilon_{j1}^{gdr} . \varepsilon_{j2}^{mig})/J \qquad (31)$$

Since ε_{j1}^{gdr} and ε_{j2}^{mig} are parameter errors measured on two independent groups, they follow two independent probability laws. Therefore, $E[\varepsilon_{j1}^{gdr}. \varepsilon_{j2}^{mig}]=0$.

$$cov_j(b_{j1}^{gdr}, b_{j2}^{mig}) \rightarrow cov_j(\beta_j^{gdr}, \beta_j^{mig}) \qquad (32)$$

One way to combine the two cross-halves covariance is to compute the corrected covariance as follows:

$$cov_j^{halves}(\beta_j^{gdr}, \beta_j^{mig}) \approx cov_j\ [(b_{j1}^{gdr},b_{j2}^{gdr}),(\ b_{j2}^{mig}, b_{j1}^{mig})] \qquad (33)$$

Correction of the covariance of categorical fixed effects with root mean square of error

We also use the root mean square error $RMSE_j$ (where $RMSE_j = [\ \Sigma_i(e_{ij}^2)/n_j]^{0.5}$) as an indicator of workplace j intra-categorical inequality and estimate its covariance with our estimates of categorical inequality. Since the measurement error of the $RMSE$ statistics is not estimated the standard outputs of regression models, we cannot use the *vcov* method to correct for the covariance and the variance. Therefore, we use the randomized halves method to obtain the same corrected estimates as above.

Hence:

$$cov_j^{halves}(\beta^{gdr}, RMSE) = cov_j\ [(b_{j1}^{gdr},b_{j2}^{gdr}),(RMSE_{j2}, RMSE_{j1})] \qquad (34)$$

Correction of the variance bias

As above, averaging the diagonal element in the variance-covariance matrices of parameters in each workplace j regression allows us to estimate $V(\varepsilon^{gdr})$.

Hence:

$$V_j^{vcov}(\beta_j^{gdr}) \approx V_j(b_j^{gdr}) - (\Sigma_j\, m_j^{gdr\times gdr})/J \qquad (35)$$

Alternatively, we can also calculate the cross-covariance of parameters estimated on randomized halves:

$$cov_j(b_{j1}^{gdr},\, b_{j2}^{gdr}) = cov_j(\beta_j^{gdr}, \beta_j^{gdr}) + \Sigma_j(\beta_{j1}^{gdr}.\varepsilon_{j2}^{gdr} + \beta_{j2}^{gdr}.\varepsilon_{j1}^{gdr} + \varepsilon_{j1}^{gdr}.\varepsilon_{j2}^{gdr})/J$$

$$cov_j(b_{j1}^{gdr},\, b_{j2}^{gdr}) \rightarrow V_j(\beta_j^{gdr}) + \Sigma_j\,(\varepsilon_{j1}^{gdr}.\varepsilon_{j2}^{gdr})/J \qquad (36)$$

Since ε_{j1}^{gdr} and ε_{j2}^{gdr} are parameter errors measured on two independent groups, they follow two independent probability laws. Therefore $E[\varepsilon_{j1}^{gdr}\,.\,\varepsilon_{j2}^{gdr}]=0$ and $cov(b_{j1}^{gdr},\, b_{j2}^{gdr}) \rightarrow V_j(\beta_j^{gdr})$.

Hence, our second method:

$$V_j^{halves}(\beta_j^{gdr}) \approx cov_j[(b_{j1}^{gdr},\, b_{j2}^{gdr}),(b_{j2}^{gdr},\, b_{j1}^{gdr})] \qquad (37)$$

Correction of the correlation bias

We can thus compute the correct correlation coefficients as follows:

$$cor_j^{vcov}(\beta_j^{gdr}, \beta_j^{mig}) \approx cov_j^{vcov}(\beta_j^{gdr}, \beta_j^{mig}) / \,[V_j^{vcov}(\beta_j^{gdr})\,.\,V_j^{vcov}(\beta_j^{mig})]^{0.5} \qquad (38)$$

$$cor_j^{halves}(\beta_j^{gdr}, \beta_j^{mig}) \approx cov_j^{halves}(\beta_j^{gdr}, \beta_j^{mig}) / \,[V_j^{halves}(\beta_j^{gdr})\,.\,V_j^{halves}(\beta_j^{mig})]^{0.5} \qquad (39)$$

