The geography of women's opportunity: evidence from Indonesia*

César Garro-Marín†

This version: November 12, 2022

Click below to get the latest version

Tables in text   Tables at the end

Abstract

This paper argues that women's birthplace strongly affects their labor force participation in adulthood. I use rich data from Indonesia and leverage variation from women moving across labor markets as children to estimate the effect on women's labor force participation of spending more time in their birthplace. My strategy compares the labor supply choices of women who currently live in the same location but who emigrated from their birthplace at different ages. I find that birthplace has strong and persistent effects on adult women's labor supply. Moreover, these effects are concentrated during the formative period between 6 and 16 years old. By the time they turn sixteen, women born in a location at the 75th of female employment will be ten p.p. more likely to work than those born in a 25th percentile location. Birthplace effects are quantitatively important. Between 21 to 45 percent of the current spatial inequality in women's labor force participation is transmitted to the next generation of women.

Keywords: gender inequality, local labor markets, place effects

JEL Codes: J16, R19, O18

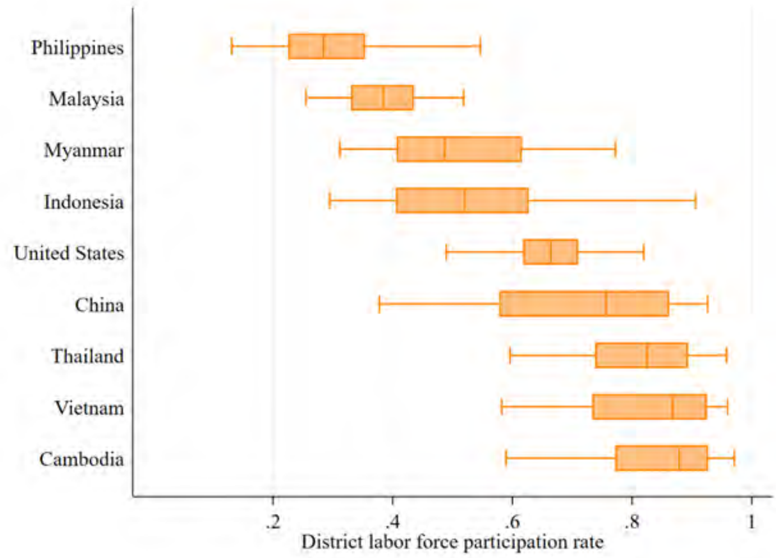
*I am indebted to my patient advisors Daniele Paserman, Kevin Lang, and Linh Tô for their constant support and helpful comments on this project. Special thanks to Bilge Erten, Dilip Mookherjee, Patricia Cortés, Johannes Schmieder, Arindrajit Dube, Silvia Prina, Nils Lehr, Masyhur Hilmy and all attendants of the Empirical Microeconomics Workshop at Boston University for helpful comments and suggestions. Special thanks to Francesco Benatti for excellent research assistance. All errors are my own.

†Boston University, email: cesarlgm@bu.edu. Mailing address: Department of Economics, Room 515, Boston University, 270 Bay State Road, Boston, MA, 02214.

1 Introduction

There are surprisingly large and persistent differences in female labor force participation (FLFP) rates within countries, and we know very little about the causes and implications of this dispersion. For example, in the United States, 76% of women work in the commuting zone of Minneapolis, while only 64% do so in the commuting zone of Los Angeles. In Italy, 59% of women work in the region of Lombardy, but only 31% in Campania, while in Indonesia, 56% of women work in the city of Denpasar, and only 43% in the country's capital of Jakarta. Beyond these three examples, figure 1 shows high dispersion sub-national labor female force participation rates for many other countries.

Figure 1: Female employment rates at the district level for selected countries



Note: The figure shows the distribution of female employment rates for a large subset of Asian countries with geographic data available in IPUMS International. Countries are ordered by median district employment rate. I use the latest available sample from IPUMS International for each country. I aggregate data at district/municipality level, except in the United States where I aggregate data for the US into Commuting Zones as in Autor and Dorn (2013).

In this paper, I start by documenting the large and persistent within-country dispersion in female labor force participation. This dispersion has generally gone unnoticed in the literature (Charles et al., 2018). Then, I argue that this dispersion in female labor force participation has strong effects on the labor market outcomes of women born across different areas within the same country. To do so, I use rich data from internal Indonesian female migrants to show that their birthplace has a strong impact on their adult labor force participation[1]. I identify the birthplace causal effect by leveraging variation coming from women living in the same labor market as adults but who left their birthplace origin at different ages as children. I exploit variation in the time spent in the birth location to distinguish the causal effect of the birthplace from variation driven

[1]Approximately a third of Indonesian women migrated at some point in their lives.

by differences in women's unobservable characteristics. My strategy boils down to comparing the labor force participation of women who emigrated in early childhood, versus those who left in their early teens. Then, if women born in places with higher female labor force participation are more likely to work the longer they stay there, I surmise that this is driven by the effect of their birth location. By comparing women living in the same location as adults, I abstract from the effect of current labor market conditions and uncover variation that is likely driven by women's labor supply. This strategy builds on that of Chetty and Hendren (2018a), and focuses it on female outcomes in a large developing country.

My results indicate that birthplace has strong and persistent effects on adult women's labor supply. I first show that women's birthplace is highly predictive of their labor supply choices. Conditional on living in the same local labor market, women born in high-female employment locations are much more likely to work than those born in places with low-female employment. This strong relationship arises even for women who emigrated young, at ages when women are unlikely to be working. These differences in labor supply could be the result of a birthplace effect, but they can also reflect differences in women's unobservable characteristics or omitted variable bias. I argue that this relationship reflects the causal effect of birthplace on women's employment by exploiting differences in the timing of emigration. I use a strategy similar to a Difference-in-Differences design, where I compare the labor force participation of women living in the same labor market but who emigrated from their birthplace at different times. Under the assumption that the omitted variable is constant for women emigrating at different ages, this strategy allows me to distinguish the causal effect from differences in women's characteristics.

I find that spending late childhood and early teen years in high-employment locations makes women more likely to work as adults. They are more likely to work the longer they stay in these high-employment locations. Staying in a place at the 75th percentile of female employment between 6 and 16 years of age makes women ten percentage points more likely to work than those born in a place at the 25th percentile. These magnitudes are quantitatively important as they imply that approximately 45% of the current spatial inequality in women's labor force participation is transmitted to the next generation of women through birthplace effects.

The causal interpretation of these birthplace effects estimates hinges on the assumption that omitted variable bias is constant across emigration ages; that is, the correlation between birthplace female employment rate and other unobserved determinants of women's labor supply is the same no matter the age they emigrated. Note that differences in factors I do not control for between women born in different locations are not enough to violate this assumption. For example, women from high-FLP locations may be more likely to work because they had parents with higher resources to invest in their education than those from low-FLFP locations. This would be a resource effect rather than a birthplace one. However, this doesn't violate the constant bias assumption. A violation would require that the resource gap be larger for women who emigrated at later ages, which is much less possible. Moreover, I find strong evidence supporting the constant selection assumption. Women from high and low LFP places show similar gaps in proxies for parental resources in other characteristics no matter the age they emigrated.

I take advantage of rich Indonesian data that stores people's birthplace and current location at a detailed geographic level. My main analyses source data from all waves of the Indonesian Family Life Survey (IFLS) and the 1985, 1995, and 2005 intercensal surveys (Statistics Indonesia, 2021; Minnesota Population Center, 2020). These representative and publicly available datasets track respondents' birthplace, current location, and migration history across midsized geographies. This level of detail allows studying differences in women's labor supply and birthplace effects at a level that is not possible in other countries from traditional sources (Bryan and Morten, 2019). Throughout the paper, I identify localities as Indonesian "regencies". There are medium-sized administrative geographies akin to counties in the United States. The average regency is approximately twice the size of the US state of Rhode Island and houses eight hundred thousand people.

This paper contributes to three strands of the literature. I contribute to the growing research showing that local labor markets can permanently affect women's labor supply, fertility, and human capital investment choices (Molina and Usui, 2022; Charles et al., 2018; Boelmann et al., 2021). I make three main contributions to this literature. First, by applying techniques borrowed from the place effects literature (Chetty and Hendren, 2018a,b; Milsom, 2021), I provide evidence causal evidence that women's birthplace has large and persistent effects on women's labor supply. This complements evidence from previous literature, which shows that exposure to current labor markets can have effects on women's expectations, labor supply, and educational investment (Molina and Usui, 2022; Boelmann et al., 2021; Milsom, 2021). Second, I also provide evidence of the ages at which birthplace is key in shaping labor supply. Although previous research has pointed out that women's childhood environment matters for their adult outcomes, this literature is mostly silent on *when* does it matter (Chetty et al., 2016). Third, my results provide new evidence that where women grow up can matter more locally. Previous research emphasizes that differences in norms, culture and other factors across large geographical areas such as states, provinces, or countries can shape women's choices (Charles et al., 2018; Boelmann et al., 2021; Alesina et al., 2013). By exploiting much more disaggregated data, my results suggest these factors can act at a more local level.

Second, this paper also contributes to the literature on place effects. Primarily using evidence from developed countries, this literature shows that where people grow up and live has important implications for intergenerational mobility (Chetty and Hendren, 2018a,b), racial inequality (Chetty et al., 2020), human capital accumulation (Molina and Usui, 2022), and criminal activity (Damm and Dustmann, 2014). I add to this literature by providing new empirical evidence linking women's birthplace to their outcomes as adults in a large developing country. In this way, my findings complement existing work showing that spatial inequality is particularly important for women's human capital investment in West Africa (Milsom, 2021) and Japan (Molina and Usui, 2022).

Finally, my paper also contributes to the literature studying the determinants of women's labor supply. This literature has primarily exploited cross-country differences in female labor supply to study its determinants and its implications (Olivetti and Petrongolo, 2008, 2014; Blau et al., 2020; Blau and Kahn, 2015). In this paper, I document the existence of large and persistent differences in female labor supply within multiple developing countries and study some of its implications. In this way, my approach is closer to the recent literature documenting that factors such as commuting and sexism can help explain the geographic differences in women's labor supply within the United States and France (Charles et al., 2018; Le Barbanchon et al., 2021; Black et al., 2014; Moreno-Maldonado, 2019).

This paper is structured as follows: after the standard description of the data in the next section, I document four facts about the within-country dispersion in female labor force participation in section 3. Then, in section 4, I study the implications of this dispersion on women growing up across different localities of Indonesia. In this section, I explain my empirical strategy and present my results. Section 6 concludes.

2 Data

2.1 Data sources

My main analyses use data from the Indonesian Family Life Survey (IFLS) and the Indonesian Intercensal Survey (SUPAS). These two datasets record detailed data on people's birthplaces, their migration histories, and their labor supply. I supplement these data with place characteristics coming from the Indonesian Census and the National Socioeconomic Survey (SUSENAS).

The IFLS is a rich panel survey tracking the information of approximately 34 thousand Indonesians across five survey years: 1993, 1997, 2000, 2007, and 2014. Overall, the IFLS is representative of 83% of the Indonesian population[2] (Hamory et al., 2021). This dataset has three advantages that make it uniquely suitable to study place effects on female labor supply. First, it records people's birthplace and location of birth in midsized geographic units. The survey tracks this information at the level of the "regency", which are administrative units similar to US counties. Research on Indonesia typically uses them to identify local labor markets (Magruder, 2013; Bazzi et al., 2022),

[2]The IFLS originally sampled households from 13 of the 27 provinces that existed in 1993. These provinces account for 83% of the Indonesian population. Appendix figure C.2 shows the geographic area covered by the original IFLS sample. Subsequent waves include data from outside this area whenever members from the original 1993 households moved to these locations. While most of my respondents are located only in these provinces, appendix figure C.3 shows a considerable amount of respondents were born outside these areas.

and their size allows me to study differences in women's employment across smaller geographic units than what I could observe in alternative datasets[3]. The typical regency is home to approximately eight hundred thousand people and covers an area roughly twice size of the US state of Rhode Island. Appendix figure C.1 depicts all the 268 regencies in my dataset.

Second, the survey tracks all people's migration episodes since they were 12 years old, their location at twelve years old, and their work history in all the years since the last survey year. With this information, I build a panel dataset tracking migration since birth and the work histories of each individual for up to 26 years, from 1988 to 2014. Finally, the IFLS also contains detailed information on childhood conditions, marriage history, and fertility, which allows me to control for a rich set of potential confounders.

The IFLS's main limitations are its relatively small sample size and the fact that I cannot observe the exact time of migration for all episodes happening before people turned 12 years old. To address these limitations, I use data from 1985, 1995, and 2005 SUPAS. SUPAS is a nationally representative cross-sectional dataset with large sample sizes. It also tracks birthplace and current location at the regency level and contains data that allows me to infer the emigration age for all migrant respondents. The survey's limitation is that it includes much less individual information than the IFLS.

I source place characteristics data from the 1980-2010 Indonesian Decennial Censuses available in IPUMS International (Minnesota Population Center, 2020) and 2012, 2013, and 2014 SUSENAS. (Statistics Indonesia, 2019, 2020). The Censuses and SUSENAS are very similar to each other. Both are large and nationally representative datasets that track all respondents' regencies of birth and current residency. However, while SUSENAS has smaller samples, it contains a richer set of information unavailable in the Census. For example, data on wage, age at first marriage, fertility, etc., is available only on SUSENAS. I compute all regency characteristics by aggregating these datasets at the regency level, restricting the sample to people aged 18 to 64. Whenever possible, I compute these aggregates from the Census.

2.2 Measurement

My main measure of women's labor supply is a dummy equal to one if she was employed during the year[4]. This is the variable that I can track most consistently across time in the IFLS. However, as a robustness check, I also show that my results carry through to alternative measures such as being a paid worker, total weekly hours worked, and being a full-time worker.

[3]Datasets available for other countries track geographic information only for states or provinces, which in most cases are either too big or too few to be interesting (Bryan and Morten, 2019)

[4]This counts unpaid family workers as employed. I also conduct robustness checks where I consider paid workers as employed.

