

A Service of



Leibniz-Informationszentrum Wirtschaft Leibniz Information Centre for Economics

Elass, Kenza; García-Peñalosa, Cecilia; Schluter, Christian

Working Paper Gender Gaps in the Urban Wage Premium

CESifo Working Paper, No. 11374

Provided in Cooperation with: Ifo Institute – Leibniz Institute for Economic Research at the University of Munich

Suggested Citation: Elass, Kenza; García-Peñalosa, Cecilia; Schluter, Christian (2024) : Gender Gaps in the Urban Wage Premium, CESifo Working Paper, No. 11374, CESifo GmbH, Munich

This Version is available at: https://hdl.handle.net/10419/305616

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



WWW.ECONSTOR.EU



Gender Gaps in the Urban Wage Premium

Kenza Elass, Cecilia García-Peñalosa, Christian Schluter



Impressum:

CESifo Working Papers ISSN 2364-1428 (electronic version) Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute Poschingerstr. 5, 81679 Munich, Germany Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de Editor: Clemens Fuest https://www.cesifo.org/en/wp An electronic version of the paper may be downloaded • from the SSRN website: www.SSRN.com

- from the RePEc website: <u>www.RePEc.org</u>
- from the CESifo website: <u>https://www.cesifo.org/en/wp</u>

Gender Gaps in the Urban Wage Premium

Abstract

We examine the economic geography of gender wage gaps to understand the role that location plays in gender earning differences. Using panelised administrative data for the universe of French workers, our findings indicate that women benefit relatively more from density than men, with an urban wage premium (return to urban density) 48% higher than for men. We identify a number of factors that explain this gap, with a large share being explained by the structure of the local labour market, notably, the extent of occupational segregation. Another important factor is commuting patterns, while childcare availability plays only a moderate role.

JEL-Codes: J310, J160, R100, R230, R120.

Keywords: gender wage gap, agglomeration economies, urban wage premium.

Kenza Elass Bocconi University Milan / Italy kenza.elass@unibocconi.it Cecilia García-Peñalosa Aix-Marseille University / France cecilia.garcia-penalosa@univ-amu.fr

Christian Schluter Aix-Marseille University / France christian.schluter@univ-amu.fr

September 2024

Appendix here: <u>https://drive.google.com/file/d/1XK1jeGmbXEah4418Fp0DoNe3CYC6LRDc/view?usp=sharing</u>

We are grateful to David Margolis, Dominique Meurs, Elie Vidal-Naquet, Eric Maurin, Pedro Mira, Eva Moreno Galbis, Fabien Petit, Giulia Giupponi, Nina Roussille, Roland Rathelot, and Sebastien Roux, as well as to audiences at the Gender Gaps Conference 2023 (GRAPE), King's College Junior Research day, ADRES 2024 Conference, COSME 2024 Workshop, 3rd Workshop on Public Policies, UEA Conference 2024, and IAAE Conference 2024, and seminars at Aix-Marseille School of Economics, Bocconi University, and the Paris School of Economics for useful comments and suggestions. We acknowledge financial support from the French government under the "France 2030" investment plan managed by the French National Research Agency Grant ANR-17-EURE-0020, the Excellence Initiative of Aix-Marseille University - A*MIDEX, and the ANR-18-CE41-0003-01 grant.

1 Introduction

Although wage differences between men and women have narrowed considerably over the past decades, they have not vanished (see e.g. Goldin, 2014). To explain this persistence in the gender wage gap, discrimination, child penalties, and gender norms are often invoked (Kleven et al., 2019a,b; Bertrand et al., 2015). While much has been learned about these factors at large, their spatial dimension has received little attention to date. The aim of this paper is thus to examine the economic geography of gender earnings gaps using data for the universe of French workers.

Geography should matter. It is well known in urban economics that wages tend to be higher in denser local labour markets, and this so-called urban wage premium is explained in part by agglomeration economies (see, for example, Combes and Gobillon, 2015, and more recently Dauth et al., 2022 and Card et al., 2024). By contrast, little is known about whether such wage gains differ between men and women and whether characteristics of the local labour market other than agglomeration economies can explain the differences. In this paper, we bring these two perspectives together and ask whether the urban wage premium, and hence the gender earnings gap, varies systematically across local labour markets, and if so, why?

Our empirical strategy starts by following standard practice in urban economics, as we estimate multi-stage AKM-style multi-way fixed effect models following Combes et al. (2008) and recent advances in the agglomeration economies literature, such as Card et al. (2024). In the first stage, wage regressions are estimated using worker and location fixed effects in order to correct for the potential selection of workers across locations. In the second stage, the measure of the urban wage premium is obtained by regressing the fitted location effects on the logarithm of location density. Adapting this procedure to our context, we estimate the first stage empirical model separately by gender and then regress the *differences* in the gender-specific fitted location effects on location density to obtain our measure of the gender gap in the urban wage premium. The credibility of this research design depends on the quality of the fitted location effects. As these are identified in the data by movers, sufficient movement across local labour markets is required. Consequently, the demands on the data are considerable. We meet this data challenge by using French administrative panelised data for the universe of workers for the period 2005-2019, yielding about 237 million observations across the entire French metropolitan territory. Moreover, our panel enables the construction of worker employment history, so we can control in the first stage for where work experience has been acquired and thus for dynamic agglomeration gains, as done by De la Roca and Puga (2016). We find that the French urban premium of female workers is 48% larger than that of men, implying that, doubling the population density of the commuting zone is associated with an approximate increase in earnings of 2.4% for women, compared to 1.6% for men.

Having obtained a significant gender gap, we then show in the second part of our analysis that this gap can be explained as we augment the second-stage regression by relevant descriptors of the local labour market. The density of urban areas can affect both labour demand as well as the availability of services, which in turn may impact the labour supply of men and women differently. In particular, we consider aspects such as childcare provision and commuting times, as well as labour market features including the diversity of occupations available and the extent of labour market concentration. The latter could affect men and women differently; for example, greater concentration could weaken the position of women when bargaining their pay, thus resulting in a greater gender earnings gap. Given the importance attributed to social norms in the recent literature on gender gaps, we also consider a proxy for gender norms to ask if more traditional norms in low-density locations are behind the observed gap. Under more traditional norms that view women as secondary earners, firms could be less likely to promote women, women themselves may be less inclined to negotiate their earnings upwards, or they may self-select into occupations seen as feminine, all of which would result in lower female earnings in less dense locations.

Augmenting the model with local characteristics, the significant baseline

estimate of the gap in the coefficients of log density of 0.0076 is fully explained by the local determinants. We find that several forces are at work. About onethird (36%) of the initial difference in the urban wage premium between men and women is due to dynamics returns to experience when we augment our first-stage with controls for where and when the experience has been acquired. Then, we assess the effect of local features. The decomposition developed by Gelbach (2016) is especially useful for our specification as it allows to disentangle the extent to which the effect of population density on the urban premium gender gap is mediated by other explanatory variables.¹

Our findings indicate that the gap remaining once we control for experience history is mainly explained by the structure of the local labour market. A key feature is the extent of occupational segregation, and our results indicate that in areas with low population density, the distinct sorting of men and women into separate occupations is more pronounced. This, in turn, is correlated with a larger wage gap, in line with an extensive literature showing that feminised occupations command lower wages (Blau and Kahn, 2003). Additionally, local disparities in road commuting time contribute considerably to the gender gap in agglomeration economies. This result is consistent with recent evidence highlighting the role of commuting in creating gender wage gaps (Le Barbanchon et al., 2021; Liu and Su, 2024), and since commuting time is greater in less dense areas this factor explains part of the effect of density. Labour market concentration and the diversity of industries are also important. In contrast, differences across locations in the availability of childcare play a significant but economically moderate role. Historical gender norms are significant too but their effect is largely wiped out once we control for occupational segregation.

Our approach enables us to make several contributions to the established literature. First and foremost, we provide a detailed examination of the economic geography of the gender gap, using exhaustive population data for French workers at the appropriate spatial scale (local labour markets). To date, researchers usually take a non-spatial perspective to explain the persis-

¹Notably, this decomposition allows to overcome the problem of sequence dependence, i.e. the order in which additional covariates enter the regression which can affect the results.

tence of the gender wage gap. For instance, a recent literature has pointed to the effect of social norms, leading to differences in chosen occupations or hours of work. Goldin (2014), for example, emphasizes the importance of workplace norms concerning hours of work which affect the degree of flexibility that women may have, while Bertrand et al. (2015) focus on households' desire to conform to the "breadwinner norm", which results in women taking labour supply decisions that ensure their husbands earn more than they do. We show, first, that these factors follow a systematic spatial pattern, as the gender gaps in the densest local labour markets are considerably smaller than in the least dense ones. Second, we find that some but not all aspects related to gender norms matter. A large share of the gender gap in the urban wage premium is explained by the higher occupational segregation observed in the least dense locations. In contrast, availability of childcare facilities plays a moderate role while the length of commuting is an important factor. The urban premium gap is also explained by other features of the local labour market, as women's earnings are negatively affected by a lower diversity of industries and greater labour market concentration, an aspect that has been largely ignored in the debate on gender gaps.

Second, we contribute to the literature on the urban wage gap by highlighting how it differs across genders, thus identifying a gender gap in the urban wage premium. The typical focus in this literature is on male workers and little attention has been given to the heterogeneity of this phenomenon between different demographic groups.² This difficulty likely arises from the challenges in tracking women's career trajectories locally and over time, especially following career breaks, which are more common among women than men due to motherhood. Our paper builds on the existing literature on agglomeration economies by computing the agglomeration gains for women as well as for men. Our finding that women experience a higher urban wage premium implies that focusing exclusively on men results in an underestimation of the

 $^{^{2}}$ A notable exception is Ananat et al. (2018) who have explored the urban wage premium gap between black and white male workers in the US and show that higher density results in a smaller increase in wages for blacks than for whites in the same metropolitan area.

overall urban wage premium. Moreover, the contrast between our results and those on black workers (in the US) indicates that the mechanisms through which density affects the wages of minority workers are highly specific to the minority group under consideration (Ananat et al., 2018).