3.5. Simulation of the biases and correction

To further explore the bias and the quality of our corrections, we rely on simulations. In our baseline simulation, we simulate 20 random datasets consisting of 100,000 workers in 2,000 workplaces (an average of 50 workers per workplace) where each workplace j has its own gender β_j^{gdr} and migration status β_j^{mig} fixed effects. We set the covariance and the correlation between β_j^{gdr} and β_j^{mig} to -0.5 and the underlying correlation of *gdr* and *mig* to 0.2. Our average true correlation between β_j^{mig} and *RMSE* is 0.43.

In the following columns, we differ from the baseline simulation by one parameter. In column 2, we double the average number of workers per unit. In columns 3 and 4, we change the correlation between β_j^{gdr} and β_j^{mig} to a stronger absolute value (-0.8) and to zero. In columns 5 and 6, we change the correlation of *gdr* and *mig* to a weaker one (0.1) and a null one. In column 7 and 8, we decrease the correlation between RMSE and β_j^{mig}.

Table A10. Correction of variance, covariance and correlation. A simulation exercise

	1. Baseline	2. Larger sample	3. *cor* $(\beta^{gdr}, \beta^{mig})$ stronger	4. *cor* $(\beta^{gdr}, \beta^{mig})$ =0	5. *cor(gdr,mig)* weaker	6. *cor(gdr,mig)* =0	7. *cor(RMSE,* β^{mig}*)* weaker	8. *cor(RMSE,* β^{mig}*)* = 0
Theoretical parameters								
Average number of observations per unit	50	100	50	50	50	50	50	50
cov$(\beta^{gdr}, \beta^{mig})$ and *cor* $(\beta^{gdr}, \beta^{mig})$	-0.5	-0.5	-0.8	0	-0.5	-0.5	-0.5	-0.5
cor(gdr,mig)	0.2	0.2	0.2	0.2	0.1	0	0.2	0.2
$V(\beta^{gdr})$ and $V(\beta^{mig})$	1	1	1	1	1	1	1	1
V(RMSE)	2.86	2.75	2.85	2.85	2.86	2.84	2.84	2.84
	(0.06)	(0.07)	(0.06)	(0.05)	(0.04)	(0.05)	(0.04)	(0.04)
cov(RMSE, β^{mig}*)*	0.73	0.72	0.49	0.84	0.74	0.73	0.36	0
	(0.03)	(0.02)	(0.03)	(0.01)	(0.02)	(0.02)	(0.02)	(0.01)
cor(RMSE, β^{mig}*)*	0.43	0.44	0.29	0.5	0.44	0.43	0.22	0
	(0.01)	(0.01)	(0.02)	(0)	(0.01)	(0.01)	(0.01)	(0.01)
Simulation estimates								
$V(b^{gdr})$	1.35	1.18	1.35	1.34	1.34	1.33	1.33	1.34
	(0.06)	(0.04)	(0.06)	(0.04)	(0.05)	(0.05)	(0.03)	(0.05)
$V(b^{gdr})$ corrected (method: *vcov*)	1	1.01	1	0.99	1	1	0.98	0.99
	(0.06)	(0.04)	(0.06)	(0.04)	(0.05)	(0.05)	(0.03)	(0.05)
$V(b^{gdr})$ corrected (method: *halves*)	1.01	1	1.01	1.01	1	1.01	0.99	1
	(0.08)	(0.04)	(0.07)	(0.05)	(0.07)	(0.05)	(0.04)	(0.06)
cov(b^{gdr}, b^{mig})	-0.56	-0.54	-0.88	-0.08	-0.54	-0.49	-0.55	-0.58
	(0.03)	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)	(0.02)
cov(b^{gdr}, b^{mig}) corrected (*vcov*)	-0.49	-0.51	-0.81	-0.01	-0.5	-0.49	-0.48	-0.51
	(0.03)	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)	(0.02)
cov(b^{gdr}, b^{mig}) corrected (*halves*)	-0.49	-0.51	-0.81	-0.02	-0.5	-0.49	-0.48	-0.51
	(0.04)	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)
cor(b^{gdr}, b^{mig})	-0.42	-0.46	-0.65	-0.06	-0.4	-0.37	-0.41	-0.43
	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)
cor(b^{gdr}, b^{mig}) corrected (*vcov*)	-0.5	-0.51	-0.81	-0.01	-0.51	-0.49	-0.48	-0.51
	(0.03)	(0.03)	(0.03)	(0.04)	(0.04)	(0.03)	(0.03)	(0.02)
cor(b^{gdr}, b^{mig}) corrected (*halves*)	-0.49	-0.51	-0.8	-0.02	-0.5	-0.49	-0.48	-0.51
	(0.04)	(0.03)	(0.04)	(0.04)	(0.04)	(0.05)	(0.04)	(0.02)
V(RMSE)	2.69	2.67	2.68	2.68	2.69	2.67	2.67	2.67
	(0.05)	(0.07)	(0.06)	(0.06)	(0.03)	(0.05)	(0.04)	(0.04)
V(RMSE) corrected (*halves*)	2.61	2.63	2.6	2.61	2.61	2.6	2.61	2.6
	(0.06)	(0.07)	(0.06)	(0.06)	(0.03)	(0.06)	(0.04)	(0.05)
cov(RMSE, b^{mig}*)*	0.69	0.71	0.47	0.83	0.7	0.72	0.37	0.03
	(0.06)	(0.08)	(0.09)	(0.06)	(0.1)	(0.09)	(0.05)	(0.07)
cov(RMSE, b^{mig} *)* corrected (*halves*)	0.69	0.71	0.47	0.83	0.7	0.72	0.37	0.03
	(0.06)	(0.08)	(0.09)	(0.07)	(0.1)	(0.09)	(0.05)	(0.07)
cov(RMSE, b^{mig} *)*	0.37	0.4	0.25	0.44	0.37	0.38	0.2	0.01
	(0.03)	(0.04)	(0.05)	(0.03)	(0.04)	(0.04)	(0.02)	(0.03)
cov(RMSE, b^{mig} *)* corrected (*halves*)	0.43	0.44	0.29	0.51	0.43	0.44	0.23	0.02
	(0.03)	(0.04)	(0.06)	(0.03)	(0.05)	(0.04)	(0.03)	(0.04)