Throughout the analysis, I associate women's labor supply choices with the characteristics of their birthplace. This exercise requires having geographic units with boundaries that are fixed across time. However, regency boundaries changed considerably from decade to decade between 1980 and 2010. Moreover, the creation of new regencies was a common event. Appendix table D.5 shows that just between 2000 and 2010, 154 new regencies were created. I address this issue using regency aggregates with consistent boundaries between 1970 and 2010. These regency aggregates were built by IPUMS International (Minnesota Population Center, 2020). There are 268 consistent-boundary regencies which are just slightly larger than the "original" regencies in the data. From now on, I will refer to these regency aggregates as regencies.

For my main analysis, I restrict my sample to internal migrants because this is the population for which I can separate the current place of residence from the birthplace. I define migration –emigration– as living outside the regency of birth. Moreover, whenever I associate women's employment with birthplace characteristics, such as FLFP or urbanicity, I source these from the 2010 Indonesian Census. In robustness checks, I show that my results are similar when I use information from other census years.

2.3 Summary statistics

In what follows, I will briefly overview my data and the Indonesian labor market using the IFLS. I obtain a similar picture with the Intercensal Survey. Overall, both sources give the same consistent picture. I start by giving a general description of the whole dataset as well as women and men separately in table D.1. There I highlight three important features of the Indonesian labor market. First, internal migration is a relatively common phenomenon. Approximately a third of Indonesians lived outside their birthplace by age 25. These internal migrants are the people I use for my primary analysis; thus, they englobe a large cross-section of the Indonesian population. Second, the Indonesian labor market is highly informal and agrarian. Forty-six percent of Indonesians are self-employed, and about a third work in agriculture[5]. There are large gender gaps in employment, type of worker, and industry of employment. Women are 34 p.p. less likely to work than men, a gap that, while large, is in line with those in Southeast Asia. Moreover, women are twice as likely as men to be unpaid/salary workers. Unpaid workers work or help earn an income but are not paid a wage or salary. Most unpaid workers work in agriculture (82%) and the retail industry (11%) (Minnesota Population Center, 2020). Finally, women are more likely to work in the service and manufacturing industries.

For my main analysis, I restrict my sample to women emigres because these are the people for whom I can distinguish between the current place of residence and the regency of birth. In table D.2 I zoom in on these populations and compare them to women that *never* left their birthplace

[5] In comparison, in the United States, only 10% of workers are self-employed, and 1% of work in agriculture

–stayers–. I show statistics for all emigres, and for those that emigrated before they turned 19. The table shows several salient differences between emigres and stayers. Despite being much more educated, women emigres are not more likely to work. Moreover, women emigrants are more likely to be salaried and display higher shares of urbanicity. This suggests that these women are predominantly migrating to places that are less rural and with more formal labor markets. The table also shows statistics for women who emigrated before they turned 19. These women are, overall, very similar to the the typical female emigre, but with just slightly lower education and employment.

Women emigration is primarily driven by family reasons. In table D.3 I breakdown women's self-reported reason for leaving their birthplace using data from the Intercensal Survey[6]. Column (1) shows that over 80 percent women's emigration are family moves. These are predominantly migrations where women follow another household member. Only thirteen percent of moves are driven by women's own work. Surprisingly, column (2) shows these figures are similar for women who emigrated before they turned 19. In comparison, column (3) shows that men are much more likely to migrate because of work.

Non-work reasons are the main drivers of women's migration. I show this in table D.3, where I describe emigration motives using data from the 1985 Intercensal Survey. This survey contains information on migration motives for all migrants in the sample. Because the IFLS is a representative panel, this table gives a good approximation of the migration motives for the migrant women that I described in table D.2. Columns (1) and (2) show that about 81% of women migrated because of family or other reasons while roughly less than 13% do so because of work, with just slight differences by emigration age. As a benchmark, I show the patterns for men in columns (3) and (4). Although non-work reasons are still important, men are much more likely to migrate because of work.

The fact that migrant women are more likely to work in the service sector could suggest that migration in Indonesia is predominantly rural to urban. However, table D.4 shows that there are substantial rural-to-rural and urban-to-urban flows. In this table, I follow Bryan and Morten (2019) and classify regencies into urban or rural according to the share of the regency's population that lives in areas that Statistics Indonesia labels as urban in the Indonesian Census. Urban regencies are those whose urban population is above a set cutoff. I chose this cutoff to match the urban share in the census. I then compute migration statistics for women born in urban and rural regencies. The table shows three salient features. First, migration is not exclusive to rural regencies. Women from urban and rural regencies migrate at similar rates. Second, migration is not just rural-to-urban. Panel A breaks down the migration episodes by origin and destination. Note that 37% percent of rural women and 72% of urban women migrate to a regency of the same category. Finally,

[6]The IFLS contains migration motive information only for people migrating at 12 years old or older. The 1985 Intercensal survey gives a better picture of migration motives for people who left very early because it contains information for migration episodes happening at all ages.

panel B shows that there is considerable heterogeneity within each regency classification. There, I show summary statistics for the female employment rates within these categories. Women are less likely to work in urban areas, but there is substantial dispersion in female employment *within* these categories. Thus, the dispersion in female employment rates I discuss in the next section is not driven by differences between urban and rural areas.

3 Four facts about women's labor supply

In this section, I use data from IPUMS International and the 1980-2010 Indonesian Censuses to present four empirical facts on female employment. First, I use data from several countries to show that large geographic differences in women's employment rates within countries are pervasive across the world. Next, I zoom in on Indonesia and (i) characterize the large dispersion in female employment across regencies, (ii) document that it is highly persistent over time, and (iii) show that it is not accounted for by variation in women's demographics or labor market characteristics across regencies. Taken together, these four facts suggest that structural differences could be driving the dispersion in women's labor supply within Indonesia.

3.1 Fact 1: within-country dispersion in women's labor supply is pervasive across countries

In table 1, I compute summary statistics of women's and men's regional employment rates for several Asian countries and the United States. For each country, I restrict the sample to people aged 18 to 64 and aggregate the employment data at the smallest geographical unit available. For most countries, this corresponds to an administrative region akin to a county or municipality, except in the United States, where I aggregate at the Commuting Zone level. In the table, I ordered countries from highest to lowest regional dispersion in women's employment[7].

This table shows three important pieces of information. Panel A highlights that despite the significant differences at the mean, these countries have large regional differences in women's employment rate *within* their borders. For most countries, the gap between the localities at the 75th and 25th percentiles (IQR) is above 15 percentage points (p.p.). A gap of 15 p.p. is large even for high female employment countries such as Vietnam, Cambodia, and Thailand. Even the smaller IQR of 9 p.p. for the United States acquires significance when we consider that this gap is equivalent to the change in the national US female employment rate during the last *thirty-eight years* (1984-2022)[8]. Second, panel A also shows that Indonesia is not an outlier. The large dispersion of female employment in Indonesia is well in line with that of countries such as China, Myanmar, and Vietnam.

[7] I give additional details about the cross-country data in section A.1

[8] This benchmark is not affected by the Covid-19 drop in women's employment. By 2022, women's employment had recovered to pre-Covid levels.

Third, panel B shows that the large geographic dispersion is mainly exclusive to women. This panel displays similar statistics for men's employment rates. In all countries but the US, the dispersion in women's employment rates is substantially larger than in men's. In fact, in six of the nine countries, the dispersion in women's employment *more than doubles* that of men's. This indicates that men work at high rates everywhere within these countries. Women, however, work at very different rates depending on the locality they live in.

Table 1: Dispersion in regional employment rates for selected countries, people aged 18-64 years

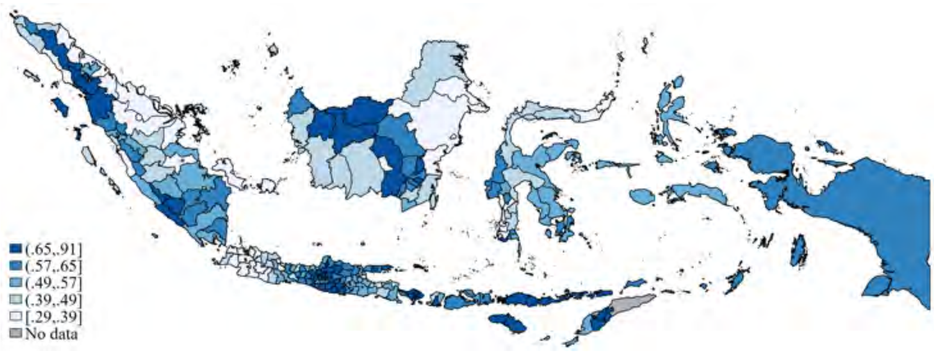
Statistic	**China**	**Indonesia**	**Myanmar**	**Vietnam**	**Cambodia**	**Thailand**	**Philippines**	**Malaysia**	**USA**
A. Women									
IQR	0.28	0.22	0.21	0.19	0.16	0.16	0.13	0.11	0.09
SD	0.17	0.14	0.13	0.12	0.11	0.11	0.10	0.07	0.07
Mean	0.71	0.53	0.51	0.82	0.84	0.81	0.30	0.38	0.67
B. Men									
IQR	0.14	0.05	0.07	0.06	0.08	0.08	0.08	0.06	0.10
SD	0.10	0.04	0.05	0.06	0.05	0.06	0.06	0.04	0.07
Mean	0.85	0.87	0.86	0.90	0.90	0.88	0.82	0.84	0.77
Mean population 18-64	266,748	533,867	83,531	79,146	50,186	58,290	40,423	91,509	202,635
No. districts	2,845	268	362	674	174	670	1,274	133	722

Notes: SD and IQR stand for Standard Deviation and Interquartile Range. The table shows statistics from a large subset of Asian countries in IPUMS International with data available at a small geographic level plus the United States. Columns are ordered from highest to lowest dispersion in women's labor supply. I use the latest available sample from IPUMS International for each country. I aggregate data at the smallest geographical unit available, except for the USA where I use Commuting Zones as in Autor and Dorn (2013). I show the unweighted cross-locality means which –might– differ from the national-level means. Section A.1 in the appendix describes the cross-country data underlying this table in extensive detail.

3.2 Fact 2: there is large within-country dispersion in women's employment rates in Indonesia

Figure 2 zooms in on the large dispersion in female employment rates in Indonesia. In this figure, I plot women's employment rates across all the 268 regencies in my dataset. Each color groups a quintile of the regencies. Darker blues denote higher employment rates. This map makes it evident that women work at very different rates across the country. Women in the top quintile of regencies see employment rates higher than 65%, while those living in the bottom quintile see rates below 29%. These low-employment regencies include important population centers, such as the Bogor regency and the city of Medan[9]. The map also highlights that the dispersion in women's employment extends across the whole country and is not driven by any particular province, island, or group of regencies.

Figure 2: Indonesia: women's employment rate by regency, 2010



Note: The figure shows regency-level employment rates for women aged 18-64. It shows all the 268 regencies with consistent boundaries between 1970 and 2010. Each color groups a fifth of the regencies. The figure uses data from the 2010 Indonesian census from IPUMS international.

3.3 Fact 3: women's employment rates are highly persistent

The large dispersion in women's employment rates could be the result of (i) temporary economic shocks that depress women's employment in some parts of Indonesia, (ii) measurement error in the employment rates, or (iii) structural differences across regencies correlated with female employment. If the dispersion arises predominantly due to temporary shocks or measurement error, we should expect very low persistence in the regencies' employment rates across years. This is because any temporary shock should dissipate after several years, and I expect measurement error to be independent across time. In contrast, high cross-year persistence can be taken as evidence that the dispersion in women's employment reflects structural differences across regencies.

[9]Medan is the capital and largest city in the province of North Sumatra. As of 2020, it is also the third most populous city in Indonesia (Brinkhoff, 2022). Bogor is a regency with more than 5 million people. It borders the Jakarta metropolitan area. See their locations in figure C.5 in the appendix.

Table 2: Indonesia: autocorrelation in regency-level men's employment rate, 1980-2010

Regressor	(1)	(2)	(3)	(4)
Female employment 10 years ago	0.80			
	(0.02)			
Female employment 20 years ago		0.72		
		(0.03)		
Female employment 30 years ago			0.70	
			(0.04)	
Same-year male employment				0.51
				(0.04)
Observations	800	534	268	1,071

Notes: The table shows the autocorrelation of regency-level employment rates across different time horizons. It also shows the simultaneous correlation between the employment of both genders. Data from 1980-2010 Indonesian Census taken from IPUMS international. Robust standard errors in parenthesis.

In columns (1) to (3) of table 2, I show estimates of the autocorrelation of the regency-level employment rates across different time horizons. For this table, I standardize the regency employment rates by year and run regressions of the form:

$$e_{rt} = \gamma_{t-j} e_{rt-j} + \varepsilon_{rt} \tag{1}$$

where e_{rt} is the standardized employment rate in regency r at time t.

The autocorrelation estimates suggest that the dispersion in women's employment rates reflects structural differences across regencies. The autocorrelations are very high, starting at 80% for ten years and staying as high as 70% for thirty years. As a benchmark, in column (4), I also show the estimate of the simultaneous correlation with men's employment rates. Note that women's employment rates are *more* correlated with themselves 30 years apart than with the *same-year* male employment rates. This high persistence indicates that the differences in female employment rates are not driven by temporary shocks or measurement error.

3.4 Fact 4: dispersion in women's employment rates cannot be accounted by differences in women's demographics alone

The large persistence in the female employment rates suggests that the dispersion in female employment rates reflects structural differences across regencies. These might be differences in the family structure or features of the labor demand across regencies. Motherhood is associated with lower attachment to the labor market by women (Angelov et al., 2016; Kleven et al., 2019). Moreover, differences in the industry mix account for up to 80% of the variation in women's labor supply in developed countries (Olivetti and Petrongolo, 2016). Therefore, the dispersion in female employment rates might just reflect differences in the family structure or the industry mix across regencies.

In table 3, I test whether permanent differences in the industry mix or women's demographics can account for most of the dispersion in employment across regencies. This table shows the R^2 from regressions of employment rates on a series of controls. They include the share of people married, the share with small children, along with measures of the age structure, the education level by gender, and the industry mix. I run the regressions separately by gender and stack data from all 1980-2010 censuses. Additionally, I include year fixed-effects to capture national trends in employment.