Third, although a small literature on gender and location exists, we advance our understanding of the reasons why the urban premium is higher for women, reconciling the recent evidence in both the urban economics and the literature on gender gaps in earnings. This empirical literature has examined how city size affects the gender wage gap with mixed evidence. Early work found a negative correlation between city size and the gender wage gap in the US and in the UK, showing that the larger the population of the standard metropolitan area, the smaller the gender wage gap is; see Oaxaca (1973), Frank (1978), and Ofek and Merrill (1997). More recently and focusing on different countries, Phimister (2005), Hirsch et al. (2013), Duranton (2016), Nisic (2017), Almeida et al. (2022) and D'Costa (2024) also found a positive correlation between city size or the rural/urban divide and women's wages relative to men's. Meekes and Hassink (2023) and De la Roca and Puga (2016), in turn, do not find a significant difference. Data limitations imply that these analyses do not correct for the sorting of workers across locations.³

Our paper's contribution to this literature is twofold. On the one hand, we provide the first analysis of gender urban wage premia that corrects for the bias associated with sorting. This avoids concerns due to unobserved worker ability and endogeneity, and our use of alternative methodologies shows that the gender gap in is robust to the estimation approach employed. On the other, this literature has barely examined the mechanisms that may explain

³The paper closest to our work is D'Costa (2024), who use the specification proposed by Glaeser and Maré (2001) to estimate the gender gap in the urban wage premium. Her approach differs from ours in that it employs a one-step approach with individual fixed effects, whereas we rely on the recent literature that has widely documented the benefits of the 2-steps approach to estimate the agglomeration economies, starting with Combes et al. (2008). Also, while D'Costa (2024) uses a binary rural/urban divide, we follow the literature by using a continuous measure of population density. This provides finer granularity, avoids arbitrary cutoffs, and better captures local externalities.

why women benefit more than men from denser locations.⁴ Our paper explores the reasons why the urban premium is higher for women, and identifies three key aspects not explored by the agglomeration economics literature: the role of returns to experience, the impact of commuting costs, and the fact that denser locations exhibit less occupational segregation, potentially due to more traditional gender norms.

The rest of the paper is structured as follows. Section 2 describes our data, while section 3 describes the empirical strategy following the methodology developed by Combes et al. (2008). We then present our key results (Section 3.2) and explore possible mechanisms (Section 3.3). Section 4 concludes. The Online Appendix contains substantial supplementary material covering data descriptives, details on estimation methods, additional empirical evidence, and robustness analyses.

2 Administrative population data

We use panelised administrative data for the *universe* of workers in France over the period 2005 to 2019 working in private or public firms, but excluding civil servants.⁵ Known as BTS (*Base Tous Salariés*), these data are provided by the French National Institute for Statistics (INSEE) as a series of exhaustive yearly files, derived from tax returns submitted by firms on their employees' payrolls. To construct the pseudo-panel, we follow the variable-based matching approach described in Babet et al. (2023).⁶ Our spatial unit, reported in the

⁴Starting with Frank (1978), the literature has simply argued that women exhibit a lower spatial elasticity of labour which in turn leads to lower wages.

⁵More specifically, the data includes workers in EPIC, 'Établissement public à caractère industriel et commercial en France', which include firms such as France Télécom, Air France, EDF-GDF, or SNCF, but exclude civil servants such as school teachers or those holding public office.

⁶In the BTS, each worker is assigned a pseudonymous identification code that changes annually. However, most job-related variables are available for both the current year and the preceding year. The construction of the exhaustive pseudo-panel is then done with the data overlap based on matched attributes (e.g. establishment ID, gender, hours worked, job duration in days, job start and end dates, work and residence municipalities, earnings, and age) between the information of the current year and the preceding one.

BTS, is the commuting zone (CZ - *zones d'emploi*), based on the place of residence, and defined by the INSEE for the year 2010 on the basis of the daily commute of workers, leading to a partition of mainland France into 297 commuting zones.⁷

While the exhaustive BTS data has not been available to researchers until recently, its sibling, the 1/12 sample known as the DADS (*Déclarations Annuelles des Données Sociales*), has been used for many years.⁸ Our empirical strategy, explained in detail in Section 3.2 below, requires sufficient movement of subgroups of workers across commuting zones and that the data be spatially representative even for the smallest locations. Since this might not be the case in the 1/12 sample, we work with the universe of French private sector employees.⁹

2.1 Variable construction

The BTS dataset allows us to construct complete labour market histories including wages, time worked, and type of job contract. 'Wages' refer to the principal occupation of the worker and are computed as monthly deflated gross earnings in 2019 prices. Monthly earnings is our main variable of interest.

Our spatial unit is the commuting zone, but we follow standard practice in urban economics and refer to 'commuting zones', 'cities' and 'locations' interchangeably. The 'density' of a commuting zone is defined as the number of inhabitants, using the INSEE's historical data of communal populations, divided by the surface area in kilometres, for each year between 2005 and 2019. The geo-coding of place of residence and place of work allows us to measure where work experience was acquired in the past, as well as to measure the current commuting distance, two aspects that we will consider in our analysis

⁷In Appendix D.3, we check the robustness of our results to an alternative spatial unit, the urban areas (*Zones Urbaines*).

 $^{^{8}}$ See, for instance, Abowd et al. (1999) Abowd et al. (1999); Combes et al. (2012); Schmutz and Sidibé (2018); Le Barbanchon et al. (2021) or Bilal (2023).

⁹Table D.2 in Appendix D.1 reproduces our main estimates using the DADS and shows that using these data results in an underestimation of the female city size earning premium and an overestimation of the male one.

following recent work that highlights the importance of these variables (De la Roca and Puga, 2016; Le Barbanchon et al., 2021).

Occupations are coded according to the French classification PCS (at the 2-digit level) with 23 categories while industry variables are coded according to the French industry classification (NAF) with 17 categories. Appendix A provides a detailed enumeration of categories.

Additional commuting zone descriptors are constructed from both the BTS and other administrative data sources. The former is used to construct measures of occupational segregation, average road commute, and labour market concentration. Appendix A describes in detail all additional variables used, which include, for example, a measure of labour market tightness constructed using data from the French unemployment agency, or the availability of childcare services, obtained from administrative data on firms and their establishments' location and main activity in France (SIRENE data).¹⁰

2.2 Sample selection

We consider prime-age workers in the private sector in mainland France.¹¹ More specifically, our age restriction is 25-55, excluding young and teenage workers since we want to focus on stable career paths. As our empirical strategy exploits the mobility of workers, the well-known 'getting-started' phenomenon would add confounding noise for the youngest workers since most teenage workers are on minimum wages with little wage change following job-to-job transitions. For the same reasons, we exclude older workers whose relocation decisions might be influenced by (pre-)retirement considerations and for whom there may be considerable selection into employment, for example if those on lower wages retire early.

¹⁰Alternative specifications using measures of local childcare services built with data obtained from the State Services for Family (*'Caisse d'Allocation Familiales'*), available only from 2013 onwards, provide similar results.

¹¹The public sector is included in the BTS only from 2010 onwards. Our robustness tests in Appendix D.3, Table D.5, confirm that our results do not change when we restrict the observation window but include these workers.

Our analysis focuses on the period 2005 to 2019, given data availability,¹² and these selection rules yield a population that comprises 106,412,914 employment spells for women and 131,336,044 for men. Overall we observe 40,380,984 workers.

2.3 Descriptive Statistics

Figure 1 maps the spatial variation across commuting zones of the gender gap in average earnings for the year 2019. Since the urban wage premium relates location effects to local population density, we depict in the second panel the densities of commuting zones. These two maps suggest that the higher the density of the population, the smaller the gender earnings gap is, especially for the 4 largest areas in France (i.e. Paris and its surroundings, Marseille, Lyon and Lille). This visual impression of a negative correlation is confirmed more formally in Table A.1 in the Appendix, where we regress the local gender gap in mean monthly earnings on local density, and obtain a negative coefficient. The same negative correlation holds when we use alternative definitions of earnings or hourly wages.

Since the urban wage premium will be identified on workers who change commuting zones, we present selected summary statistics that compare movers and stayers. A map and transition matrix of movers' flows can be found in Appendix A.3. Over the period 2005 to 2019, our dataset contains 5,525,132 commuting zone changes, with 15% of men and 12% of women moving at least once. Of all the workers included in the sample, around 20% of female employment spells correspond to women who migrate at least once, while approximately 24% of male employment spells pertain to men who have moved at least once. A substantial share of such moves involves career changes, as approximately 29% of female movers have a different occupation after the move, the corresponding percentage being 31% for men.

¹²Although the data is available starting in 1994, we start only in 2005 to ensure we have sufficient prior years of data to accurately reconstruct the workers' dynamic experience histories to replicate the estimation of De la Roca and Puga (2016), as we will discuss below.



Figure 1: Local gender gaps in earnings and density in 2019

(a) Gender gaps in earnings

(b) Urban density

Notes: BTS data, 2019. This map plots the **Notes:** INSEE data, 2019. This map plots mean gross monthly deflated earnings gap by the demographic density (population within commuting zone. A darker area denotes a one square kilometre) by commuting zone. higher gender earnings gap and a lighter area A darker area denotes a higher density and denotes a lower gender earnings gap. a lighter area denotes a lower density.