Note: For each simulation, we randomly generate 20 databases of 100,000 observations composed of workplaces with on average 50 (or 100 in column 2) employees (sd=7 or 10). We create workplace wage variable according to the equation $log(w_{ij}) = \beta_j^{gdr}.gdr_{ij} + \beta_j^{mig}.mig_{ij} + a_j + u_{ij}$ where parameters and variables follow the correlation structure listed in the first part of the table. We then estimate these parameters with our two-stage uncorrected approach and implement the *vcov* and *halves* corrections. We display the average parameters estimated on 20 different datasets followed by their standard deviation in parentheses.

The simulations show the following results:

- The variance and covariance of the fixed effects are substantially biased.

- The bias in the variance and covariance of the fixed effects is weaker when the unit size is large (column 2).
- The bias in the covariance decreases as the absolute value of the correlation between *gdr* and *mig* decreases (column 5). When uncorrelated, the bias is close to 0 (column 6).
- The two methods of correction allow us to correctly estimate the variance, the covariance and the correlation coefficient. They give very similar results.
- The *vcov* correction method is relatively more powerful than the randomized *halves* one (as shown by the smaller standard deviations) which justifies using the former when available.
- The variance of the RMSE and its covariance with β^{mig} does not seem to be substantially biased. However, their correlation coefficient is biased mainly due to the bias in the variance of the β^{mig} fixed effects.

4. References

Woolridge, Jeffrey M. (2009) 2013. *Introductory Econometrics: A Modern Approach*. Mason: South Western, Cengange Learning.