Table 3: Indonesia: share of employment rate dispersion accounted for observed regency characteristics, 1980-2010

	Women's					**Men's**				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
R^2	0.13	0.26	0.30	0.31	0.47	0.01	0.41	0.60	0.69	0.79
Year FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Age structure		✓	✓	✓	✓		✓	✓	✓	✓
Women's education level			✓	✓	✓					
Men's education level								✓	✓	✓
Share married				✓	✓				✓	✓
With child under 5				✓	✓				✓	✓
Industry shares					✓					✓
N	804	804	804	804	804	804	804	804	804	804

Notes: The table reports the R^2 of a regression of regency employment rates on regency-level aggregates. Age structure controls are shares aged 30-49 and 50-64. Education level measures are shares who attended at most middle school, high school, and college. When indicated, the regressions include 1-digit industry shares. Sample restricted to people aged 18-64. Data from IPUMS International.

Table 3 shows that differences in women's demographics or the industry mix account for, at most, a moderate share of the dispersion in female employment across regencies. Column (4) shows that controlling for women's education level and the regency's family and age structure accounts for only a third of the dispersion in employment rates. Adding a complete set of industry shares takes the R^2 to 47%. Although these R^2 indicate that these factors can explain part of the cross-regency variation in employment rates, collectively, they still leave 53% of women's dispersion unaccounted for. In contrast, column (10) shows that when I perform an analogous exercise for men, these same variables can account for 80% of the variation in men's employment rates. Therefore, the dispersion in female employment rates reflects variation in *other* factors that are *specific* to women.

4 Empirical strategy and results

In this section, I present evidence of the large and persistent effects of birthplace on women's labor supply using data from women migrants in Indonesia. I start by illustrating how I identify these effects using information from women who currently live outside their birthplace. Next, I present the empirical evidence in three steps. First, I show that, conditional on the current place of residence, birthplace is highly predictive of women's labor supply in adulthood. This persistence can reflect the causal effect of birthplace or a spurious correlation driven by women's unobserved characteristics. Next, I build on this result and show that birthplace also has high predictive power for women who left their birthplace before they turned 18, a sample for which migration driven by current employment opportunities is less prominent. Finally, using a strategy similar to (Chetty and Hendren, 2018a), I show that birthplace has stronger predictive power for these early women migrants the longer they stay in their birthplace. These estimates assume that the degree of selection is independent of the age at which these women left their birthplace. I present evidence providing considerable support for this assumption.

4.1 Place and women's labor supply: the identification challenge

The place of residence can, directly and indirectly, affect women's labor supply. Direct effects affect the labor supply of all the current female residents. There is considerable empirical evidence documenting these effects. These might arise, for example, from factors such as the levels of childcare availability (Compton and Pollak, 2014), commuting costs (Le Barbanchon et al., 2021; Farre and Ortega, 2021), the industry makeup (Olivetti and Petrongolo, 2014), or the level of sexual discrimination in the local labor market (Charles et al., 2018). Differences across localities in any of these factors will cause geographic differences in women's labor supply. However, place can also affect women indirectly by affecting their preferences and the skills they acquire. Women born and brought up in locations where many women work can internalize these norms and thus be more likely to work as adults (Charles et al., 2018; Boelmann et al., 2021). Moreover, environments with high female employment could make women more likely to invest in acquiring the skills they need to participate in the labor market (Molina and Usui, 2022). These permanent indirect effects will create differences in labor supply across women born in different locations *irrespective* of where they currently reside. Evidence on indirect effects is much more scarce in the literature (Charles et al., 2018).

The omitted variable problem

In this paper, my main interest lies in determining what women's labor supply would be if, conditional on the current place of residence, she was born in an area where more women work. This counterfactual exercise keeps the woman, her family, and her place of residence fixed and varies only her childhood experience. To answer this question, I study the labor supply of women residing outside their birthplace. Because for these women, the place of residence is different from their birthplace, I can separate the indirect effects from the direct impact of place. More formally, let us consider the following model for women's probability of employment e_i,

$$e_{it} \quad = \quad \delta_c + \boldsymbol{\sigma} p_b + \eta_{it} \tag{2}$$

In this model, women's employment choices depend on three main factors. First, a place-of-residence fixed effect δ_c captures all the direct effects of location c on female labor supply. These might include commuting costs, childcare availability, and gender discrimination. Second, the birthplace female employment p_b is intended to capture the causal effect of growing up in a location where p_b percent of the women work. Finally, the error term η_{it} captures all other factors making some women migrants more likely to work than others.

Model (2) follows closely the tradition brought forth by the "epidemiological" approach literature (Fernández and Fogli, 2006). Women's birthplace could have multiple impacts on women's behavior as adults. Including the prevailing female employment rates as the main regressor in equation (2) relies on the idea that these rates capture the place-driven factors vital in determining women's employment choices. This specification also facilitates testing whether alternative channels are driving the relationship with the birthplace employment rates (Fernández, 2007).

In model (2), $\boldsymbol{\sigma}$ captures the birthplace effects. It gives the counterfactual increase in women's employment if they had been born in a place with a one p.p. higher female employment rate. In the ideal, but unfeasible experiment, I would reassign women's birthplace randomly while keeping their family and the current residency fixed. Random assignment would guarantee that women's birthplace is uncorrelated with the error term. Thus an OLS regression of (2) would give a consistent estimate of $\boldsymbol{\sigma}$. In observational data, however, it is likely that the unobserved factors imbued in the error term are correlated with birthplace labor supply. Therefore, the OLS will conflate the causal effects of birthplace with omitted variable bias:

$$\begin{aligned} \text{plim } \hat{\boldsymbol{\sigma}} &= \boldsymbol{\sigma} + \frac{\text{cov}(\tilde{p}_b, \tilde{\eta}_{it})}{\text{var}\,(\tilde{p}_b)} \\ &= \boldsymbol{\sigma} + \boldsymbol{\gamma} \end{aligned} \tag{3}$$

where tilde accented variables are residualized from regency fixed effects (Angrist and Pischke, 2009). Expression (3) shows that the OLS coefficient $\boldsymbol{b}$ reflects two factors: first, the causal effect of birthplace $\boldsymbol{\sigma}$, but also differences in unobservable characteristics across women from different origins $\boldsymbol{\gamma}$. The critical identification challenge is separating the selection term $\boldsymbol{\gamma}$ from the birthplace effect $\boldsymbol{\sigma}$.

The selection term $\boldsymbol{\gamma}$ makes explicit that even in the absence of a causal effect, my birthplace could be capturing characteristics about me or my family that are relevant for my work decision. I will be arguing later that the causal effect of place is positive ($\boldsymbol{\sigma} > 0$). That is, being born in a place where more women work, makes you more likely to work. In these circumstances, I will be more concerned with omitted variable –or selection– bias making women from high-employment birthplaces more likely to work than their low-employment counterparts. For example, previous research shows that daughters from working mothers are more likely to work (Fernández, 2007). Even in the absence of a causal effect, a positive $\hat{\boldsymbol{\sigma}}$ could simply be reflecting that, in places where more women work, girls are more likely to be brought up by a working mother.

Using emigration age data to identify causal effects

Under additional assumptions, data on the age of emigration allows me to distinguish selection from the causal effect of place. The argument is similar to that of Chetty and Hendren (2018a). I assume that place effects are stronger the longer women stay there. Thus, the employment choice for women who emigrated at age a is determined as follows:

$$e_{it} \quad = \quad \delta_c + \lambda_a + \boldsymbol{\sigma}_a p_b + \eta_{it} \tag{4}$$

Here $\boldsymbol{\sigma}_a$ captures the cumulative effect of birthplace up to age a[10]. The age of emigration fixed-effects λ_a absorb differences in labor force participation across women who emigrated at different ages. The causal impact of staying in the birthplace at age a is then $\boldsymbol{\pi}_a = \boldsymbol{\sigma}_a - \boldsymbol{\sigma}_{a-1}$.

By an argument analogous to that in expression (3), the OLS estimates will conflate the causal effects of birthplace $\boldsymbol{\sigma}_a$ with the omitted variable bias for women migrating at age a $\boldsymbol{\gamma_a}$[11]:

$$\text{plim } \hat{\boldsymbol{\sigma}}_a \quad = \quad \boldsymbol{\sigma}_a + \boldsymbol{\gamma}_a \tag{5}$$

Assumption 1. ***Constant omitted variable bias***
Omitted variable is the same no matter the age of emigration, that is $\boldsymbol{\gamma}_a = \boldsymbol{k}$

This assumption requires that, conditional on the location and age of emigration fixed effects, the correlation between the birthplace employment rate and the error term is the same for women who emigrated at different ages. To illustrate what this assumption entails, let us consider the motive for emigration. It is natural to expect women who emigrated for their work to be more likely to be working in their destination and women from high-employment locations to be more likely to relocate because of their job. Moreover, I would expect seventeen-year-old women to be more likely to migrate for their own work than fourteen-year-olds. However, this does not *necessarily* violate my estimation strategy because I need something much weaker and akin to a parallel-trends

[10] The causal effect $\boldsymbol{\sigma}$ in the previous subsection can be interpreted as a weighted average of age-specific causal effects.

[11] You can find the full derivation of this expression in appendix section B. I defined γ_a as the a-th element in the vector plim $\left[(\widetilde{P}'\widetilde{P})^{-1}\widetilde{P}'\tilde{\boldsymbol{\eta}}\right]$. Here, P is the matrix containing the interaction between the age of emigration dummies and the birthplace female employment rates. $\boldsymbol{\eta}$ is the vector of error terms. Tilde-accented variables are residualized from current location and age of emigration dummies.

condition. I require that, if fourteen-year-olds from high-employment locations are 5 p.p. more to emigrate because of work than those from low-employment places, then this gap must be the same for those emigrating as seventeen-year-olds.

Under the constant omitted variable bias assumption, I can isolate the birthplace causal effect from the omitted variable bias. By subtracting the OLS estimates across different emigration ages, the constant selection term γ goes away, leaving only the causal effects:

$$\begin{aligned} \text{plim } \hat{\boldsymbol{\sigma}}_{\boldsymbol{a}} - \hat{\boldsymbol{\sigma}}_{\boldsymbol{a-1}} &= \boldsymbol{\sigma}_a - \boldsymbol{\sigma}_{a-1} \\ &= \boldsymbol{\pi}_a \end{aligned} \tag{6}$$

this expression also shows that identification does not necessarily require constant bias across all *all* emigration ages. If, instead, selection is constant only within some age ranges, I can still identify the effects within those ranges. For example, suppose there is reason to believe that the bias for women who emigrated between 0 to 6 years old is different than for those who emigrated between the ages of 7 to 15. If constant selection holds *within* these ranges, I can still identify the place effects within the 0 to 6 and 7 to 15 ranges. In section 4.4, I present estimates from the birthplace effects along with evidence that women emigrating at different ages are similar to each other along multiple dimensions.

Although the specification and identification strategy is similar to Chetty and Hendren (2018a)'s seminal paper on intergenerational mobility, there is a significant difference. While I focus on understanding the effect of the birthplace, Chetty and Hendren (2018a) concentrate on how exposure to a specific destination affects your outcomes. There are two main reasons for this choice. First, focusing on the birthplace allows me to abstract from issues like labor demand and wages, which are likely to affect women's decision to work. Second, in the context of female labor supply, women's birthplace can capture factors such as culture and gender norms, which can permanently shape women's decisions (Fernandez and Fogli, 2009).

4.2 Birthplace is highly predictive of women's labor supply

I start by comparing the labor supply of women who *live in the same location* but were born in different regencies. I do this by regressing a dummy equal to one if the person is employed at year t (e_{it}) on the regency of current residency fixed-effects (δ_c), year fixed-effects (t), women's employment rate in her regency of birth (p_b), and a set of individual and regency-level controls X_{it}. These controls might include age, religion, education, number of books at home when growing up.

$$e_{it} = \delta_c + \theta_t + \boldsymbol{b} p_b + X_{it}\kappa + \varepsilon_{it} \tag{7}$$

I source the regency female employment rates from the 2010 census. Because these employment rates are highly persistent, I obtained similar results using data from other census years[12].

In this regression, $\boldsymbol{b}$ is the parameter of interest. It captures the *association* between women's labor supply and the prevailing female employment rate in their birthplace. For now, I will refer to $\boldsymbol{b}$ as the birthplace persistence coefficient. Because the model includes regency of residency and year fixed-effects, $\boldsymbol{b}$ is –primarily– identified out of differences in labor supply of women who live in the same regency, in the same year, but who were born in different localities. In this model, δ_c absorbs differences in women's labor supply driven by permanent differences in the localities of residency, such as average wages, industry mix, healthcare availability, etc. At the same time, θ_t captures national secular trends in women's labor supply.

I call the slope of the birthplace employment rate $\boldsymbol{b}$ –rather than $\boldsymbol{\sigma}$– to emphasize that it generally differs from the causal effect discussed in section 4.1. A positive $\boldsymbol{b}$ could reflect differences in factors utterly unrelated to the birthplace, making women from high-female employment regencies more likely to work than their counterparts from low-employment regencies. For example, women from high-employment locations can just have characteristics that I do not observe that make them more likely to work.

[12]See table 2. See appendix table D.9 for more details.

Table 4: Indonesia: estimates birthplace persistence on women's labor supply ($\boldsymbol{b}$)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Women's employment rate at birthplace (p_o)	0.38***	0.39***	0.35***	0.37***	0.34***	0.34***	0.29***	0.24***
	(0.04)	(0.04)	(0.05)	(0.04)	(0.04)	(0.04)	(0.08)	(0.08)
Mean employment rate	0.54	0.54	0.54	0.54	0.54	0.54	0.51	0.51
Implied IQR gap	0.08	0.09	0.08	0.08	0.08	0.08	0.06	0.05
Sample	Full	Full	Full	Full	Full	Full	Known mother	Known mother
Year FE	✓	✓	✓	✓	✓	✓	✓	✓
Regency FE	✓	✓	✓	✓	✓	✓	✓	✓
Age		✓	✓	✓	✓	✓	✓	✓
Religion			✓	✓	✓	✓	✓	✓
Education				✓	✓	✓	✓	✓
Childhood SES					✓	✓		
Siblings						✓		
Mother worked								✓
Observations	64,501	64,501	64,501	64,501	64,501	64,501	18,135	18,135
N individuals	6,115	6,115	6,115	6,115	6,115	6,115	2,640	2,640
R^2	0.10	0.12	0.13	0.14	0.14	0.14	0.14	0.14

Notes: Uses data from IFLS. Sample restricted to people residing outside their birthplace. Implied IQR gap shows the implied employment gap between someone born at a regency at the 75th percentile of employment rate and someone born at the 25th percentile. The IQR of the female employment rate across regencies is of 22 percentage points. Standard errors clustered by regency of origin. When indicated, the regressions control for a quadratic polynomial in age, and fixed-effects for seven religion and for education categories. Standard errors clustered by regency of origin.