Table 1: Summary statistics: Average values for movers and non-movers

		Men	Women	Gap
Monthly earnings	Non-movers	3,041	2,488	553 ***
In euros	Movers	$3,\!133$	2,694	439 ***
Part-time	Non-movers	0.10	0.34	-0.24 ***
Share of population	Movers	0.10	0.24	-0.14 ***
Experience in the	Non-movers	3.76	3.50	0.31 ***
establishment - In years	Movers	2.57	2.50	0.07 ***
Experience in the firm	Non-movers	4.44	4.10	0.34 ***
In years	Movers	3.30	3.22	0.09 ***
Short-term contract	Non-movers	0.18	0.17	0.01 ***
Share of population	Movers	0.21	0.20	0.01 ***

Note: BTS data (2005-2019). Table 1 displays the average characteristics (and gaps) of workers, measured separately for male and female workers, as well as for movers across different commuting zones and stayers within the same commuting zone during the period 2005-2019.

Table 1 presents summary statistics for our data. It shows that spatial

mobility tends to increase earnings. Women tend to earn less than men, but the earnings gender gap is lower for movers than for stayers $(553 \in \text{per month})$ for the former and $439 \in \text{for the latter}$, the resulting earnings gap being 16% and 22%, respectively). The incidence of part-time workers is considerably larger for women, but lower among female movers (24% vs. 34%), while about 10% of male workers are part-timers. By construction, movers have shorter job durations in the same establishment or firm than stayers, typically by around one year, while the incidence of non-permanent contracts is higher among movers. The respective durations are somewhat shorter for women than for men.

Table 2: Descriptive statistics of movers before and after the CZ migration

			All			25-35			35-45			45-55	
		Men	Women	Gap	Men	Women	Gap	Men	Women	Gap	\mathbf{Men}	Women	Gap
Log density	Just before moving	5.77	5.82	0.05	5.75	5.86	0.11	5.84	5.82	-0.02	5.72	5.64	-0.08
	Just after moving	5.65	5.65	0	5.68	5.74	0.06	5.63	5.57	-0.06	5.56	5.46	-0.10
Wage difference	Following the migration	4%	5%	1	5%	5%	0	3%	5%	2	3%	5%	2

Note: BTS data (2005-2019). Table 2 reports the average log population density in the commuting zones of origin and destination, for male and female movers, and the gender gap across age groups before and after moving to a different commuting zone. Additionally, Table 2 reports the wage differential following a move to a different commuting zone, separately measured for men and women and age groups.

Table 2 focuses on movers and confirms that movers tend to experience earnings gains. Following a move, the earnings of female workers increase by about 5% across all age groups, while younger male workers enjoy slightly higher relative gains than older male workers (5% v. 3%). Across all age and sex groups, movers tend to end in slightly less dense locations. Maps of movers' flows by gender and age groups are provided in Appendix A.3, showing that mobility patterns do not exhibit important variations throughout the life cycle.

Further descriptive statistics can be found in Appendices A and B. In Appendix A.4, we provide evidence that the negative selection bias occurring in less dense CZ is not greater for women than for men, at least as far as the educational characteristics of the non-employed are concerned. Figures B.7 and B.8 in Appendix B, map the workforce shares by municipality and gender, as well as the associated gender gap, showing an absence of substantial differences in the geographical distribution of men and women.

3 Gender gaps in the urban wage premium

3.1 Empirical strategy

3.1.1 Agglomeration economies

Recent work on the urban wage gap employs a two-step procedure, put forward by Combes et al. (2008). In the first step, a two-way fixed effects model for (log) wages of the AKM-type (Abowd et al., 1999) is estimated, which includes individual fixed effects and location effects. In the second stage, the prediction of the location effect is regressed on (the log of) local population density, the slope coefficient being the measure of the urban wage premium. A positive coefficient is usually interpreted as evidence of agglomeration gains, making similar workers more productive in denser local labour markets. In contrast, the positive correlation between worker fixed effects and location fixed effect is seen as evidence of sorting, capturing the fact that more productive workers tend to work in denser locations which, because of agglomeration gains, are more productive.

Specifically, the empirical two-stage model for log wages is given by

$$w_{ijt} = \beta x_{ijt} + \alpha_i + \nu_{jt} + \epsilon_{ijt}, \qquad (1)$$

$$\hat{\nu}_{jt} = \delta D_{jt} + \xi_{jt}, \qquad (2)$$

where w_{ijt} denotes the log monthly earning of worker *i* in commuting zone *j* at time *t*, α_i is the individual worker fixed effect, the vector x_{ijt} captures the worker's observed time-varying characteristics, D_{jt} is the log of density in CZ *j* at *t*, while ϵ_{ijt} and ξ_{jt} denote errors uncorrelated with the regressors. ν_{jt} is the potentially time-varying location (commuting zone) effect, and $\hat{\nu}_{jt}$ its prediction. In the second stage regression, the predicted location fixed effects are regressed on local density. This empirical model allows for systematic pat-

terns of mobility related to the fixed effects and characteristics ("exogenous" mobility), but rules out "endogenous" mobility driven by unobserved locationworker match-specific components.¹³ This is satisfied if the assignment process is $E(\nu_{it}^l|X,\epsilon) = \Pr\{J(i,t) = l|X,\epsilon\} = \Pr\{J(i,t) = l|X\}$, where the latter is a function of the fixed effects and characteristics $G_{il}(\alpha_i, \nu_{1t}, \nu_{2t}, \ldots, x_{ilt})$. This condition permits systematically different mobility patterns by e.g. worker productivity, characteristics, or location amenities, but rules out a correlation with a match-specific as part of ϵ . This latter scenario could arise from unobserved human capital accumulation, so that past residuals are predictive of future mobility. Our implementation limits this effect, since we estimate the first stage regression for 5 year windows, thus allowing α_i to vary, and we further control for dynamic agglomeration gains relating to where experience was acquired (as described in detail in Appendix D.2). Finally, we observe that in our adaption of the second stage estimation, we can even accommodate endogenous mobility provided it is invariant to sex, as we consider the difference $\hat{\nu}_{it}^F - \hat{\nu}_{it}^M$ and the distortion is then subtracted out. In Appendix E we consider an event-study design to provide further evidence of exogenous mobility. As is common in the literature, we do not include industry and occupation fixed effects.¹⁴

To analyse the gender gap in agglomeration returns, we estimate the first stage, equation (1), separately for women and men so as to obtain, for each commuting zone, two location effects, $\hat{\nu}_{jt}^F$ and $\hat{\nu}_{jt}^M$. This allows the effect of density to be gender specific, so that we have different values for δ^F and δ^M .

¹³More specifically, let J denote the assignment of worker i to location l, J(i,t) = l, ν_{it}^{l} the associated dummy variable, L the matrix collecting these, and ϵ the vector of all residuals. OLS requires $E(L'\epsilon|X) = 0$.

¹⁴Many analyses of urban premia do not include either industry or occupation fixed effects (Card et al., 2013, 2024; Dauth et al., 2022) while others control for industry (Combes et al., 2008; Abowd et al., 1999). Card et al. (2024) argue that such controls are not desirable. First, occupations tend to reflect the skills of the individual and are hence captured by the worker fixed effects. Second, when workers change CZs, they will have access to a different set of occupations and industries and the resulting wage changes should be interpreted as part of the location effect. Because of this argument, and given our interest in the effect of local labour market structure on the gender wage gap, we include neither control. However, in the Appendix D.3, Table D.6 we show that our core estimates are robust to the inclusion of both industry fixed effects (10 categories) and broad group of occupation (6 categories).

To capture the gender gap, equation (3) regresses the difference in the fitted gender-specific location effects across genders, $\hat{\nu}_{jt}^F - \hat{\nu}_{jt}^M$, on the logarithm of city density. That is,

$$\hat{\nu}_{jt}^F - \hat{\nu}_{jt}^M = \gamma D_{jt} + \tilde{\xi}_{jt}.$$
(3)

where $\gamma \equiv \delta^F - \delta^M$ and $\hat{\gamma}$ is our measure of the gender gap in the urban wage premium. A positive (negative) $\hat{\gamma}$ indicates that female workers gain relatively more (less) than male workers from agglomeration effects.¹⁵

Two caveats are in order. First, it is well known that the implementation of the model is challenging since the location effects are identified only by movers, i.e. workers changing locations, and that a low incidence of movement yields a so-called limited-mobility bias. Moreover, the data might not be spatially representative of small locations, resulting in further distortions. The use of panel data for the *universe* of workers in France for a sufficiently long period allows us to overcome these challenges since the data yield a large number of movers as discussed above.

Second, although the method we employ is now widely used when looking at agglomeration economies, Card et al. (2024) have recently raised concerns that models using individual fixed effects underestimate the causal effects of location. As an alternative, they propose to estimate, instead of equation (1), the classic AKM two-way fixed effects model (Abowd et al., 1999) that includes both individual and establishment fixed effects. They then compute the location wage premium associated with a CZ as a weighted average of the establishment effects in that CZ. This allows them to consider the unobserved heterogeneity in the premiums paid by different firms in the same CZ. We will use this approach as a robustness check and, as we will see below, find that although it results in larger agglomeration effects, as argued by Card et al. (2024), the increase is equivalent for men and women and consequently the estimated gender gap is unaffected.

¹⁵Equation (2) implies a linear regression model. In Appendix C, we replicate our main result using a spline regression approach, allowing for non-linearity.

3.1.2 Decomposing the effect of density

Equation (3) above yields an estimate of the effect of density on the gender wage gap. In order to understand what drives this effect, we include a set of covariates X_{jt} that are deemed potential explanations of why density matters for the gender wage gap, and try to understand the contribution of each of these factors to the overall effect.

Gelbach (2016) highlights that sequentially adding covariates to a baseline model and comparing the resulting coefficient estimates does not typically allow us to identify the contribution of different covariates as the decomposition of the overall effect depends on the order in which these are introduced. To overcome this problem, Gelbach (2016) propose a decomposition using the omitted variable bias formula that clearly separates the impact of each omitted variable on the change in the coefficient of the variable of interest, in our case, log density.