Table 4 shows estimates of the birthplace persistence coefficient $\boldsymbol{b}$. Column (1) shows results from a baseline specification that includes year and regency of residency fixed-effects only. The coefficient of 0.37 indicates that birthplace is highly predictive of women's employment. To see how large this coefficient is, let us consider two women: Putri and Amanda. Putri was born in the city of Probolinggo in East Java, which has a female employment rate of 40%. In contrast, Amanda was born in the regency of Sukoharto in Central Java, with a female employment rate of 62%. This gap places these two regencies at approximately the 75th and the 25th percentiles of the distribution of female employment rates. The 0.37 coefficient implies that Putri is 8 percentage points less likely to work than Amanda. This is a difference of 15% relative to the employment rate of the average Indonesian woman.

The additional estimates in table 4 also allow me to rule out several potential drivers of the birthplace persistence. Columns (2) and (3) show that controlling for women's age and religion barely modifies the estimate. Thus, this persistence is not explained by geographic differences in age or religion at the birthplace. Column (4) adds education level as a control. Recent research suggests that exposure to low-employment places can affect women's labor supply through the expectations and education channel (Molina and Usui, 2022). In areas with low female employment rates, women set low labor market expectations and thus invest less in skills. However, there is little support for this in Indonesia because controlling for education barely moves the estimate.

Columns (5) to (8) rule out childhood socioeconomic status and maternal labor supply as drivers of my results. In columns (5) and (8), I study the role of childhood economic conditions. The birthplace persistence could be capturing differences in socioeconomic status (SES) across labor markets. However, I can rule out this possibility using IFLS data on self-reported childhood socioeconomic information. They come from a battery of questions where respondents reported information on their household when they were 12 years old. These include wealth and education proxies such as the number of books, the number of people per room, and whether their father was in formal employment, among others. Adding these additional controls has little effect on the childhood persistence estimate.

In addition, in columns (7) and (8), I rule out the possibility that the birthplace persistence is driven by differences in maternal labor supply. Previous literature shows that women with working mothers are more likely to work (Fernandez and Fogli, 2009). Therefore, the birthplace persistence might just be reflecting the fact that in places where more women work, there are higher shares of working mothers. Because of the panel nature of the IFLS, I can compute the maternal labor supply for a subset of women in my sample. Column (7) re-estimates the birthplace persistence for this sample. Column (8) shows the persistence estimate when I control for maternal labor supply. Although the point estimate is slightly smaller, I can rule out that maternal labor supply drives my results.

Table 5: Indonesia: estimates birthplace persistence on men's labor supply (***b***)

	(1)	(2)	(3)	(4)
Women's employment rate at birthplace (p_o)	0.01	0.04	0.05*	0.04
	(0.03)	(0.03)	(0.03)	(0.03)
Mean employment rate	0.90	0.90	0.90	0.90
Implied IQR gap	0.00	0.01	0.01	0.01
Year FE	✓	✓	✓	✓
Regency FE	✓	✓	✓	✓
Age		✓	✓	✓
Religion			✓	✓
Education				✓
Observations	60,126	60,126	60,126	60,126
N individuals	6,293	6,293	6,293	6,293
R^2	0.05	0.17	0.17	0.18

Notes: Uses data from IFLS. Sample restricted to people residing outside their birthplace. Implied IQR gap shows the implied employment gap between someone born at a regency at the 75th percentile of employment rate and someone born at the 25th percentile. The IQR of the female employment rate across regencies is of 22 percentage points. Standard errors clustered by regency of origin. When indicated, the regressions control for a quadratic polynomial in age, and fixed-effects for seven religion and for education categories. Standard errors clustered by regency of origin.

Moreover, table 5 shows that birthplace persistence is specific to women. In this table I show estimates from regressions where I relate men's employment in adulthood to their birthplace's *female employment rate*. Note that all these estimates are close to zero and are precise enough to rule out any considerable persistence for men. For example, the estimate in column (8) implies an IQR gap of just one p.p. Given the very large employment rates for men, these differences are minimal.

4.3 There is large persistence for those who migrated young

The strong persistence of birthplace in women's labor supply could reflect the fact that working and non-working women have different migration patterns. If working women migrate only when a job is likely at the destination, even conditioning in the current location, women from high-employment regencies would be more likely to work than their low-employment counterparts. This correlation would arise even in the absence of any birthplace causal effect.

Table 6: Indonesia: estimates of birthplace persistence on labor supply for early migrants ($\boldsymbol{b}$)

	Women		**Men**	
	Baseline	**Young**	**Baseline**	**Young**
	(1)	(2)	(3)	(4)
Women's employment rate at birthplace (p_o)	0.36***	0.39***	0.04	0.08*
	(0.04)	(0.06)	(0.03)	(0.04)
Mean	0.54	0.52	0.90	0.86
Implied IQR gap	0.08	0.09	0.01	0.02
Age at emigration	All	< 19	All	< 19
Observations	64,727	27,977	60,119	23,016
N individuals	6,133	2,629	6,291	2,389
N movers	6,133	2,629	6,291	2,389
R^2	0.14	0.16	0.18	0.25

Notes: Uses data from IFLS. Columns (2) and (4) restrict the sample to people who left their birthplace before they turned 19. Implied IQR gap shows the implied employment gap between someone born at a regency at the 75th percentile of employment rate and someone born at the 25th percentile.. All regressions controls for year, regency of residency, religion, and education FE, and a quadratic polynomial on age. Standard errors clustered by the regency of birth.

If my results were driven by unobservable differences between working and non-working women, then there should be little birthplace persistence for women who emigrated before they entered the labor market. I test this by restricting the sample to women who emigrated before they turned 19. These women represent 42% of my sample and, as figure D.1 illustrates, they were unlikely to be working at the time they emigrated.

In column (2) in table 6, I show the birthplace persistence estimates for these emigres. For comparison, the table also shows estimates for the full sample –Baseline–. Even though these women were unlikely to be working at the time they emigrated, I obtain similar estimates to those of the full sample. Thus unobservable differences between working and non-working women are unlikely to be driving these results. Table 6 also shows estimates of a similar exercise for men. Again, there is little evidence of persistence for men.

4.4 The birthplace persistence is stronger the longer you stay

The strong birthplace persistence in women's employment could still reflect unobservable differences between women from different origins. Here, I address this concern by exploiting differences in the timing of migration to argue that this persistence reflects the causal effect of women's birthplace. To do so, I use a Difference-in-Differences strategy and augment expression (7) by (i) allowing the coefficient on female employment rate to vary by the emigration age ($\boldsymbol{b_a}$), and (ii) adding age of emigration fixed-effects (λ_a).

As I discuss in section 4.1, I can decompose the OLS estimates of age specific-slopes into a cumulative causal effect σ_a, and a selection term γ:

$$\boldsymbol{b}_a = \boldsymbol{\sigma}_a + \boldsymbol{\gamma}$$

Under the assumption that omitted variable bias is constant across emigration ages, I can identify the causal effect of place at any given age (π_a) by subtracting the persistence coefficients across emigration ages:

$$\boldsymbol{\pi}_a = \boldsymbol{b}_{a+1} - \boldsymbol{b}_a$$

Moreover, the coefficient for least exposed cohort gives as estimate of the omitted variable bias: $\gamma = b_0$

To estimate this model, I leverage age of emigration data from the IFLS and the Intercensal survey. The structure of the IFLS questionnaire and its sample size imposes several restrictions. First, for all migrants who emigrated before they turned 12, I do not know the exact age at which they left their birthplace. The IFLS only tracks the age at migration for episodes after 12 years old. Second, this IFLS sample size forces me to group emigration age into bins. Figure C.6 shows that the number of regencies is large relative to the number of women emigrating at any given age. Because the regression identifies the age slopes out of comparisons between women who (i) live in the same regency but who (ii) left their birthplace at different ages, I must ensure that, within each regency, I have women who emigrated at different ages[13]. Therefore, in the estimation, I group women into five age of emigration brackets: 11 or less, 12-14, 15-16, 17, and 18 years old. Appendix figure C.7 shows that this grouping roughly balances the number of women migrants across the age bins.

Longer stay does make you more likely to work

Table 7 shows estimates of the birthplace persistence by the age of emigration $\boldsymbol{b_a}$. My sample remains restricted to people who left their birthplace before they turned 19. All regressions in the table control for education and religion fixed-effects, along with a quadratic polynomial on age. In column (1), I reproduce the baseline estimates for these women, while Column (2) allows the coefficient on birthplace female employment to vary by age bracket.

[13] On average, only 185 women emigrated in each age between 12 and 16 years old. Because there are 268 regencies, in many places, there will be little within-regency variation to identify different coefficients for each age cell if I do not bin the emigration ages

Table 7: Indonesia: estimates of birthplace persistence ($\boldsymbol{b_a}$) by age of emigration

	Women		**Men**	
	(1)	(2)	(3)	(4)
p_b	0.387		0.082	
	(0.058)		(0.042)	
Age of emigration interactions				
11- × p_b		0.087		0.109
		(0.113)		(0.082)
12-14 × p_b		0.217		0.103
		(0.145)		(0.065)
15-16 × p_b		0.533		0.047
		(0.112)		(0.078)
17 × p_b		0.572		0.201
		(0.126)		(0.101)
18 × p_b		0.545		-0.042
		(0.134)		(0.076)
Year FE	✓	✓	✓	✓
Current regency FE	✓	✓	✓	✓
Age of migration FE		✓		✓
Observations	27,977	27,977	23,014	23,014
No. individuals	2,629	2,629	2,389	2,389
No. migrants	2,629	2,629	2,389	2,389
r2	0.16	0.16	0.25	0.25

Notes: Uses data from IFLS. Table restricts the sample to people residing outside their birthplace with known age of emigration. All regressions controls for religion, and education FE; and a quadratic polynomial on age.

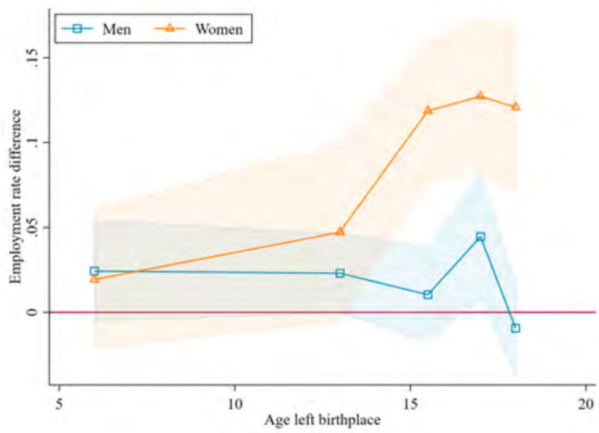
These results show a striking pattern in the birthplace coefficients. The coefficients progressively increase from 0.09 p.p. for women emigrating before 12 to approximately 0.53 for those emigrating between 15 and 16 years old and stay roughly constant after that. This pattern points to late childhood and early teens as the key ages where birthplace matters. Under the constant omitted variable bias, the 11 or less slope approximates the size of omitted variable bias. This is the cohort least exposed to the birthplace. On average, these women stayed only six years in their birthplace[14]. Thus, we can attribute any systematic differences between women from different origins to factors I do not control for. For subsequent age brackets, I attribute the increase relative to the nine p.p. baseline as coming from the birthplace causal effect. Consequently, the increasing slope pattern points to a positive causal effect of women's birthplace on their adult labor supply

[14]According to data from the Intercensal Survey, people emigrate at roughly constant rates at each age before 12. Therefore, the average woman in this bracket left their birthplace when she was six years old

during childhood and adolescence. Staying after turning 16 has little additional effect.

Moreover, the birthplace effects arise for women only. In column (4), I show birthplace persistence estimates for men. These are interactions between the emigration age and the *birthplace female labor force participation*. These estimates are imprecise, primarily flat, and do not display women's slopes' consistent pattern. If anything, these estimates suggest that men's slopes are driven mainly by omitted variable bias.

Figure 3: Indonesia: implied gap IQR in employment rate by age of emigration



Note: The figure shows estimates of $\boldsymbol{b_a}\times$ regency-level IQR in female employment (22 p.p.). This is the implied gap in employment at different ages of out-migration between a person born in a regency at the 75th percentile in female employment, and another one born in a regency at the 25th percentile of female employment. Series for women use $\boldsymbol{\sigma}_a$ estimates in column (2) of table 7. Series for men use $\boldsymbol{\sigma}_a$ estimates from column (4) of table 7. Point estimates are placed at the mean point of the respective age interval. Shaded areas show 90% confidence intervals. The figure uses data from IFLS.

My results imply that place effects are important drivers of the geographic differences in women's labor supply. I illustrate this in figure 3, where I show the counterfactual gaps in the employment between two women, one born in a regency at 75th percentile of employment and the other at the 25th percentile[15], if they left their birthplace at different ages. The figures places the gap estimates at the midpoint of each of the age brackets in table 7. If both of these women had emigrated when they were six years old, I would observe a gap of 2 p.p. in their labor supply mostly driven by unobservable differences between these two women. In contrast, if they stayed in their birthplace up to 13 and 15 years old, this gap would widen to 5 and 12 p.p. respectively. This widening is driven by the birthplace causal effect. In all, birthplace causes a gap of 10 p.p. in the likelihood of employment of these two women. Note that the existing gap of 22 p.p. in FLFP between the birthplaces of these women translated into a gap of 10 p.p. for them, which implies that approximately 45% of the existing inequality in female labor supply is transmitted to the next generation of women growing up in these locations.

[15]These two regencies have a gap of 22 p.p. in the female employment rate.

The data supports the constant selection assumption

The causal interpretation of the birthplace persistence coefficients hinges on the strong assumption that selection is independent of emigration age. More precisely, conditioning on the current location and other controls, I require that the relationship between women's unobserved characteristics and the birthplace female employment rate be the same for women who emigrated at different ages. Below, I provide results showing that selection along several observable dimensions is constant across emigration age. This suggests that the identification assumption is likely to hold in my data.

One can think of the identification assumption as an analog of the parallel trends in Difference in Differences. I expect women coming from high and low-employment regencies to be different from each other. This is not an issue. However, even in the absence of a birthplace causal effect, if there are factors correlated with female employment that change differently across emigration ages for these two groups of women, I would mistakenly assign this variation to the causal effect. In other words, the lack of parallel trends would lead me to find a causal effect where there is none.

Figure 4 shows that women emigrating at different ages display similar selection patterns on (i) reasons for emigrating, (ii) proxies of parental wealth, (iii) number of siblings, and (iv) the characteristics of the destination regency. This figure shows estimates of the slopes on the interaction between age of emigration and birthplace FLFP β_a in regression of the form:

$$y_i = \lambda_a + \beta p_b + \sum_a \beta_a 1_a \times p_b + X_i\kappa + \varepsilon_{it} \tag{8}$$

where y_i is a woman characteristic, and 1_a is an age of emigration dummy. For all regressions, I chose the latest age –18– as the base category, thus all the β_a are readily interpreted as the slope difference relative to women who emigrated at 18. *Under constant selection all the interactions must be zero.*

Panel (a) shows estimates where I use indicators of the reason for emigrating as outcomes. This information is available in the IFLS only for people who emigrated at 12 years old or older, thus the figure does not show estimates for the youngest emigre cohort. Although imprecise, these estimates suggest constant selection by emigration motive. For all motives, I cannot reject that all interactions are jointly zero[16].

Panels (b) and (c) of figure 4 show estimates for proxies of parental wealth and measures of family size. IFLS respondents provide retrospective information about their household during their

[16] See point estimates and tests of hypothesis in table D.10 in the appendix.

childhood. In panel (b), I use as outcomes whether her father had formal employment, an indicator of having more than 11 books at home, and the number of people per room in the household[17]. In all cases, I cannot reject that all the interactions are jointly zero. Panel (c) performs a similar exercise with the number of siblings. Previous research argues that the number of siblings is an important determinant of parental investment's in women's education. Although there seems to be a difference in the number of brothers for 17 year old emigres, reassuringly I cannot reject that all estimates are jointly zero at conventional significance levels.

Finally, in panel (c), I explore whether women's destination changes by emigration age. Estimates in this panel come from regressions where I use destination regency characteristics as outcomes. Because my birthplace persistence regressions control for the current location fixed-effects, selection due to the destination is less of a concern. It is reassuring, however, that I cannot reject that all the interactions are jointly zero. Thus, I do not see evidence that emigres' age changes the destination they choose.

5 Robustness

In this section I show that my results survive multiple robustness checks. I start by showing that I obtain similar conclusions when exploiting emigration age data from the Intercensal Survey. Next, I show my results are robust to (i) my choice of dependent variable, and (ii) the year from which I source the female employment rates.

5.1 I obtain similar results in alternative datasets

While very rich in information, the IFLS has two main limitations for the estimation strategy: (i) relatively small sample size, (ii) and lack of age information for emigres who left before they turn 12 years old. In figure 5, I take advantage of data from the 1985, 1995, and 2005 Indonesian Intercensal Surveys. This survey is similar to the census. It contains a more limited array of information than the IFLS. However, it has a larger sample size, and it contains information from which I can extract the exact age of emigration for all migrants. This additional information allows me to estimate place effects for earlier and smaller age brackets.

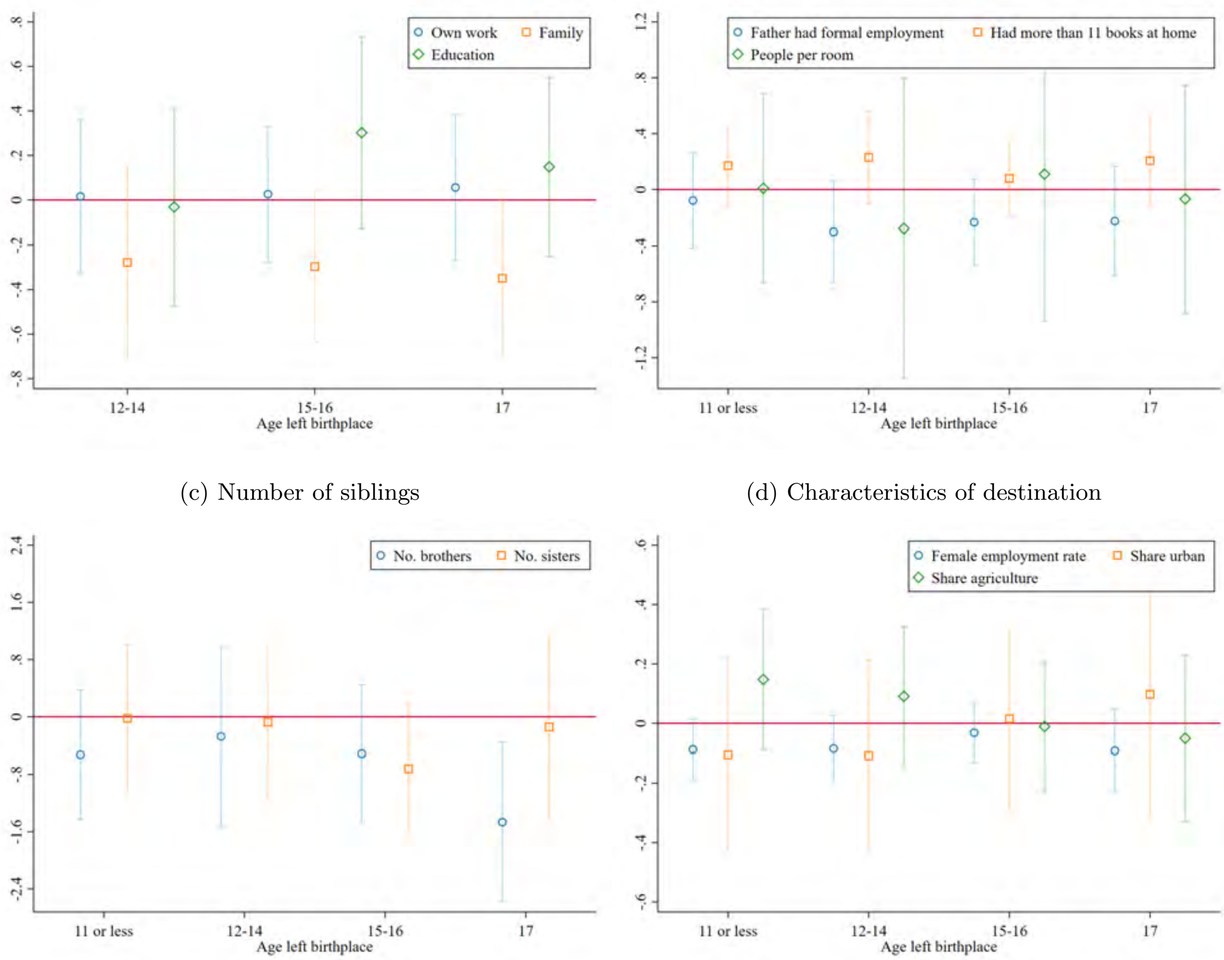
Panel A in figure 5 shows birthplace persistence estimates using the Intercensal Survey. Reassuringly, the results are similar in this alternative dataset. Women's birthplace persistence increases in the age of emigration and then levels off in their late teens. In contrast, men's persistence is mostly flat. Both patterns are consistent with a positive causal effect of birthplace that is exclusive

[17]Only 21% of the women in this sample declare having 11 or more books when they were 12 years old. I define a respondent's parent as formally if he was employed as a private or government worker. Only 26% of the women in this sample report having a formally employed father when they were 12 years old.

Figure 4: Indonesia: women and selection by age of emigration in the IFLS

(a) Reason of migration (b) Wealth proxies

(c) Number of siblings (d) Characteristics of destination



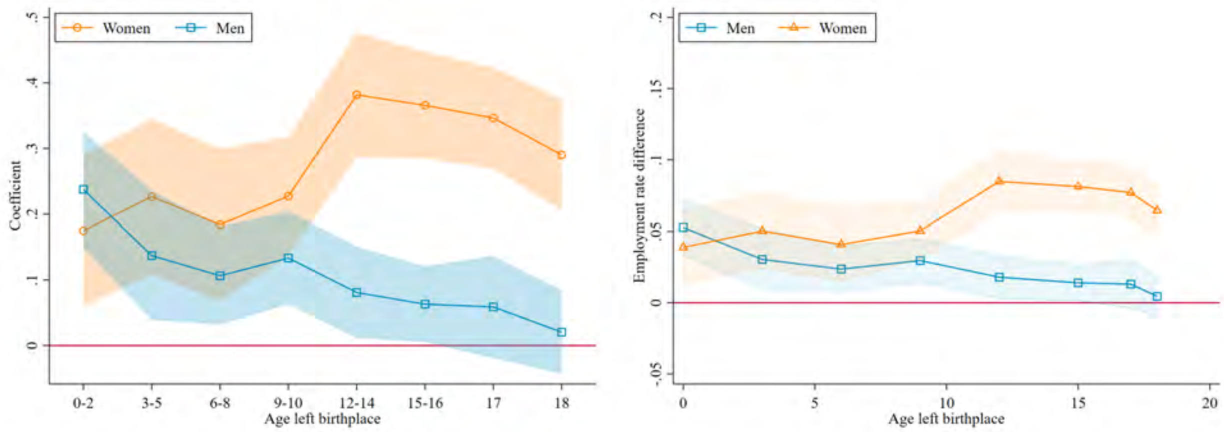
Note: The figure show the coefficients on the interactions between age of emigration and birthplace female labor force participation. The regressions also control for (i) age of emigration dummies, (ii) birthplace female employment rate (iii) interactions between the emigration age and birthplace female employment rate. I chose 18 years old as the base category so that the interactions test whether the slopes at each age are different from that for women emigrating at 18. Data on reasons for emigrating is available only for people emigrating at 12 years old or older. Error clustered by regency of birth. The figure shows 90% confidence intervals. Data from the IFLS.

to women.

Nevertheless, there are two important differences relative to the IFLS results. Here, (i) selection is more important and thus I obtain smaller causal effects, and (ii) the causal effects level off slightly earlier. If we take the earliest point estimate figure 5 as the selection term, it is around 0.17. Although with a large standard error, this is about twice the size implied by the IFLS. The birthplace persistence goes up between 6 and 14 years old and then levels off at around 39 p.p. This implies smaller but still sizable birthplace effects. Panel (B) in the figure shows that the

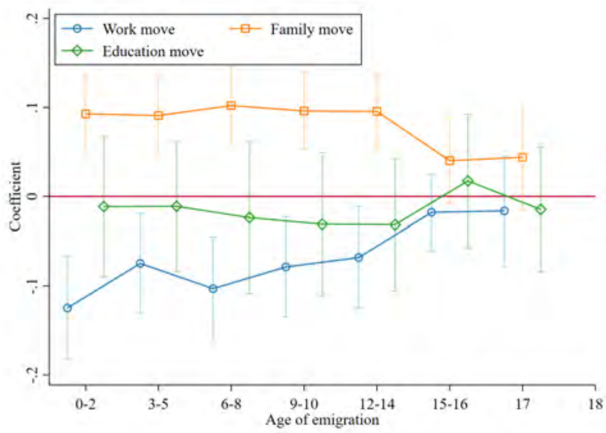
Figure 5: Indonesia: birthplace

(a) Regression coefficients (b) Implied IQR gap



Note: Panel (a) shows estimates of the birthplace persistence coefficients by age of emigration $\boldsymbol{b_a}$. Panel (b) shows estimates of $\boldsymbol{b_a} \times$ regency-level IQR in female employment (22 p.p.). This is the implied gap in employment at different emigration ages between a person born in a regency at the 75th percentile in female employment, and another one born in a regency at the 25th percentile of female employment. The regression controls for a quadratic polynomial in age, education FE, and current regency FE. Shaded areas show 90% confidence intervals. The figure uses data from the 1985, 1995, 2005 Incercensal Surveys.

Figure 6: Indonesia: estimates of birthplace slopes by age of emigration ($\boldsymbol{k_a}$)



Note: The figure shows estimates of the interactions between age of emigration dummies and the birthplace female employment rates. I chose 18 years old as the base category so that the interactions test whether the slopes at each age are different from that for women emigrating at 18. Sample restricted to women residing outside their birthplace. All regressions controls for year by bithplace regency fixed-effects, age of emigration fixed-effects, education fixed effects, and a quadratic polynomial on age. Standard errors clustered by regency of origin. Uses data from 1985, 1995, and 2005 Intercensal Surveys available in IPUMS international.

cumulative effect for two women on opposite sides of the IQR in female employment at 14 years old is of 4.6 p.p. This is 9% of the average female employment rate and it implies that 21% of the current spatial inequality is transmitted to the next generation of women.

The Intercensal survey also provides additional data to test the validity of the identification assumption. A potential concern with the tests in figure 4 is that these estimates are imprecise. Thus, I may not have the power to detect changes in selection with the IFLS sample sizes. To alleviate this concern, in figure 6, I show interaction estimates coming from the regressions where I use the emigration motive as dependent variable. This figure uses the larger samples from the Intercensal Survey. This figure suggests selection by emigration varies in late teens. Seventeen and eighteen year old women are more likely to be working than those at younger ages, so this change in selection is not surprising. However, the slope jumps at 15-16 years old are more concerning for the identification assumption. Note, however, that there is little evidence of changing selection between the ages of 6 and 14 years old. In fact, I cannot reject they are jointly equal. These are exactly the ages for which the Intercensal Survey estimates in figure 5 yield sizable place effects. Thus, in all, figure 6 supports the causal interpretation of the birthplace persistence estimates.

5.2 Results are similar for alternative measures of labor supply

All my main results use being employed as my main measure of labor supply. However, table 8 shows that my main results carry through when I use alternative measures of labor supply. Column (1) shows my baseline estimates. In column (2), I use being a paid worker as an outcome. This adjustment is important because 23% of women are unpaid workers. It is thus reassuring that similar persistence patterns arise under this alternative measure. In columns (3) and (4), I use total weekly hours worked, and a dummy of being a full-time worker. I define full-time as working more than 35 hours per week. Because weekly hours information is not available in waves 4 and 5 of the IFLS, the samples for these two outcomes are substantially reduced. While admittedly less clear, the general persistence patterns are similar for these two outcomes: spending more time in high employment regencies increases women's labor supply in adulthood.