In particular, assume that the gender gap is explained by an expression of the form

$$\hat{\nu}_{jt}^F - \hat{\nu}_{jt}^M = \gamma_d D_{jt} + \mathbf{X}_{jt} \gamma_x + \epsilon_{jt}, \qquad (4)$$

where the vector γ_x are the coefficients on the vector of covariates \mathbf{X}_{jt} and γ_d captures the remaining, or direct, effect of density. Equation (3) then exhibits an omitted variable bias if population density is correlated with the variables in \mathbf{X}_{jt} .

Gelbach (2016) suggests carrying out a decomposition in which we can compute the contribution of variable x to the total impact of density as follows. First, estimate the full model defined in equation (4) to get the vector of $\hat{\gamma}_x$. Then, regress the variables in \mathbf{X}_{jt} on D_{jt} to obtain a vector of coefficients $\hat{\beta}_x$. Since the omitted variable bias due to the absence of variable x, b_x , can be expressed as $\hat{b}_x = \hat{\gamma}_x \hat{\beta}_x$ and given that the total bias is $\hat{b} = \sum_x \hat{b}_x$, then the estimate \hat{b}_x can be seen as a measure of the contribution of variable x to explaining the impact of density on the gender wage gap.

Our analysis will proceed in two steps. Section 3.2 reports estimates of the gender gap in the urban wage premium, as defined by γ , focusing on the second

stage of the estimation as defined by equation (3). In Section 3.3 we turn to the question of whether the observed gender gap in returns to density can be explained by observable city differences. To do so, we perform a decomposition of the effect of the various variables using the methodology just described.

3.2 Baseline estimates

The location effect is obtained from the first stage equation (1) as a prediction, estimated for three non-overlapping 5-year periods (2005-2009, 2010-2014, 2015-2019) by gender.¹⁶ As the predicted gender-specific location effect may be affected by which time-varying individual characteristics x_{ijt} are included, we estimate several specifications by progressively adding controls. First-stage regressors include experience in the firm, working time (distinguishing full-time and part-time workers), whether the workers have a short-term contract (indicating precariousness of the job), and the (log) commuting distance.¹⁷ The second stage regression is weighted by the 2005 population in the CZ to give greater weight to locations with a larger population.

¹⁶There are two reasons for doing this. First, the coefficients in the first regression may vary over time; second, given that we have 237 million observations, it is computationally difficult to produce estimates that pool all periods.

¹⁷To adhere closely to the specifications of the existing literature, we do not consider the potential duration of unemployment preceding the beginning of the contract. However, in Appendix D.3, we introduce an additional control in the first stage, incorporating the duration of unemployment (in days) before the employment contract.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Women	Men	Women	Men	Women	Men	Women	\mathbf{Men}
Log Density	0.0177^{***}	0.0147^{***}	0.0193^{***}	0.0141^{***}	0.0236^{***}	0.0160^{***}	0.0747^{***}	0.0532^{***}
	(0.0009)	(0.0007)	(0.0008)	(0.0007)	(0.0009)	(0.0007)	(0.0020)	(0.0013)
Observations	4,455	4,455	4,455	4,455	4,455	4,455	4,455	4,455
R-squared	0.7153	0.7315	0.7068	0.7047	0.7368	0.7131	0.8141	0.6664
			1st stage c	ontrols				
Workers fixed effects	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark		
Experience, contract type \mathscr{C}			/	/	/	/	/	/
working time			\checkmark	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark
Log commute					\checkmark	\checkmark	\checkmark	\checkmark

Table 3: Second stage estimation - Urban premium and location density

Note: ***p < 0.01, **p < 0.05, *p < 0.1. Coefficients are reported with robust standard errors in parenthesis. Second stage estimation of the city-size earnings premium. Controls include year dummies. All columns denote a first-stage regression of monthly deflated gross earning with both time-varying commuting zone fixed effects and individual fixed effects. Columns (1) and (2) without further controls, (3) and (4) controlling for experience in the firm and in the establishment, short-term contract & working-time, (5) and (6) controlling additionally for log commute, and (7) and (8) controlling for all these aforementioned variables but without workers fixed-effects. Appendix C.1 reports the first-stage coefficients.

Table 3 reports the sex-specific results of the second stage regression, δ^{g} , as defined by equation (2), with the various columns differing in the explanatory variables included in the first stage regression. Columns (1) and (2) include only worker fixed effects, and we sequentially include additional individual characteristics. The urban wage premium is positive for both genders, indicating that both benefit from agglomeration gains. For men, this premium is fairly stable across all four specifications, the range being 0.014-0.016. In contrast, for women the premium grows across the four specifications from 0.018 to 0.024, indicating the importance of part-time work and commuting time for female workers. For completeness, we have also reported in columns (7) and (8) the results for a first-stage specification that does not include individual worker fixed effects. The finding shows the resulting substantial distortion in the urban wage premium estimate, which arises from a wrong attribution of systematic worker differences to location differences, leading to an underestimation of the gender gap in agglomeration economies. According to our preferred specification (columns (5) and (6)), a doubling of the commuting zone's density is associated with an increase in earnings of 2.4% for women and 1.6% for men.¹⁸

	(1)	(2)	(3)	(4)	(5)	(6)
				IV	Card et al. (2024) 's	Dynamic exp.
				11	method	gains
Log Density	0.0030***	0.0052***	0.0076***	0.0078***	0.0076***	0.0049***
	(0.0003)	(0.0002)	(0.0003)	(0.0003)	(0.0010)	(0.0005)
Observations	4 455	4 455	4 455	4 455	891	4 455
R-squared	0.5156	0.6459	0.6895	0.6892	0.2306	0.4791
		1st s	tage contro	ols		
Workers fixed effects	\checkmark	\checkmark	~	\checkmark	\checkmark	\checkmark
Experience, contract type &		\checkmark	\checkmark	\checkmark	1	1
working time						
Log commute			V	~	V	V
Dynamic exp. gains						\checkmark

Table 4: The gender gap in the urban wage premium: Baseline results

Note: *** p < 0.01, ** p < 0.05, * p < 0.1. Coefficients are reported with robust standard errors in parenthesis. Estimates of γ , the gender gap in the urban wage premium, given by equation (3), based on estimation, separate by sex, of the first stage regression (1). The dependent variable in the first stage is monthly deflated gross earnings, and all specifications include worker fixed effects and time-varying commuting zone effects. Additional worker-level controls are in column (2) experience in the firm and in the establishment, short-term contract & working-time, column (3) further controls for log commute. The first stage regression work C.1. Column (4) uses historical data to instrument for density and column (5) the estimation method proposed by Card et al. (2024). Column (6) expands the first stage regression by including dynamic experience gains, which are described in detail in Appendix D.2. In all specifications, the second stage regression is weighted by 2005 local population size. Number of observations for columns (1), (2), (3), (4) and (6): 297 commuting zones × 15 years = 4,455. Number of observations for columns (5): 297 × 3 periods = 891.

We next compute the gender gap in the urban wage premium for each location, $\hat{\nu}_{jt}^F - \hat{\nu}_{jt}^M$, and regress it on urban density. The results are reported in Table 4. The first three columns differ in the explanatory variables included in the first stage regression, and indicate that the gender gap in the urban wage premium is always significant and positive, implying that female workers tend to benefit relatively more from agglomeration gains. Our preferred specification is the one that includes all individual-level controls and is reported in column (3). The wage gap is 0.0076, more than twice as large as the one

¹⁸These results are in line with the literature which reports recent estimates of the elasticity of about 0.02 to 0.05, irrespective of the gender Duranton and Puga (2014); Graham and Gibbons (2019). Estimates of this elasticity vary depending on the identification strategy, period, spatial scale and sample restrictions.

without controls, indicating the importance of composition effects as the male and female labour forces have different characteristics across locations.

Table 4 also considers the possibility that the second-stage estimates of the gender gap might be affected by an endogeneity bias arising from the density of different locations changing due to systematic worker location choices. To this end, we follow Combes et al. (2011) and use historical values of local employment density as an instrument for current density, in our case as of 1881. The data are obtained from historical municipal population data.¹⁹ The results are reported in column (4) and show that the slope coefficient estimate is virtually unchanged, suggesting no endogeneity bias, in line with the literature (Ciccone and Hall, 1996; Combes et al., 2011). Consequently, we use OLS in all subsequent estimations.

Column (5) reports the elasticity estimate, with the identification strategy developed by Card et al. (2024), as described in subsection 3.1.1. As argued by Card et al. (2024), using firm rather than location fixed effects results in a higher estimates of agglomeration economies, however, it does not impact the estimates of the gender gap in agglomeration economies.²⁰

Lastly, column (6) of Table 4 presents the coefficient obtained when we consider dynamic agglomeration gains in the first stage. De la Roca and Puga (2016) argue that agglomerations gains may be dynamic, as experience gained in large cities may increase earnings even once the worker leaves the city. The returns to experience should hence account not only for when, but also where it has been acquired. To this end, we augment the first stage equation with a set of variables which relate past and current employment locations in terms of location-specific accumulated experience. Estimates allowing for dynamics gains are reported and discussed in Appendix D.2. We find, for instance, that women exhibit a consistently higher value for their experience

¹⁹They are available on the INSEE website at this link.

²⁰The method proposed by Card et al. (2024) increases the agglomeration elasticity estimate from 0.016 to 0.022 for men and from 0.024 to 0.030 for women. The number of observations falls with this estimation strategy since it does not allow for time-varying local fixed effects. Card et al. (2024)'s strategy thus provides estimates for each of our periods (2005-2009, 2010-2014, 2015-2019) and 297 commuting zones.

acquired in Paris regardless of their current location, while men tend to benefit from past experience in Paris only if they stay there. The resulting gender gap in agglomeration effects is given in column (6) of Table 4. As expected, including dynamic gains captures part of the differences between men and women, resulting in a still highly significant, gender gap, that falls from 0.0076 to 0.0049.