Table 8: Indonesia: birthplace persistence effects for alternative measures of labor supply

	Employed	**Paid worker**	**Weekly hours**	**Full time**
	(1)	(2)	(3)	(4)
Age of emigration interactions				
11- × p_b	0.09	-0.00	13.96	0.19
	(0.11)	(0.11)	(11.04)	(0.20)
12 × p_b	0.21	-0.01	15.03	0.28
	(0.15)	(0.15)	(10.94)	(0.21)
15 × p_b	0.53***	0.41***	10.84	0.14
	(0.11)	(0.11)	(8.93)	(0.17)
17 × p_b	0.57***	0.42***	20.84**	0.35*
	(0.13)	(0.14)	(8.44)	(0.19)
18-19 × p_b	0.54***	0.59***	31.97***	0.41**
	(0.13)	(0.12)	(9.34)	(0.18)
Observations	27,977	27,977	8,599	8,599
No. individuals	2,629	2,629	1,156	1,156
No. migrants	2,629	2,629	1,156	1,156
r2	0.16	0.13	0.17	0.15

Notes: I define full-time work as working more than 35 hours per week. Weekly hours data is not available for waves 4 and 5 of the IFLS. This substantially reduces the sample in columns (4) and (5). All regressions control for religion, and education FE; and a quadratic polynomial on age. Data from the ILFS.

6 Conclusions

In this paper, I document large and persistent spatial inequality in women's labor supply in Indonesia, a country with more than 118 million women. I argue that a substantial portion of this inequality is driven by the local environment women are born into. To identify the causal effect of place, I leveraged variation coming from the age women emigrated from their birthplace. I compared the labor supply choices of women who currently live in the same location, but who emigrated from their birthplace at different ages as children. If the omitted variable bias is independent of the age of emigration, this strategy allows me to distinguish the causal effect of place from variation driven by differences in women's unobserved characteristics.

I show that women's birthplace is particularly important during the formative childhood and teen years. Staying in a location at the 75th percentile of female employment between 6 and 16 years of age makes women between 4 to 10 percentage points more likely to work than those born in a location at the 25th percentile. These magnitudes mean that between 21 and 45 percent of the current spatial inequality in women's employment transmits to the next generation of women. Therefore, these women-specific place effects can be an important driver of the large and persistent differences in women's labor force participation within countries.

An important question left for subsequent work is what drives these place effects. The effects I uncover can be the consequence of continual exposure to gender norms less amenable to women's work (Boelmann et al., 2021). They can also be result of heterogeneity in schooling quality. Future research should focus on distilling the mechanisms behind these large geographic disparities. The form of effective policies aiming to encourage women's inclusion on the labor market depends crucially on what the main mechanism behind these results is.

References

Alesina, A., Giuliano, P., and Nunn, N. (2013). On the origins of gender roles: Women and the plough. *Quarterly Journal of Economics*, 128(2):469–530. 4

Angelov, N., Johansson, P., and Lindahl, E. (2016). Parenthood and the gender gap in pay. *Journal of Labor Economics*, 34(3):545–579. 13

Angrist, J. D. and Pischke, J.-S. (2009). *Mostly harmless econometrics : an empiricist's companion*. Princeton University Press. 16

Autor, D. H. and Dorn, D. (2013). The Growth of Low-Skill Service Jobs and the Polarization of the US Labor Market. *American Economic Review*, 103(5):1553–1597. 2, 11, 37

Bazzi, S., Hilmy, M., Marx, B., and Po, S. (2022). Religion, Education, and Development. 5

Black, D. A., Kolesnikova, N., and Taylor, L. J. (2014). Why do so few women work in New York (and so many in Minneapolis)α Labor supply of married women across US cities. *Journal of Urban Economics*, 79:59–71. 5

Blau, F. D. and Kahn, L. M. (2015). Immigration and the distribution of incomes. In *Handbook of the Economics of International Migration*, volume 1. 5

Blau, F. D., Kahn, L. M., Comey, M., Eng, A., Meyerhofer, P., and Willén, A. (2020). Culture and gender allocation of tasks: source country characteristics and the division of non-market work among US immigrants. *Review of Economics of the Household*, 18(4):907–958. 5

Boelmann, B., Raute, A., and Schonberg, U. (2021). Wind of Change? Cultural Determinants of Maternal Labor Supply. *SSRN Electronic Journal*. 4, 15, 33

Brinkhoff, T. (2022). Indonesia: Administrative Division (Provinces, Regencies and Cities), retrieved from https://www.citypopulation.de/en/indonesia/admin/. 12

Bryan, G. and Morten, M. (2019). The Aggregate Productivity Effects of Internal Migration: Evidence from Indonesia. *Journal of Political Economy*, 127(5):2229–2268. 4, 6, 8, 48

Charles, K. K., Guryan, J., and Pan, J. (2018). The Effects of Sexism on American Women: The Role of Norms vs. Discrimination. *SSRN Electronic Journal*. 2, 4, 5, 15

Chetty, R. and Hendren, N. (2018a). The impacts of neighborhoods on intergenerational mobility I: Childhood exposure effects. *Quarterly Journal of Economics*, 133(3):1107–1162. 3, 4, 15, 17, 18, 40

Chetty, R. and Hendren, N. (2018b). The impacts of neighborhoods on intergenerational mobility II: County-level estimates. *Quarterly Journal of Economics*, 133(3):1163–1228. 4

Chetty, R., Hendren, N., Jones, M. R., and Porter, S. R. (2020). Race and economic opportunity in the United States: An intergenerational perspective. 4

Chetty, R., Hendren, N., and Katz, L. F. (2016). The effects of exposure to better neighborhoods on children: New evidence from the moving to opportunity experiment. *American Economic Review*, 106(4):855–902. 4

Compton, J. and Pollak, R. A. (2014). Family proximity, childcare, and women's labor force attachment. *Journal of Urban Economics*, 79:72–90. 15

Damm, A. P. and Dustmann, C. (2014). Does growing up in a high crime neighborhood affect youth criminal behavior? *American Economic Review*, 104(6):1806–1832. 4

Farre, L. and Ortega, F. (2021). Family Ties, Geographic Mobility and the Gender Gap in Academic Aspirations. *SSRN Electronic Journal*. 15

Fernández, R. (2007). Alfred Marshall lecture women, work, and culture. In *Journal of the European Economic Association*, volume 5, pages 305–332. 16, 17

Fernández, R. and Fogli, A. (2006). Fertility: The role of culture and family experience. In *Journal of the European Economic Association*, volume 4, pages 552–561. 16

Fernandez, R. and Fogli, A. (2009). Culture: An empirical investigation of beliefs, work, and fertility. *American Economic Journal: Macroeconomics*, 1(1):146–177. 18, 21

Fogli, A. and Veldkamp, L. (2011). Nature or Nurture? Learning and the Geography of Female Labor Force Participation. *Econometrica*, 79(4):1103–1138.

Hamory, J., Kleemans, M., Li, N. Y., and Miguel, E. (2021). Reevaluating Agricultural Productivity Gaps with Longitudinal Microdata. *Journal of the European Economic Association*, 19(3):1522–1555. 5

International Labour Organization (2021). ILOSTAT database: Country profiles. 37

Kleven, H., Landais, C., and Søgaard, J. E. (2019). Children and gender inequality: Evidence from Denmark. *American Economic Journal: Applied Economics*, 11(4):181–209. 13

Le Barbanchon, T., Rathelot, R., and Roulet, A. (2021). Gender Differences in Job Search: Trading off Commute against Wage. 5, 15

Magruder, J. R. (2013). Can minimum wages cause a big push? Evidence from Indonesia. *Journal of Development Economics*, 100(1):48–62. 5

Milsom, L. (2021). The Changing Spatial Inequality of Opportunity in West Africa. 4, 5

Minnesota Population Center (2020). *Integrated Public Use Microdata Series, International: Version 7.3 [dataset]*. MN:IPUMS. 4, 6, 7

Molina, T. and Usui, E. (2022). Female Labor Market Opportunities and Gender Gaps in Aspirations. *SSRN Electronic Journal*. 4, 5, 15, 21

Moreno-Maldonado, A. (2019). Mums and the City Female labour force participation and city size. 5

Olivetti, C. and Petrongolo, B. (2008). Unequal pay or unequal employment? A cross-country analysis of gender gaps. *Journal of Labor Economics*, 26(4):621–654. 5

Olivetti, C. and Petrongolo, B. (2014). Gender gaps across countries and skills: Demand, supply and the industry structure. *Review of Economic Dynamics*, 17(4):842–859. 5, 15

Olivetti, C. and Petrongolo, B. (2016). The Evolution of Gender Gaps in Industrialized Countries. *Annual Review of Economics*, 8(1):405–434. 13

Statistics Indonesia (2019). *Survei Sosial Ekonomi Nasional (Susenas), 2012 Core, https://doi.org/10.7910/DVN/12TVW1*. Harvard Dataverse, V1. 6

Statistics Indonesia (2020). *Survei Sosial Ekonomi Nasional (Susenas), 2016 Kor, https://doi.org/10.7910/DVN/BOXZWU*. Harvard Dataverse, V2. 6

Statistics Indonesia (2021). The Indonesian Population Census. 4

A Data appendix

A.1 Cross-country data

I use harmonized data from IPUMS International to build figure 1 from the introduction and table 1 from section 3. They show local employment rates for men and women aged 18-64 for a subset of Asian countries and the United States. For all these countries, I use the latest decennial census sample available. This corresponds to 2010 or a year close to it, except in Malaysia, Thailand, and China where the latest censuses are from 2000. In all cases, I restrict the samples to people aged 18-64.

I define employment using the harmonized employment status (`empstat`). However, this variable is not available in Thailand, Philippines, and Vietnam. In Thailand and Philippines, I say a person is employed if they report being self-employed, a salary worker, or an unpaid worker (`classwkr`). In China, employed workers are those who reported working at least 1 day in the past week. Despite these slight definition differences, table A.1 shows that the employment rates I obtain are in line with the female labor force participation rates reported by the International Labor Organization and the World Bank (International Labour Organization, 2021)[18]. The differences in the age ranges I consider drive the discrepancies for the United States, Vietnam, Thailand and China.

For all countries, I compute subnational employment rates at the lowest geographic unit available. For most countries, this corresponds to a geographic area akin to a district, a country, or a municipality. The only exception is the United States, where I compute these rates by commuting zone (Autor and Dorn, 2013). I winsorize the employment rates at the 5th and 95th percentiles by country. This reduces the possibility that very small regions drive the dispersion I observe within countries.

[18]The only exception is the Philippines, where the data from IPUMS International implies much lower employment rates. In my data, I obtain a female employment rate of 33% for women aged 18-64. The ILOSTAT database reports a female labor force participation rate of 48% for 15+ women in 2010. The gap between these two figures cannot be accounted for by female unemployment which is of the order of 4%. That said, I am interested in within-country dispersion, these discrepancies are second order as long data collection is consistent within the country.

Table A.1: Female labor force participation rages by country: IPUMS vs ILOSTAT

Country	**IPUMS (ages 18-64)**	**ILOSTAT (ages 15+)**	**Difference**
Cambodia	0.82	0.81	0.01
China	0.74	0.64	0.10
Indonesia	0.50	0.51	-0.01
Malaysia	0.43	0.43	-0.00
Myanmar	0.50	0.53	-0.03
Philippines	0.33	0.48	-0.15
Thailand	0.77	0.64	0.13
United States	0.67	0.58	0.10
Vietnam	0.79	0.72	0.07

Notes: Uses data from IPUMS international and ILOSTAT. I restrict the sample in IPUMS to people aged between 18-64 years old.

B The Empirical Strategy

When I introduce the age of emigration data in section 4.1, I assume that women's employment decisions are determined by place of residence fixed effects δ_c, age of emigration fixed effects λ_a, female labor force participation at birthplace p_b, and an error term η_i:

$$e_{it} = \delta_c + \lambda_a + \boldsymbol{\sigma}_a p_b + \eta_{it} \tag{9}$$

The error term embodies factors that are potentially important in determining women's decision to work, but which I do not observe. These factors could be correlated with the woman's birthplace employment rate. For example, generally I do not observe whether a woman's mother worked. This variable is naturally correlated with the birthplace female employment rate.

To simplify the discussion, I write this model in its matrix form as follows:

$$E = F\boldsymbol{\omega} + P\boldsymbol{\sigma} + \boldsymbol{\eta}$$

here F is a matrix containing place of residence and age of emigration indicators, P contains interaction between the age of emigration fixed effects and FLFP at birthplace, $\boldsymbol{\omega}$ stacks the location and age of emigration fixed-effects, $\boldsymbol{\sigma}$ is a vector containing the age of emigration effects $\boldsymbol{\sigma}_a$, and η is error term vector.