A possible explanation for these results is differences in promotions. It has been shown that women have lower internal promotion rates than men, a fact partly attributable to them having smaller networks (Kunze and Miller, 2017). If career progression is more likely when there is the possibility for external promotions, women may benefit more from high-density locations where the opportunities for external promotions are greater. Such a mechanism could explain the greater returns to experience gained in Paris that women exhibit we identify.

Figure 2 plots the gender gap in the urban premium for the year 2019, that is $\hat{\nu}_{j,2019}^F - \hat{\nu}_{j,2019}^M$, against the (log) population density, using our preferred specification (column (3)). Note that in the least dense commuting zones the gender gap differs considerably across locations and is often negative. France is a highly centralized country. The capital, Paris, is a gravitational centre and Paris and the commuting zones in its surrounding region (Île-de-France) are the densest locations in France. The Figure shows that these locations are both the densest and those in which the gender gap in agglomeration returns is the highest, raising the question of whether they are driving our results. Table D.4 in Appendix D.3 reports a number of robustness exercises concerning this question. Dropping Paris from the regression reduces $\hat{\gamma}$ from 0.0076 to 0.0054, further dropping all other commuting zones in the wider Paris region leads to a further reduction to 0.0039, and dropping the largest 5% of commuting zones to a further fall to 0.002. However, the coefficients remain highly significant.

Further robustness tests are reported in Appendix D.3 (see Tables D.5, D.6 and D.4), where we confirm that the gender gap persists in all our specifications. Specifically, we consider different samples of workers: adding public sector workers and consequently restricting the analysis to 2010-19 has no ef-



Figure 2: The gender gap in the urban wage premium in 2019 by commuting zone

Notes: The depicted regression is given by equation (3), for the year 2019. The specification used is that corresponding to Table 4, column (3).

fect while including younger (20-24) and older workers (55-60) and restricting the sample to full-time workers both reduce the coefficient of interest though it remains significant. The last result confirms our priors as, given the importance of part-time jobs for female workers, we expected to find a lower gap when considering full-time workers only. We also segment our sample into various age brackets to show the presence of the urban wage premium gap across the entire lifespan. The results suggest that the latter amplifies as individuals progress through life, that is, the relative advantage obtained by women when working in denser locations is greater for older individuals. Lastly, Table D.6 considers additional specifications of the location effect, using alternative administrative data, including firm fixed effects, and changing the dependent variable to the hourly wage. The gender wage gap persists across all alternative specifications. For instance, the gender gap is 0.0056 when considering hourly wages as the dependent variable in the first stage.

Additionally, Combes et al. (2018) argue that the analysis of agglomeration economies should be linked to urban costs. Unfortunately, price indices at the CZ level are not available. As land and house prices are the main drivers of differential urban costs, these costs are unlikely to be different for men and women. Indeed, Figures B.7 and B.8, in the Appendix B illustrate a clustering of both men and women in the same municipalities within the commuting zone, as the difference in the spatial distribution of the workforce between men and women does not exceed 0.12 percentage points. As a result, the likelihood of encountering distinct urban costs for each gender is low.

3.3 Explaining the gender gap by observable city differences

Several mechanisms could explain the higher city-size earnings premium for women, and this section considers a number of possibilities. The density of urban areas can affect both labour demand and the availability of services, which in turn may impact the labour supply of men and women differently. In particular, certain characteristics of labour demand and of services that are more prevalent in larger urban areas may disproportionately benefit women, resulting in a higher urban wage premium for women compared to men. For expositional simplicity, our discussion of the mechanisms in Table 5 starts from the baseline specification for the gender-gap of the urban wage premium, $\hat{\gamma}$, and discussed in follows Table 4, which is now reported in column (1). In column (2) we control for dynamic experience gains in the first-stage regression, and then investigate the roles of several mechanisms. Appendix A.1 gives details of how we have constructed our regressors. Our discussion primarily focuses, for the sake of brevity, on the change in and decomposition of the estimate of $\hat{\gamma}$.

Columns (1) and (2) indicate that a substantial component of the difference in the urban premium across genders stems from dynamic experience gains, as when we control for these in the first-stage regression, the estimate of $\hat{\gamma}$ falls from 0.0076 to 0.0049. Columns (3) and (6) report regression coefficients when we add add several covariates that may explain why women benefit more than men from population density. Columns (4) and (7) present the results of the Gelbach (2016) decomposition, which breaks down how these different factors contribute to the change in the log-density coefficient. The underlined coefficients in columns (4) and (7), 0.0035 and 0.0062, respectively, report the difference between the unadjusted log-density coefficient in column (2), 0.0049, and the adjusted coefficients in columns (3) and (6), 0.0014 and -0.0013. The other coefficients in columns (4) and (7) quantify the contribution of each factor to the change in the coefficient on log-density, showing how much of the variation can be attributed to each observed characteristic. The percentages reported next to the coefficients–columns (5) and (8)–indicate the proportion of the coefficient change explained by each factor.

The covariates used capture various aspects related to childcare, gender norms, and local labour markets. The difference between columns (3)-(5) and columns (6)-(8) is that the latter include an additional explanatory variable, occupational segregation by gender. As we will see below, including this covariate has a major effect on both the coefficient of density and those on several other variables. It is not obvious whether or not we should think of occupational segregation as an explanatory variable to understand the differences in the effect of density, as in a sense, it implies explaining a gender gap (in earnings) by another gender gap (in the allocation of men and women to jobs). Hence, we start by discussing the results when occupational segregation is not included, and then add this variable.

The literature on gender wage gaps has emphasised the role of child penalties, and given its prominence, we start by adopting a spatial gendered laboursupply perspective and consider the availability of childcare facilities in the

	OLS Specification		OLS Specification	Gelbach De Expl	ecomposition ained	OLS Specification	Gelbach Decomposition Explained	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log Density (gap)	0.0076*** (0.0003)	0.0049^{***} (0.0005)	0.0014^{**} (0.0005)	$\frac{0.0035}{(0.0003)}^{***}$	<u>69%</u>	-0.0013^{***} (0.0004)	$\frac{0.0062}{(0.0004)}^{***}$	<u>127</u> %
Childcare Facilities Public			24.0333***	0.0010***	20%	8.9795*	0.0004*	8%
			(4.4012)	(0.0003)		(4.5913)	(0.0002)	
Childcare Facilities Private			-30.9973***	-0.0004***	-8%	-20.5492***	-0.0002**	-4%
1962 Fertility rate			(5.9296) 0.0037* (0.0023)	(0.0001) 0.0008^{***} (0.0001)	16%	(5.5320) -0.0049*** (0.0017)	(0.0001) 0.0001 (0.0001)	2%
Mean male commuting time			0.1316***	0.0029***	59%	0.1177***	0.0025***	51%
U			(0.0124)	(0.0002)		(0.0103)	(0.0002)	
Labour Market Tightness			0.0040***	0.0000	0%	0.0036***	-0.0001***	-2%
			(0.0003)	(0.0000)		(0.0003)	(0.0000)	
Labour Market Concentration			-0.0063***	-0.0023***	-47%	-0.0100***	-0.0021***	-43%
Diversity of industry			(0.0017) -0.0006*** (0.0000)	(0.0002) 0.0024^{***} (0.0002)	49%	(0.0018) -0.0006*** (0.0000)	(0.0002) 0.0022^{***} (0.0002)	45%
Diversity of occupation			-0.0017***	-0.0010***	-20%	-0.0003	-0.0016***	-33%
Occupational Segregation Duncan Index			(0.0002)	(0.0000)		-0.1453***	0.0050***	102%
						(0.0084)	(0.0004)	
				First st	tage controls			
Multiple variables on experience - Where it has been acquired and used		\checkmark		\checkmark	0		\checkmark	
Observations R-squared	$4,455 \\ 0.6895$	$4,455 \\ 0.4791$		$4,455 \\ 0.6006$			$4,455 \\ 0.6624$	

Table 5: Density earning premium gap estimation and Gelbach decomposition

Note: ***p < 0.01, **p < 0.05, *p < 0.1. Coefficients are reported with robust standard errors in parenthesis. Columns (1), (2), (3) and (6) display the second-stage estimation of the density earning premium gap (women-men), given by equation (4) with a different set of covariates. Columns (4) and (7) present the Gelbach (2016)'s decomposition results, and columns (5) and (8) the corresponding proportion of the log-density coefficient's change explained by the variable. Controls include year dummies. Controls in the first-stage regression of monthly deflated gross earnings as per Table 4. The measures of public and private childcare facilities are the number of facilities per capita. Labour Market Tightness is the ratio of job vacancies to job seekers. Labour Market Concentration is the HHI index for concentration of workers, with the set of firms hiring in occupations (PCS) at the 4-digit level. Diversity of occupations and industries denote the HHI index for occupations. Mean commuting male road time is computed from commute information in the BTS data and the METRICS application developed by INSEE. The 1962 fertility rate is defined as the ratio of births to fertile-age women.

CZs.²¹ Women are usually the main child-care provider, hence greater avail-

²¹The academic literature provides contradictory results on the impact of childcare facility availability or expansion on female wages.Hermes et al. (2024) and Lefebvre and Merrigan (2008) find a positive effect on female outcomes and earnings. On the inverse, Havnes and Mogstad (2011) analyses a large-scale expansion of subsidized child care in Norway and do not find a causal impact on maternal employment estimated as working hours and participation. However, they do not analyse the impact on wages although it could have improved the job matching. Similarly, Pora (2024) do not find an impact of the expansion of subsidized childcare facilities on parents' labour outcomes. Nevertheless, the analysis focuses solely on the expansion of public childcare facilities, whereas our study also includes the availability of private childcare facilities. Our specification differ from the existing literature by using rich administrative data on firms' activities and locations to consider both public and private childcare facilities. ability of child-care support would increase their possibilities for having a full-time job, for having a longer commute, or for more flexible working time, thus leading to a better job match and/or higher wages. Denser CZs could be characterized by a greater number of childcare facilities, which, moreover, are likely to be closer to the worker's residence than in less dense locations (see Appendix B, Figure B.10). We hence examine whether the advantage conferred to women by density is due to the greater availability of childcare facilities, both public and private. The Gelbach (2016) decomposition shows that public childcare facilities explain 20% of the narrowing log-density coefficient as working mothers in high-density locations benefit relatively more from greater childcare availability.²² We believe this to be a key empirical result.