My main interest is estimating the birthplace effects vector $\boldsymbol{\sigma}$ consistently. For simplicity I can express the model in terms of the birthplace effects and the unobserved components by residualizing it from the age and residency fixed-effects. Let $\tilde{Z} = I - F(F'F)^{-1}F'$. Then,

$$\tilde{E} \quad = \quad \widetilde{P}\boldsymbol{\sigma} + \tilde{\boldsymbol{\eta}}$$

now let us consider the OLS estimate of the birthplace effects $\hat{\boldsymbol{\sigma}} = (\widetilde{P}'\widetilde{P})^{-1}\widetilde{P}'E$. Note that from the above expression it follows that,

$$\hat{\boldsymbol{\sigma}} = \boldsymbol{\sigma} + (\widetilde{P}'\widetilde{P})^{-1}\widetilde{P}'\tilde{\boldsymbol{\eta}}$$

Therefore:

$$\begin{aligned} \text{plim } (\hat{\sigma}) &= \boldsymbol{\sigma} + \text{plim } \left[(\widetilde{P}'\widetilde{P})^{-1}\widetilde{P}'\tilde{\boldsymbol{\mu}}\right] \\ &= \boldsymbol{\sigma} + \boldsymbol{\gamma} \end{aligned} \tag{10}$$

now let us examine the meaning of expression (10) in detail. This expression implies that the OLS estimate for the birthplace persistence at age a is the sum of two terms: (i) the birthplace effect effect at that age $\boldsymbol{\sigma}_a$, and (ii) a term that captures the correlation between the residualized employment rate at the origin and the residual $\boldsymbol{\gamma}_a$:

$$\boldsymbol{b}_a \quad = \quad \boldsymbol{\sigma}_a + \boldsymbol{\gamma}_a \tag{11}$$

B.1 Identification

Assumption 1 (constant selection):

Selection on unobservables does not depend on the age of emigration, that is $\boldsymbol{\gamma}_a = \boldsymbol{k}$

Assumption 1 essentially requires that correlation between women's –unobservable– characteristics and women's origin is the same no matter the age at which they move. To see this, note that γ is driven two components:

$$\boldsymbol{\gamma} = \text{plim } \left[(\widetilde{P}'\widetilde{P})^{-1}\right] \text{plim } \left[\widetilde{P}'\tilde{\boldsymbol{\eta}}\right]$$

The term plim $\left[\widetilde{P}'\tilde{\boldsymbol{\eta}}\right]$ captures the correlation between women's birthplace and the unobserved characteristics. This is transparent by examining the general term of the vector $\widetilde{P}'\tilde{\boldsymbol{\eta}}$:

$$\sum_{i=1}^{N} \widetilde{p_b^a}\tilde{\eta}_i \tag{12}$$

where p_b^a denotes the –residualized– interaction between the birthplace female employment rates and age of emigration dummies. By law of the large numbers, this element converges to:

$$\mathbb{E}\left(\widetilde{p_b^a}\tilde{\eta}_i\right)$$

Two sufficient but not necessary conditions for the constant selection to be satisfied are:

$$\mathbb{E}\left(\widetilde{p_b^a}\tilde{\eta}_i\right) = \boldsymbol{c} \tag{13}$$

$$plim(\widetilde{P}'\widetilde{P})^{-1} = Q \tag{14}$$

where Q non-singular matrix with: (i) diagonal elements equal to each other, and (ii) off-diagonal elements equal to each other.

The first condition requires the correlation women's unobserved characteristics and birthplace female labor force participation to be the same for women migrating at different ages as children. For instance, this condition allows for the fact that in places where more women work they were more likely to have working mothers. A violation of (14) would occur if, for example, women with working mothers stayed longer in their birthplace.

Moreover, note that equation (13) only contains variables residualized from current location and age of emigration fixed effects. It is likely that women of different ages migrate due to different motives. For example, women migrating at 10 years old would be less likely to migrate for school than 12 year old women because secondary in Indonesia starts at 13. However, this is not necessarily violates condition (13). This is only a problem if, after conditioning on current location and age of emigration fixed-effects, women migrating for family reasons came from different origins than women migrating for education. While I cannot fully test for this condition, I can provide supporting evidence by correlating birthplace female employment rate with observed women's characteristics for different emigration age cohorts.

Condition (14) imposes restrictions on the correlations between birthplace female employment for women migrating at different ages as children. It will be generally satisfied if women migrating at different ages came from roughly the same origins.

B.2 From OLS to causal effects

The constant selection assumption allows ne to identify the causal effects of spending more time at the birthplace. Identification follows the same intuition as in Chetty and Hendren (2018a). Because $\boldsymbol{\sigma}_a$ captures the birthplace effect accumulated up to age a, the effect of spending age a is just the difference across consecutive ages:

$$\boldsymbol{\pi}_a = \boldsymbol{\sigma}_a - \boldsymbol{\sigma}_{a-1}$$

under the constant selection assumption, equation (11) shows that simple subtraction of the OLS estimates identify the causal effects of place:

$$\boldsymbol{b}_a - \boldsymbol{b}_{a-1} = \boldsymbol{\sigma}_a - \boldsymbol{\sigma}_{a-1} \tag{15}$$

With an additional normalization, the OLS estimates can also identify the size of the selection term $\boldsymbol{\gamma}$. If we normalize the causal effect for the children with the least exposure to birthplace to zero ($\boldsymbol{\sigma}_0 = 0$), the OLS coefficient for this children is an estimate of the selection term:

$$\boldsymbol{b}_0 = \boldsymbol{c} \tag{16}$$

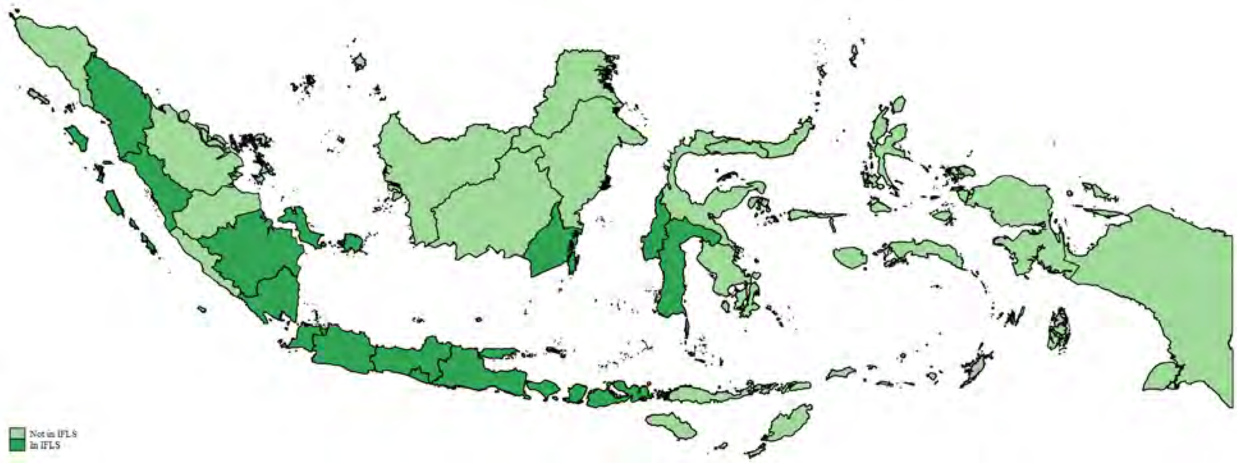
Equations (15) and (16) provide a full guide for estimating the causal effects. OLS estimates for women migrating at the earliest ages provide an estimate for the selection terms. Cross-age differences in the OLS estimates render the causal effect of spending a given age or period at the birthplace.

C Figures

Figure C.1: Indonesian regencies

Note: The figure shows the 268 regency aggregates with consistent boundaries between 1970 and 2018. Boundaries obtained from IPUMS International.

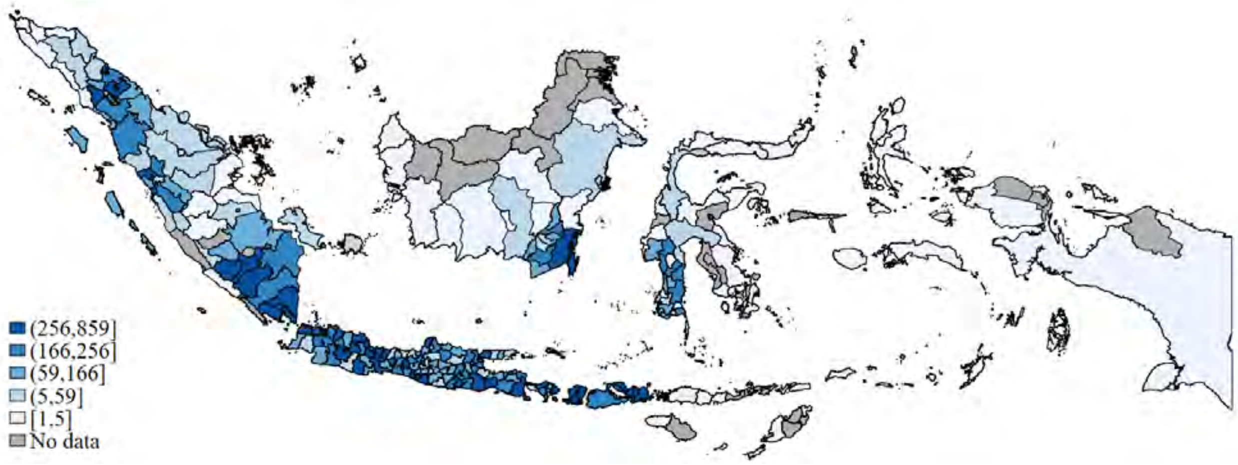
Figure C.2: Provinces in the original 1993 IFLS sample



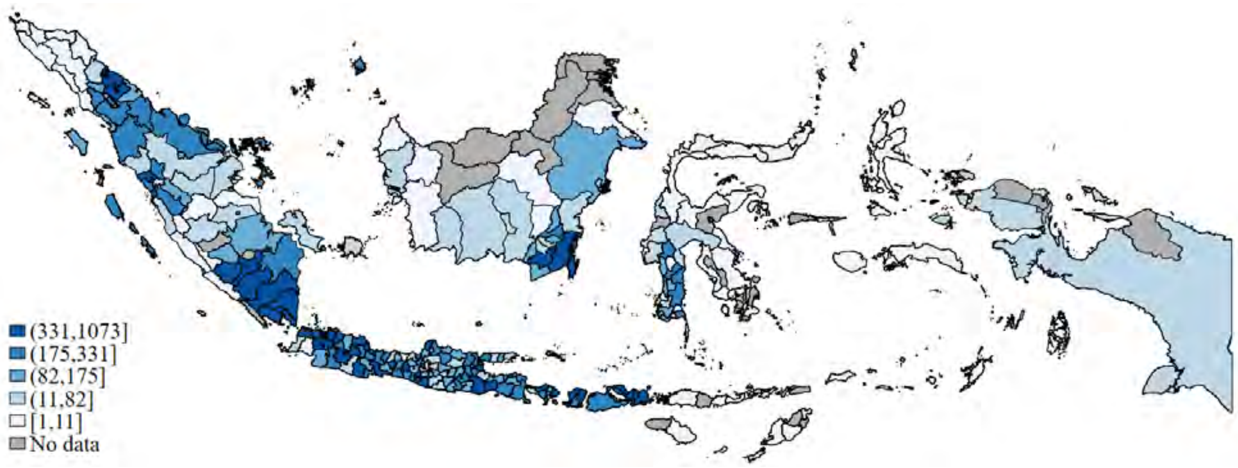
Note: The provinces from which the original 1993 IFLS sampled households. Because of migration, subsequent years can include individuals living outside these provinces. *Source:* RAND corporation.

Figure C.3: Indonesia: current and birthplace location of IFLS respondents

(a) Birthplace

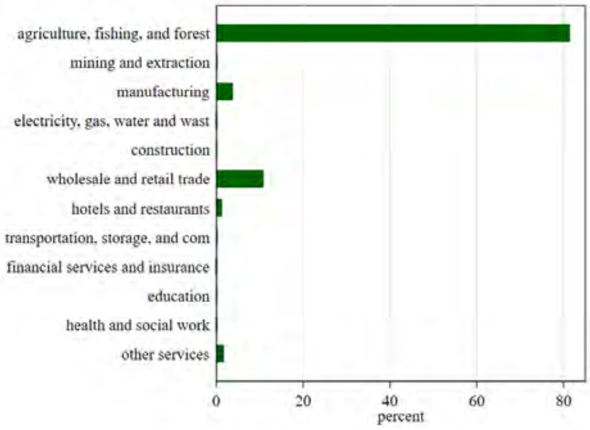


(b) Current location



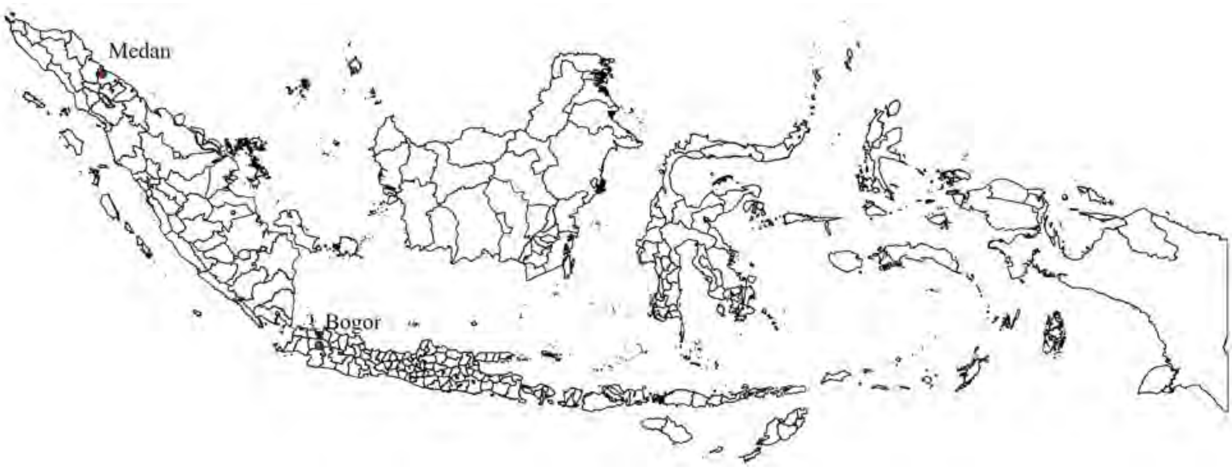
Note: The figure shows the counts of respondents by regency. It shows all the 268 regencies with consistent boundaries between 1970 and 2010. Each color groups a fifth of the regencies. The figure uses data from the IFLS.

Figure C.4: Indonesia: industry of employment of unpaid workers



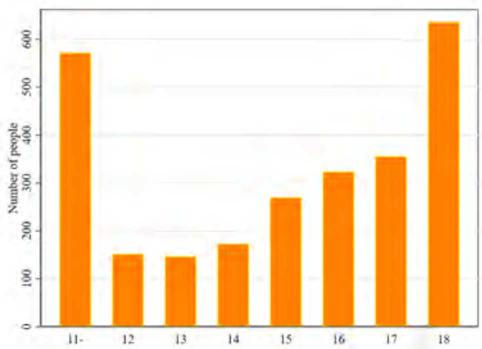
Note: Data from Indonesian Census 2010.

Figure C.5: Indonesia: location of selected regencies



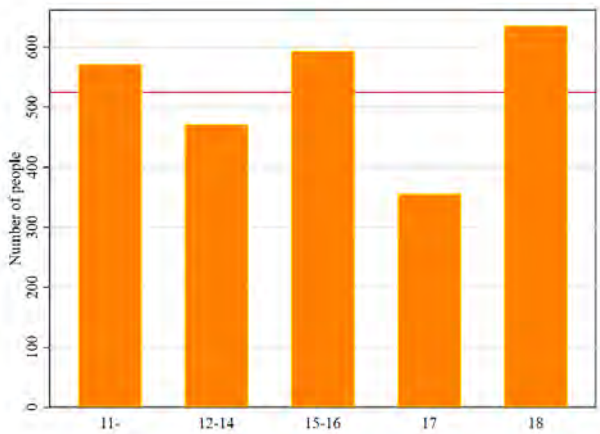
Note: The map shows with red dots the locations of the city of Medan and Bogor regency.

Figure C.6: Indonesia: number of women by age they left their birthplace



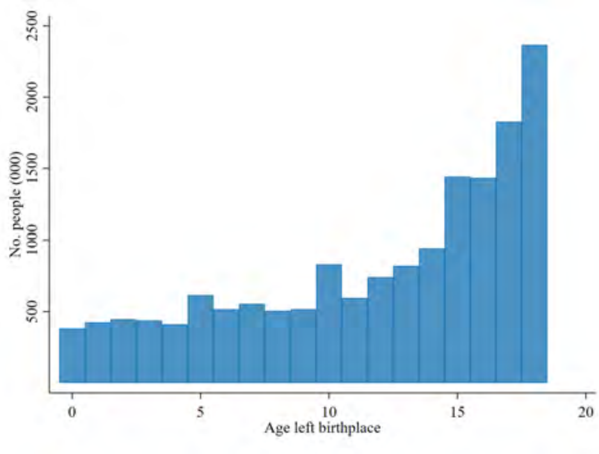
Note: Sample restricted to women who outmigrated before they turned 19 years old. Data from the IFLS.

Figure C.7: Indonesia: number of women by age they left their birthplace



Note: Sample restricted to women who outmigrated before they turned 19 years old. The red line shows the average count across all age brackets. Data from the IFLS.

Figure C.8: Indonesia: number of emigres by emigration age, 0-19 years old



Note: Data from 1985, 1995, and 2005 Intercensal Surveys.

D Tables

Table D.1: IFLS: summary statistics by gender and migration status

	All	**Women**	**Men**
	(1)	(2)	(3)
Age	35.54	35.27	35.85
Married	0.84	0.74	0.97
Attended at least high school	0.37	0.32	0.42
Urban	0.28	0.27	0.28
Muslim	0.89	0.90	0.89
Share left birthplace by age 25	0.31	0.29	0.34
Employed	0.71	0.55	0.89
Type of worker			
Self-employed	0.46	0.40	0.50
Salaried	0.42	0.37	0.46
Unpaid / family worker	0.12	0.23	0.04
Industry of employment			
Agriculture	0.31	0.31	0.32
Services	0.40	0.44	0.38
Manufacturing	0.14	0.16	0.13
Construction	0.05	0.01	0.09
Observations	516,670	276,986	239,684
Number of individuals	37,440	19,074	18,366

Notes: data from IFLS. Urban shows the share of people who report living in a town or a big city. This definition differs from the BPS defined urban classification available in the Indonesian Census and SUSENAS.

Table D.2: Indonesia: women's characteristics by migration status

	Stayers	**Emigres**	**Emigrated young**
	(1)	(2)	(3)
Age	36.07	35.34	33.27
Married	0.76	0.77	0.73
Attended at least high school	0.24	0.44	0.40
Urban	0.12	0.58	0.58
Muslim	0.91	0.87	0.88
Share left birthplace by age 25		0.82	1.00
Employed	0.57	0.53	0.51
Type of worker			
Self-employed	0.43	0.36	0.36
Salaried	0.31	0.45	0.45
Unpaid / family worker	0.25	0.19	0.19
Industry of employment			
Agriculture	0.37	0.20	0.20
Services	0.37	0.55	0.55
Manufacturing	0.16	0.15	0.15
Construction	0.00	0.01	0.01
Observations	169,669	68,619	29,871
Number of individuals	11,555	6,769	2,933

Notes: Non emigres are women who *never* left their birthplace. Migrant shows women who live outside their birthplace. Data from IFLS.

Table D.3: Indonesia: emigration reason by age and gender

	Women		**Men**	
	All	**Left young**	**All**	**Left young**
	(1)	(2)	(3)	(4)
Family	0.81	0.83	0.54	0.64
Work	0.13	0.10	0.37	0.23
Education	0.05	0.07	0.09	0.12

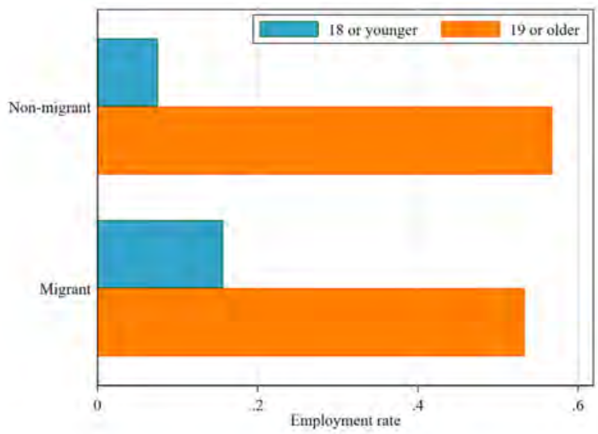
Notes: data from 1985 Intercensal Survey. Columns (1) and (3) breakdown emigration motive for all migrants. Columns (2) and (4) do so for women and men who emigrated before they turned 19.

Table D.4: IFLS: women's migration patterns and regency characteristics by urbanicity of regency of origin

	Birth regency		
	Rural	Urban	Total
	(1)	(2)	(3)
Number of regencies	135	94	229
Share of IFLS women born in these regencies	0.49	.51	100
Migration rate	0.30	0.27	0.28
A. Share of emigres living in:			
Rural regencies	0.37	0.28	0.67
Urban regencies	0.63	0.72	0.67
B. Characteristics of origin regency			
Women's employment rate			
Average	0.57	0.46	0.52
STD	0.14	0.11	0.14

Notes: I define migration as living outside the regency of birth. In this table, for people observed in more than one regency during their adulthood, I keep the most recent place of residence only. Following Bryan and Morten (2019) I classify regencies as urban if the share of population living in an urban area is above a cutoff. I choose the cutoff to match the urban share at the national level. Data from IFLS and IPUMS International.

Figure D.1: IFLS: women's employment rates by age and migration status



Note: The figure shows women's employment rates by age and migration status. I define as migrants women who reside outside their birthplace. Non-migrants are those who *never* left their birthplace. Data from IFLS.

Table D.5: Indonesia: number of existing rengencies by year, 1980-2010

	1980	1990	2000	2010
Number of regencies	286	295	339	493

Notes: These regencies have chaning borders across decades. In my analysis, in each year, I aggregate these unis into 268 consistent-boundary regencies. Data IPUMS international.

Table D.6: Regency-level summary statistics

	Mean	**Std. Dev.**	**Min**	**Max**	**Observations**
	(1)	(2)	(3)	(4)	(5)
Population	533,867	525,307	18,430	3,909,730	268
Share urban	0.45	0.30	0.07	1.00	268
Employment rate					
Women	0.53	0.14	0.29	0.91	268
Men	0.87	0.04	0.70	0.94	268
Industry composition					
Agriculture	0.43	0.23	0.00	0.81	268
Mining	0.01	0.03	0.00	0.26	268
Manufacturing	0.08	0.08	0.01	0.42	268
Construction	0.05	0.03	0.01	0.14	268
Services	0.35	0.13	0.12	0.68	268
Share married	0.75	0.05	0.56	0.93	268
Employment rate					
Share with at least high school					
Women	0.32	0.15	0.06	0.80	268
Men	0.36	0.15	0.10	0.82	268
Share literate					
Women	0.92	0.07	0.59	1.00	268
Men	0.95	0.05	0.72	1.00	268

Notes: table aggregates regencies to keep boundaries consistent across time. Table uses information from Indonesian Census and SUSENAS 2012.

Table D.7: Autocorrelation in women's employment rate for different regency boundaries, 1980-2010

	Fixed boundaries			**Raw boundaries**		
	(1)	(2)	(3)	(4)	(5)	(6)
$t - 10$ years	0.80			0.71		
	(0.02)			(0.02)		
$t - 20$ years		0.72			0.56	
		(0.03)			(0.03)	
$t - 30$ years			0.70			0.46
			(0.04)			(0.05)
Observations	800	534	268	878	541	278

Notes: sample restricted to women aged 18-64. Data IPUMS international. Robust standard errors in parenthesis. Columns (1) to (3) aggregate regencies to keep geographic units with consistent boundaries across time. Columns (4) to (6) the boundaries from the raw data.

Table D.8: Indonesia: autocorrelation in regency-level women's employment rate, 1980-2010

	(1)	(2)	(3)
$t-10$ years	0.81		
	(0.06)		
$t-20$ years		0.73	
		(0.07)	
$t-30$ years			0.78
			(0.05)
Observations	800	534	268

Notes: The table shows the autocorrelation in regency-level employment rates. Data from IPUMS international. Robust standard errors in parenthesis.

Table D.9: Indonesia: estimates birthplace persistence for alternative employment rate measures

Source of birthplace employment rates	(1)	(2)	(3)	(4)
Census 2010 (baseline)	0.36***			
	(0.04)			
Census 2000		0.29***		
		(0.04)		
Census 1990			0.27***	
			(0.04)	
Census 1980				0.26***
				(0.04)
Observations	64,727	64,727	64,741	64,727
N individuals	6,133	6,133	6,133	6,133
R^2	0.14	0.14	0.14	0.14

Notes: The table shows estimates of the birthplace persistence coefficients when I source the birthplace female employment rate from different census years. Uses data from IFLS and IPUMS international. Sample restricted to people residing outside their birthplace. All regressions controls for year, regency of residency, religion, and education FE, and a quadratic polynomial on age. Standard errors clustered by regency of origin.

Table D.10: Indonesia: tets for no difference in selection in destination characteristics by age of emigration

	Childhood information					Characteristics of destination		
	Formal worker father	**11+ books**	**Ppl. per room**	**No. brothers**	**No. sisters**	**FLFP**	**% agriculture**	**% urban**
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
p_b	-0.13	-0.26**	0.32	0.43	0.47	0.28***	0.10	-0.09
	(0.14)	(0.12)	(0.45)	(0.62)	(0.51)	(0.05)	(0.11)	(0.15)
Age of outmigration and p_b interactions								
11 or less	-0.08	0.17	0.01	-0.53	-0.02	-0.09	0.15	-0.11
	(0.21)	(0.18)	(0.41)	(0.55)	(0.63)	(0.06)	(0.14)	(0.20)
12-14	-0.30	0.23	-0.28	-0.27	-0.08	-0.08	0.09	-0.11
	(0.22)	(0.20)	(0.65)	(0.76)	(0.66)	(0.07)	(0.14)	(0.19)
15-16	-0.23	0.08	0.11	-0.51	-0.72	-0.03	-0.01	0.02
	(0.19)	(0.16)	(0.63)	(0.58)	(0.55)	(0.06)	(0.13)	(0.19)
17	-0.23	0.21	-0.07	-1.47**	-0.14	-0.09	-0.05	0.10
	(0.24)	(0.20)	(0.49)	(0.68)	(0.77)	(0.08)	(0.17)	(0.25)
Controls								
Emigration age FE	✓	✓	✓	✓	✓	✓	✓	✓
Year FE						✓	✓	✓
Age						✓	✓	✓
Age^2						✓	✓	✓
Interactions 0-17 are equal to zero								
F-statistic	0.62	0.49	0.15	1.82	0.69	0.68	0.54	0.40
p-value	0.65	0.74	0.96	0.13	0.60	0.61	0.71	0.81
Interactions 0-16 are equal to zero								
F-statistic	0.83	0.58	0.19	0.46	0.91	0.84	0.50	0.22
p-value	0.48	0.63	0.90	0.71	0.43	0.48	0.68	0.88
Observations	2,629	2,624	2,622	2,629	2,629	27,977	27,977	27,977
Number of people						2,629	2,629	2,629

Notes: This table shows test of the hypothesis that selection is constant across emigration age. It shows estimates of the interactions between age of emigration dummies and the birthplace female employment rates in regression (8). I chose 18 years old as the base category so that the interactions test whether the slopes at each age are different from that for women emigrating at 18. Columns (1) to (5) use one observation per individual. Standard errors clustered by birthplace regency. Data from the IFLS.

Table D.11: Indonesia: tests for no difference in selection in reason of emigration

	Work (1)	**Family** (2)	**Education** (3)
p_b	0.11***	-0.08***	0.03
	(0.03)	(0.02)	(0.04)
Age of outmigration and p_b interactions			
0-2	-0.12***	0.09***	-0.01
	(0.03)	(0.03)	(0.05)
3-5	-0.07**	0.09***	-0.01
	(0.03)	(0.03)	(0.04)
6-8	-0.10***	0.10***	-0.02
	(0.03)	(0.03)	(0.05)
9-11	-0.08**	0.10***	-0.03
	(0.03)	(0.03)	(0.05)
12-14	-0.07*	0.10***	-0.03
	(0.03)	(0.03)	(0.04)
15-16	-0.02	0.04	0.02
	(0.03)	(0.03)	(0.05)
17	-0.02	0.04	-0.01
	(0.04)	(0.04)	(0.04)
Emigration age FE	✓	✓	✓
Destination regency by year FE	✓	✓	✓
Age	✓	✓	✓
Age^2	✓	✓	✓
Interactions 0-17 are equal to zero			
F-statistic	2.73	3.83	0.91
p-value	0.01	0.00	0.50
Interactions 0-17 are equal			
F-statistic	2.81	2.26	1.05
p-value	0.01	0.04	0.39
Interactions 3-14 are equal			
F-statistic	0.55	0.99	0.31
p-value	0.65	0.40	0.82
Observations	39,803	39,803	39,803

Notes: This table shows test of the hypothesis that selection is constant across emigration age. It shows estimates of the interactions between age of emigration dummies and the birthplace female employment rates in regression (8). I chose 18 years old as the base category so that the interactions test whether the slopes at each age are different from that for women emigrating at 18. Data from the 1985, 1995 and 2005 Intercensal Survey.