We also include a measure of historical gender norms in order to capture potential variations in social attitudes between urban and rural areas. Less dense locations may be more traditional making workers living there more likely to adhere to gender norms. As a result, women may be perceived as secondary earners by households and firms, impacting earnings. For example, employers could be less likely to promote women, resulting in stagnant earnings, or women themselves may be less inclined to negotiate their wages upwards. Thus, weaker gender norms could be a reason why women's earnings are positively affected by density. As a proxy for gender norms, we use historical data on fertility, in particular the 1962 fertility rate, which is the earliest wave available in the French Census at the CZ level.²³ This proxy of gender norms explains 16% of the reduction in the log-density coefficient.

Commuting patterns may be important as differences in preferences over commuting time have been shown to be an important determinant of gender wage gaps in France by Le Barbanchon et al. (2021). In particular, they find that women are more willing than men to tradeoff wages for commuting

 $^{^{22}}$ The effect seems to stem mainly from public childcare facilities, with private ones barely affecting positively the coefficient of density, a result due to the fact that private childcare facilities are well less correlated with density (see Figure B.10).

 $^{^{23}}$ To capture gender norms, we also used measures of the historical share of women amongst university graduates and the historical female employment rate (both measured in 1968, the earliest date for which data are available at the suitable level) and obtained equivalent results to those for fertility.

time, and that the resulting employment choices can explain a considerable share of the gender wage gap. Moreover, Liu and Su (2024) show that gender gaps in commuting time are spatially different and depend on the geographical concentration of jobs. Hence, we compute a measure of the male commuting road time in each CZ, capturing the extent to which a CZ requires a long commute. Male commuting road time plays an important role in the gender gap in urban earning premium, as it explains 59% of the change in log-density coefficient.²⁴

Several features of the local labour market may be relevant, namely labour market concentration, tightness, and the diversity of industries and occupations. Local labour market tightness (LMT) is an important aspect affecting labour market outcomes (Bilal, 2023; Domash and Summers, 2022), and it may affect the wages of women and men differently. It has been shown to be a good proxy for labour demand and hence affects wage setting. Greater LMT should make employers willing to offer more attractive compensation packages, including higher wages and better working conditions, in order to attract and retain workers. With greater competition for workers, employers will be less likely to discriminate against women in terms of wages, hiring or promotions, resulting in a smaller gender pay gap. We hence consider the effect of LMT, defined as the ratio between the number of job vacancies and the number of unemployed actively looking for a job. Whether denser locations exhibit higher LMT than less dense CZs is not a priori clear, and there is no evidence of this. In denser CZ, a more dynamic labour market should result in more vacancies and lead to greater tightness. Tightness, however, also depends on local unemployment, and in France, high-density areas are characterized by high unemployment rates.²⁵ Our data does not indicate that any of the effects dominates, as we do not find a significant correlation between the CZ's density and its degree of tightness. This is reflected in column (4), where we can see that although labour market tightness matters for the local gender gap in

²⁴Results with an alternative measure of car commute distance instead of commute time provide similar results. Detailed information about the construction of these variables can be found in Appendix A.1.

²⁵See Figure B.9 in Appendix B.

premium wage, it does not explain why this gap is negatively correlated with population density (explaining 0% of the change in the coefficient).

Concentration in the local labour market refers to the degree to which a small number of employers dominate a specific job market within a given CZ. Dodini et al. (2024) find that women are more likely to work in highly concentrated occupations. In highly concentrated labour markets, employers may have more power to set wages and working conditions, reducing the bargaining power of workers, especially that of women who have been shown to be less willing to negotiate (Leibbrandt and List, 2015). This argument is captured by Hirsch et al. (2013) who develop a spatial model of monopsonistic competition in which urban areas have thick labour markets that give rise to a more competitive environment. This constrains employers' ability to engage in monopsonistic Robinsonian discrimination, which favours women. Discriminatory practices may also be more prevalent in concentrated markets if fewer opportunities are available for workers.

To measure labour market concentration, we follow Marinescu et al. (2021) and use the labour market Herfindahl-Hirschman Index (HHI) which computes concentration from employment shares. For a given occupation, CZ, and year, the labour market share of a firm is determined by the number of workers it employs relative to the total number of employed workers. The HHI of an occupation is then calculated by summing the squared labour market shares of all firms hiring in that occupation, location, and year. To obtain an index for each commuting zone, we average the HHI across occupations for the CZ, weighting by the share of workers in the occupation. As Figure B.10 and Marinescu et al. (2021) show, local labour market concentration is smaller in denser areas, explaining why this feature of the labour market contributes negatively and substantially (-47%) to the gender gap in the density earning premium.

A greater *diversity of industries and occupations* should allow workers to find better matches, and men and women may benefit differently from this diversity. Papageorgiou (2022) has shown that the greater number of occupations in large cities is one of the main factors that explain the city-size earnings premium. Hence we use the inverse-Herfindahl as a measure of the diversity of industries (using the NAF 1 digit categories) and occupations (using the PCS 1 digit categories) in a location. Diversity of industry increases the gender gap in agglomeration economies while diversity of occupations contributes to decreasing it.

Overall, the results reported in columns (3) to (5) indicate that the covariates we include reduce the direct effect of density by 69%. Childcare availability, historical gender norms, commuting patterns and labour market features all play a role. The single largest effect is that of commuting patterns, but labour market concentration and industry diversity are also key factors. To our surprise, the availability of childcare and historical gender norms have only a moderate explanatory power in understanding the role of density.

The last feature we consider is occupational segregation. Occupational sorting has been shown to be a major determinant of gender wage gaps (Cortes and Pan, 2018; Blau and Kahn, 2017), hence in columns (6), (7) and (8) we additionally introduce a measure of occupational segregation. We measure occupational segregation using the Duncan dissimilarity index, a commonly used measure (see e.g Cortes and Pan (2018)). The Duncan index captures the uneven distribution of men and women across different occupations or industries in a given CZ in a given year, and our data indicate that in areas with low population density, the distinct sorting of men and women into separate occupations is more pronounced, as shown in Figure B.11.²⁶

Occupational segregation has a large explanatory power. Once we control for it, our coefficient of interest falls from 0.0049 (column (2)) to -0.0013(column (6)), leading to a change of 0.0062 (column (7)), and this variable accounts for over 100% of the change in the coefficients. The negative coefficient on density indicates that once we take into account the extent of segregation in jobs, the urban premium is higher for men than for women. In our final specification, the strongest effect comes from occupational segregation, with a contribution that is twice as large as that of the second most significant vari-

²⁶Alternative measures of segregation based on more detailed occupation classification (2, 3 or 4 digits) yield to similar results.

able, average male commuting time. This suggests that a major portion of the initial gender gap in agglomeration economies is driven by greater segregation in low-density CZs. Introducing this variable also has a considerable effect on the role played by some of the other covariates. Notably, the role of historical gender norms falls considerably (from explaining 16% of the change to 2%), a result likely due to more traditional norms leading to greater segregation of men and women across jobs. In contrast, the contributions of other labour market features remain relatively stable.

To summarize, our main finding is that the gender gap in the urban premium is largely explained by two features, dynamic returns to experience and observable differences between locations, which vary systematically with location density. Starting from an initial estimate of 0.0076, dynamic returns to experience explain about 35% of the gap and local features can explain much or all of the remaining effect, depending on the specification. A key feature is the extent of occupational segregation, although this can be thought of as a joint outcome with gender earnings gaps rather than as an explanatory variable. When we exclude this variable, local features explain almost half (46%) of the effect of density. Commuting time and local labour market characteristics play a major role. In contrast, the availability of childcare services and historical gender norms, although significant, provide a moderate contribution to explaining the effect of density.

4 Conclusion

Our analysis of the economic geography of gender wage gaps in France has demonstrated that: (i) the gender wage gap varies systematically with urban population density, female workers benefiting relative more from urban density than male workers, (ii) the returns from urban density are explained by our set of controls, which capture different mechanisms, each of which makes a distinct contribution that we decompose. Quantitatively, a large portion of the gender gap in returns to density is explained by the structure of the local labour market, notably the extent of occupational segregation which accounts for a majority of the gap. Commuting time also plays an important role in agglomeration economies. Childcare facilities and historical gender norms are important but not the dominant mechanisms. The results help our understanding of the persistence of the gender gap, as they indicate that female workers in less dense locations loose out, notably because the sorting of female workers in feminised occupations paying lower wages is more pronounced in those locations.

References

- Abowd, J. M., Kramarz, F., and Margolis, D. N. (1999). High wage workers and high wage firms. *Econometrica*, 67(2):251–333.
- Albanese, A. and Gallo, G. (2020). Buy flexible, pay more: The role of temporary contracts on wage inequality. *Labour Economics*, 64(C):S0927537120300208.
- Almeida, E. R., Araújo, V., and Gonçalves, S. L. (2022). Urban wage premium for women: evidence across the wage distribution. *World Development*, 159:106059.
- Ananat, E., Shihe, F., and Ross, S. L. (2018). Race-specific urban wage premia and the black-white wage gap. *Journal of Urban Economics*, 108:141–153.
- Babet, D., Godechot, O., and Palladino, M. G. (2023). In the land of akm: Explaining the dynamics of wage inequality in France. *Working paper*.
- Bertrand, M., Kamenica, E., and Pan, J. (2015). Gender identity and relative income within households. *The Quarterly Journal of Economics*, 130(2):571–614.
- Bilal, A. (2023). The Geography of Unemployment. The Quarterly Journal of Economics, 138(3):1507–1576.
- Blau, F. D. and Kahn, L. M. (2003). Understanding international differences in the gender pay gap. *Journal of Labor economics*, 21(1):106–144.
- Blau, F. D. and Kahn, L. M. (2017). The gender wage gap: Extent, trends, and explanations. *Journal of economic literature*, 55(3):789–865.
- Card, D., Heining, J., and Kline, P. (2013). Workplace Heterogeneity and the Rise of West-German Wage Inequality. *The Quarterly Journal of Economics*, 128(3):967–1015.
- Card, D., Rothstein, J., and Yi, M. (2024). Location, location, location. American Economic Journal: Applied Economics. Forthcoming.
- Ciccone, A. and Hall, R. E. (1996). Productivity and the density of economic activity. *The American Economic Review*, 86(1):54–70.
- Combes, P.-P., Duranton, G., and Gobillon, L. (2008). Spatial wage disparities: Sorting matters! *Journal of Urban Economics*, 63(2):723–742.

- Combes, P.-P., Duranton, G., and Gobillon, L. (2011). The identification of agglomeration economies. *Journal of Economic Geography*, 11(2):253–266.
- Combes, P.-P., Duranton, G., and Gobillon, L. (2018). The Costs of Agglomeration: House and Land Prices in French Cities. *The Review of Economic Studies*, 86(4):1556–1589.
- Combes, P.-P., Duranton, G., Gobillon, L., Puga, D., and Roux, S. (2012). The productivity advantages of large cities: Distinguishing agglomeration from firm selection. *Econometrica*, 80(6):2543–2594.
- Combes, P.-P. and Gobillon, L. (2015). The empirics of agglomeration economies. In Duranton, G., Henderson, J. V., and Strange, W. C., editors, *Handbook of Regional and Urban Economics*, volume 5, pages 247–348. Elsevier.
- Cortes, P. and Pan, J. (2018). Occupation and Gender. In Averett, S. L., Argys, L. M., and Hoffman, S. D., editors, *The Oxford Handbook of Women* and the Economy, pages 424–452. Oxford University Press.
- Dauth, W., Findeisen, S., Moretti, E., and Suedekum, J. (2022). Matching in Cities. Journal of the European Economic Association, 20(4):1478–1521.
- De la Roca, J. and Puga, D. (2016). Learning by Working in Big Cities. *The Review of Economic Studies*, 84(1):106–142.
- Dodini, S., Lovenheim, M., Salvanes, K., and Willén, A. (2024). Monopsony, Job Tasks and Labour Market Concentration. *The Economic Journal*, page ueae002.
- Domash, A. and Summers, L. H. (2022). How tight are U.S. labor markets? Working Paper 29739, National Bureau of Economic Research.
- Duranton, G. (2016). Agglomeration effects in Colombia. Journal of Regional Science, 56(2):210–238.
- Duranton, G. and Puga, D. (2014). The Growth of Cities. In Aghion, P. and Durlauf, S., editors, *Handbook of Economic Growth*, volume 2 of *Handbook of Economic Growth*, chapter 5, pages 781–853. Elsevier.
- D'Costa, S. (2024). Re-evaluating the urban wage premium: the changing roles of geographical and job transitions for women and men1,2. *Regional Science and Urban Economics*, page 104038.

- Frank, R. H. (1978). Why women earn less: The theory and estimation of differential overqualification. The American Economic Review, 68(3):360– 373.
- Gelbach, J. B. (2016). When do covariates matter? and which ones, and how much? Journal of Labor Economics, 34(2):509–543.
- Glaeser, E. and Maré, D. (2001). Cities and skills. Journal of Labor Economics, 19(2):316–342.
- Goldin, C. (2014). A grand gender convergence: Its last chapter. American Economic Review, 104(4):1091–1119.
- Graham, D. J. and Gibbons, S. (2019). Quantifying wider economic impacts of agglomeration for transport appraisal: Existing evidence and future directions. *Economics of Transportation*, 19:100121.
- Havnes, T. and Mogstad, M. (2011). Money for nothing? universal child care and maternal employment. *Journal of Public Economics*, 95(11):1455–1465. Special Issue: International Seminar for Public Economics on Normative Tax Theory.
- Hermes, H., Krauß, M., Lergetporer, P., Peter, F., and Wiederhold, S. (2024). Early child care, maternal labor supply, and gender equality: A randomized controlled trial. Technical Report 10178, CESifo Working Paper.
- Hirsch, B., König, M., and Möller, J. (2013). Is there a gap in the gap? Regional differences in the gender pay gap. Scottish Journal of Political Economy, 60(4):412–439.
- Kleven, H., Landais, C., Posch, J., Steinhauer, A., and Zweimuller, J. (2019a). Child penalties across countries: Evidence and explanations. AEA Papers and Proceedings, 109:122–26.
- Kleven, H., Landais, C., and Søgaard, J. E. (2019b). Children and gender inequality: Evidence from denmark. American Economic Journal: Applied Economics, 11(4):181–209.
- Kunze, A. and Miller, A. R. (2017). Women helping women? Evidence from private sector data on workplace hierarchies. *The Review of Economics and Statistics*, 99(5):769–775.
- Lagrosa, I. (2022). Income dynamics in dual labor markets. Working Paper.

- Le Barbanchon, T., Rathelot, R., and Roulet, A. (2021). Gender differences in job search: Trading off commute against wage. *The Quarterly Journal of Economics*, 136(1):381–426.
- Lefebvre, P. and Merrigan, P. (2008). Child-care policy and the labor supply of mothers with young children: A natural experiment from canada. *Journal* of Labor Economics, 26(3):519–548.
- Leibbrandt, A. and List, J. A. (2015). Do women avoid salary negotiations? Evidence from a large-scale natural field experiment. *Management Science*, 61(9):2016–2024.
- Liu, S. and Su, Y. (2024). The Geography of Jobs and the Gender Wage Gap. The Review of Economics and Statistics, 106(3):872–881.
- Marinescu, I., Ouss, I., and Pape, L.-D. (2021). Wages, hires, and labor market concentration. *Journal of Economic Behavior & Organization*, 184:506–605.
- Meekes, J. and Hassink, W. H. J. (2023). Endogenous local labour markets, regional aggregation and agglomeration economies. *Regional Studies*, 57(1):13–25.
- Nisic, N. (2017). Smaller differences in bigger cities? Assessing the regional dimension of the gender wage gap. *European Sociological Review*, 33(2):292– 3044.
- Oaxaca, R. (1973). Male-female wage differentials in urban labor markets. International Economic Review, 14(3):693–709.
- Ofek, H. and Merrill, Y. (1997). Labor immobility and the formation of gender wage gaps in local markets. *Economic Inquiry*, 35(1):28–47.
- Papageorgiou, T. (2022). Occupational matching and cities. American Economic Journal: Macroeconomics, 14(3):82–132.
- Phimister, E. (2005). Urban effects on participation and wages: Are there gender differences? *Journal of Urban Economics*, 58(3):513–536.
- Pora, P. (2024). Keep Working and Spend Less? Collective Childcare and Parental Earnings in France. Technical report.
- Schmutz, B. and Sidibé, M. (2018). Frictional Labour Mobility. The Review of Economic Studies, 86(4):1779–1826.

Appendix

This Appendix describes the construction of the data. The accompanying Online Appendix consists of several sections. Appendix B provides further details on the data and descriptive statistics. We report next some of the underlying estimates used in our specifications, with Appendix C providing the estimations of the first stage. We then discuss several robustness exercises in Appendix D. In Appendix E we perform an event study to examine the validity of the exogeneity of spatial mobility.

A Data and summary statistics

A.1 Construction of variables

We start by providing details on the variables used in our estimation of the urban wage premium, which are available at the BTS.

The variable occupations is coded according to the French classification PCS (at the 2-digit level) with the following categories: Agricultural managers, Artisans, Shopkeepers and related professions, Business leaders with 10 employees or more, Liberal professions and related professions, Civil service executives, intellectual and artistic professions, Corporate executives, Intermediate professions in education, health, civil service and related professions, Intermediate administrative and commercial professions in companies, Technicians, Foremen, supervisors, Civil servants, Administrative employees, Commercial employees, Direct service personnel to individuals, Skilled workers, Unskilled workers and Agricultural workers.

The variable industry is coded according to the French industry classification (NAF) with 17 sub-sections: 1. Agriculture, hunting, forestry, 2. Fishing, aquaculture 3. Mining and quarrying, 4. Manufacturing, 5. Production and distribution of electricity, gas and water, 6. Construction, 7. Wholesale and retail trade, 8. Hotels and restaurants, 9. Transport and communications, 10. Financial activities, 11. Real estate, renting and business activities, 12. Public administration and defence, 13. Education, 14. Health and social work, 15. Other social and personal service activities, 16. Activities of households as employers and 17. Extra-territorial organizations and bodies.

The densest locations (the top 20%) are, in reverse order: Aix-en-Provence, Bourgoin-Jallieu, Houdan, Lorient, Maubeuge, Metz, Annecy, Toulouse, Forbach, Orange, Chambéry, Colmar, Agde - Pézenas, Flanders - Lys, Nice, Les Sables-d'Olonne, Rouen, La Rochelle, Saint-Nazaire, Bordeaux, Brest, Saint-Malo, Le Genevois Français, Fréjus - Saint-Raphaël, Avignon, Nantes, Melun, Nîmes, Meaux, Calais, Boulogne-sur-mer, Perpignan, Mulhouse, Thionville, Le Havre, Salon-de-Provence, Mantes-la-Jolie, Dunkerque, Montpellier, Toulon, Istres - Martigues, Béthune - Bruay, Cannes - Antibes, Sète, Plaisir, Douai, Roissy - Sud Picardie, Valenciennes, Lyon, Evry, Cergy, Strasbourg, Marseille -Aubagne, Poissy, Créteil, Saclay, Lille, Marne-la-Vallée, Lens - Hénin, Roubaix - Tourcoing and Orly, Paris.

In the second stage, we use several variables obtained from both the BTS and other sources. We compute two measures of childcare availability, the ratio between the number of places in public or in private childcare facilities and the population. Data for facilities is obtained from French administrative data on firms and their establishment location and main activity ('SIRENE data') to gather information, for each commuting zone, on the number of childcare facilities for young children. The database is publicly available at https://www.sirene.fr/sirene/public/accueil?sirene/cale = en. We only select firms whose main activity is childcare for young children (NAF Rév 2 88.91A). Childcare facilities are classified as public or private depending on their legal classification defined by the French Institute of Statistics and Economic Studies. We exclude the firms whose legal classification is that of sole trader. Public childcare facilities as defined as firms which are public legal entities subject to commercial law or legal entities and organizations subject to administrative law. Private childcare facilities include firms with another type of legal status.

Information on several labour market characteristics was obtained from the BTS dataset. We measure the diversity of occupations and diversity of industries available in a location by computing inverse-Herfindahl indices for both. For occupations, we use the PCS 1-digit categories and for industries the NAF 1-digit. To measure labour market concentration, we follow Marinescu et al. (2021) and use the labour market Herfindahl-Hirschman Index (HHI) which computes concentration from employment shares. For a given occupation, CZ, and year, the labour market share of a firm is determined by the number of workers it employs relative to the total number of employed workers. The HHI of an occupation is then calculated by summing the squared labour market shares of all firms hiring in that occupation, location, and year. To obtain an index for each commuting zone, we average the HHI across occupations for the CZ, weighting by the number of workers in the occupation.

We measure occupational segregation with the Duncan dissimilarity index, a commonly used measure; see, for example Cortes and Pan (2018). The Duncan index captures the uneven distribution of men and women across different occupations or industries in a given CZ in a given year. It is computed as

$$D = \frac{1}{2} \sum_{i=1}^{n} |\text{Share of men in occupation i} - \text{Share of women in occupation i}|$$

and ranges from 0 to 1, with 0 denoting a perfectly equal distribution of men and women across all occupations, and 1 denoting complete occupational segregation. Our indices are built using PCS 2-digit as described above.

Labour market tightness is defined as the ratio between the number of job vacancies and the number of unemployed actively looking for a job. We compute this measure using data from the French Unemployment Services (STMT) for the period 2005-2016, and publicly accessible data from the French Unemployment Services website covering the period from 2017 to 2019.²⁷ Information on vacancies is derived from data from the French governmental agency for unemployment (Pôle Emploi, the Statistiques du Marché du Travail). We use occupations (PCS) at the 4-digit level.

The geo-coding of place of residence and place of work allows us to mea-

²⁷Data are publicly available online at this link.

sure where work experience was acquired in the past, as well as to measure commuting distance. The variable commute distance is created following the methodology used in Le Barbanchon et al. (2021). When workers reside and work in different municipalities, the commute distance is computed as the distance between the centroids of the workplace municipality and the municipality of residence. When workers reside and work in the same municipality, the variable is calculated as the average distance between two random locations within the municipality. The mean commuting gender gap in road time is computed based on the commute information of the BTS data and the METRICS application developed by INSEE.

Social norms are measured by the 1962 fertility rate. The fertility rate is defined as the number of births in the CZ divided by the number of women aged between 15 and 45. This measure is derived from the 1962 Census ('*Recensement de la population - 1962*') for the year 1962, which is the earliest available dataset that allows us to compute fertility rates at the CZ level.

A.2 Gender gaps in earnings and local density

Figure 1 has displayed the map of the gender earnings gap by location and has juxtaposed the map of local population density. The visual impression is that the higher the density of the population, the smaller the gender wage gap. Here we confirm this formally, by regressing the gender earnings gap on local year-by-year density. We also demonstrate that the results are consistent across various possible earnings definitions (e.g. monthly earnings, hourly wage rates). Table 1 shows that the gender earnings gap and location density are negatively correlated.

	(1) Hourly gross	(2) Hourly net	(3) Monthly gross	(4) Monthly net	(5) Hourly net total	(6) Hourly gross total
	0.011000***	0.000504***	0.010105***	0.010929***	0.000170***	0.000044**
(Log) Density	-0.011280^{***} (0.000358)	-0.009524^{***} (0.000351)	-0.012127^{***} (0.000367)	-0.010363^{***} (0.000360)	-0.006170^{***} (0.000345)	-0.008644^{***} (0.000347)
Observations	3,267	3,267	3,267	3,267	3,267	3,267
R-squared	0.3553	0.3165	0.3554	0.3157	0.1908	0.2656

Table A.1: Regression of log-gender wage gap on log-density

Note: ***p<0.01, **p<0.05, *p<0.1. Coefficients are reported with robust standard errors in parenthesis. The wage gap is measured in log for the commuting zone of residence, representing the difference between the mean wages of men and women. Column (1) corresponds to the gap in hourly gross wages, column (2) in hourly net wages, column (3) in monthly gross earnings, column (4) in monthly net earnings, column (5) illustrates the gap in hourly net wages for the total of contracts during the year, and column (6) depicts the gap in hourly gross earnings for the total of contracts during the year. All regressions are controlled for the year.

As for Figure 2, these regressions show that the largest commuting zones cluster nicely around the fitted regression line.



Figure A.1: Impact of density on gender wage gap

Notes: *BTS* database, 2019. This graph plots the linear regression of the gross earnings mean deflated gender wage gap (in log) on log density, weighted by the share of the population in the CZ.

A.3 The geography of movers flows between the different commuting zones in France

The identification of the urban wage premium relies on the movers. If moves in our data are concentrated in certain areas, our estimates may not account for existing patterns in certain sparsely populated areas, for instance, and may only highlight patterns between Paris and the French provinces. Moves are defined as a change in the worker's commuting zone of residence. Figure A.2, however, illustrates that internal migrations are fairly evenly distributed across France, although naturally, a higher volume of moves is observed between Paris and other major CS in France (such as Lyon, Lille, Marseilles, Bordeaux, etc.). Figure A.2: Map of movers' flows



Notes: BTS database, 2005-2019. This map plots the number of moves between the commuting zones. The darker and the bigger the line, the higher the number of moves.

			Departure Commuting Zone							
			Paris	Marseille	Lille	Lyon	Bordeaux	Toulouse		
	Dania	Nb of obs.		13000	12428	23351	13158	14684		
	Paris	%		0.16	0.15	0.28	0.16	0.18		
	Manasilla	Nb of obs.	18412		1128	4877	1536	2853		
	Marseille	%	0.22		0.01	0.06	0.02	0.03		
A	T III.a	Nb of obs.	12777	657		1636	854	921		
Commuting	Commuting Zone Lyon	%	0.15	0.01		0.02	0.01	0.01		
Zana		Nb of obs.	33670	4608	2420		2263	3642		
Zone		%	0.41	0.06	0.03		0.03	0.04		
	Dondoouu	Nb of obs.	24566	1752	1455	2660		7314		
	Bordeaux	%	0.30	0.02	0.02	0.03		0.09		
	Taulausa	Nb of obs.	22042	2987	1509	3530	6643			
	rouiouse	%	0.27	0.04	0.02	0.04	0.08			

Table A.2: Number of moves

Notes: BTS data, 2005-2019. The table presents the number of moves between various French commuting zones, along with the respective shares in the overall number of moves. These moves represent a change in the worker's commuting zone of residence.

Table A.2 reports the number of moves between various French commuting zones, along with the respective shares in the overall number of moves that each of them represents. Moves represent a change in the worker's commuting zone of residence. In addition, Figures B.1, B.2, and B.3 in subsection B display that mobility patterns exhibit relatively consistent trends throughout the life cycle and across gender, despite a higher frequency of moves among men.



Figure A.3: Transition matrix of movers' flows

Notes: *BTS* database, 2005-2019. This matrix plots the share of moves between two groups of commuting zones. The darker the square, the higher the share of moves.

Figure A.3 complements the aforementioned descriptive statistics by illustrating the distribution of moves across various combinations of the largest CZs, either in terms of population density or grouped by their population density. The visualization reveals that moves occur not solely between Paris and other areas.

A.4 Characteristics of the unemployed and local density

Selection bias could alter our results. Indeed, our results could be driven by gender gap in workforce composition difference between denser and less dense areas. If women with high-wage characteristics are more likely to become un-

employed when moving to less dense areas compared with men, our results could be explained by a higher negative selection bias in less dense areas for women compared with men. Hence, using administrative data of the unemployed who registered at the French Unemployment Services (STMT data) for the period 2005-2016, we compute measures of educational composition separately for men and women, for each CZ. Figure A.4 plots the share of unemployed with a higher education degree according to the population density of the CZ, and figure A.5 the breakdown by education degree. Red lines refer to women and green to men. Figure A.4 reflects population density-dependent composition effects. Both unemployed men and women are more likely to hold higher education degrees in denser CZ, reflecting the overall higher education level of the population in these areas. However, the composition effects related to CZ size appear similar for both genders. As seen in the general population, the proportion of unemployed women exceeds that of unemployed men. Besides, the gender gap in p.p in the percentage of unemployed individuals with higher education degrees is smaller in less dense areas. This suggests there is no higher negative selection bias for women compared with men in less dense areas. If such a bias existed, the positive gap between men and women would have decreased with increasing population density.

Figure A.4: Share of unemployed with a higher education degree in the CZ and density



Notes: STMT Data 2005 - 2016. This figure plots the share of unemployed registered at the French Unemployment Services who hold a higher education degree by CZ and log-density. The red line represents the share of women and the green line the share of men.



Figure A.5: Share of unemployed by education degree and density

Notes: STMT Data 2005 - 2016. This figure plots the share of unemployed registered at the French Unemployment Services who hold a higher education degree by CZ and log-density. The red line represents the share of women and the green line the share of men.