

NBER WORKING PAPER SERIES

MEASURING INFORMATION FRICTIONS IN MIGRATION DECISIONS:  
A REVEALED-PREFERENCE APPROACH

Charly Porcher  
Eduardo Morales  
Thomas Fujiwara

Working Paper 32413  
<http://www.nber.org/papers/w32413>

NATIONAL BUREAU OF ECONOMIC RESEARCH  
1050 Massachusetts Avenue  
Cambridge, MA 02138  
May 2024, Revised September 2024

We thank Treb Allen, Stéphane Bonhomme, Lorenzo Casaburi, Rafael Dix-Carneiro, Dave Donaldson, Cécile Gaubert, Gabriel Kreindler, Melanie Morten, Diego Puga, and James Tybout, for insightful conversations, and seminar participants at Aarhus, UC Berkeley, Carlos III, CEMFI, Chicago-Princeton Conference on Spatial Economics, Cornell, Dartmouth, Johns Hopkins SAIS, MIT, Penn State, Princeton, Stanford, University of Toronto, University of Virginia, University of Zurich, UPF, Yale, and the NBER Summer Institute for helpful comments. We especially thank Milena Almagro for her thoughtful discussion, and Mark Bamba, Juanma Castro-Vincenzi, Pedro Cabra-Acela, and Anaïs Galdin for excellent research assistance. All errors are our own. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2024 by Charly Porcher, Eduardo Morales, and Thomas Fujiwara. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Measuring Information Frictions in Migration Decisions: A Revealed-Preference Approach  
Charly Porcher, Eduardo Morales, and Thomas Fujiwara  
NBER Working Paper No. 32413  
May 2024, Revised September 2024  
JEL No. F10, J61, R23

**ABSTRACT**

We investigate the role of information frictions in migration. Using novel moment inequalities and data on internal migration in Brazil, we estimate worker preferences and migration costs while allowing for unobserved worker-specific information sets. We find that common estimation procedures overestimate migration costs and underestimate the importance of expected wages in migration decisions. Model specification tests indicate that workers often have limited information on location-specific wages. However, those living in regions with better internet access and larger populations have more precise wage information, and information precision decreases with distance. According to our estimated model, workers' limited wage information plays a quantitatively important role in reducing migration flows and worker welfare, and limits the effect of policies that reduce migration costs.

Charly Porcher  
McDonough School of Business  
Georgetown University  
37th & O Street, NW  
Washington, DC 20057  
cp1105@georgetown.edu

Thomas Fujiwara  
Department of Economics  
Princeton University  
131 Julis Romo Rabinowitz Building  
Princeton, NJ 08544  
and NBER  
fujiwara@princeton.edu

Eduardo Morales  
Department of Economics  
Princeton University  
291 Julis Romo Rabinowitz Building  
Princeton, NJ 08544  
and NBER  
ecmorale@princeton.edu

# 1 Introduction

Migration is among the most effective ways for workers to improve their economic conditions. Yet, even within countries, migration rates are low. Both a lack of information about potential destinations and high migration costs may reduce migration, but disentangling their roles is difficult because researchers rarely observe what workers know. Studies often place strong assumptions on migrants' information and focus on estimating migration costs, sometimes interpreting them as accounting for information frictions. However, information frictions affect migration decisions differently than costs do. While costs affect how beneficial a move is, lack of information does not, but may lead to mistakes in people's choices. Increasing access to information will thus help improve workers' location decisions, and may enhance the benefits of policies that reduce migration costs.

We introduce a moment inequality procedure to separately identify the role of information frictions and migration costs in workers' location choices. Crucially, our procedure is valid even if workers differ in their information sets in ways unobserved to the researcher. We use it to answer four questions. What do workers know about wages in different locations? How does allowing for unobserved information sets affect estimates of how workers trade off expected wage gains against migration costs when choosing their location? How would workers' migration decisions change if their information changed? How does workers' information mediate the impact of changes in migration costs?

Our estimation procedure extends the treatment of information sets in the moment inequality literature to multinomial logit models with potentially large choice sets and choice-specific fixed effects. We apply our estimator to data on wages and location choices of formally employed workers in Brazil, obtaining four main results. First, workers generally only have coarse information on location-specific wages. However, those living in areas with better internet access or larger populations are better informed. Second, our estimates of the migration elasticity to expected wages are three times larger than those obtained using common estimation procedures that place stronger assumptions on workers' information, whereas our migration cost estimates are 21% lower on average. Third, migration rates would increase significantly if all information frictions were removed. Fourth, researchers assuming workers' information is better than it truly is will overestimate the welfare gains from reductions in migration costs; e.g., due to improvements in transportation infrastructure.

Our baseline analysis is based on a static migration model that incorporates expectations on wages, migration costs, amenities, and prices, as well as idiosyncratic preferences, as drivers of workers' location choices.<sup>1</sup> We impose no restriction on wage expectations beyond

---

<sup>1</sup>In Appendix G, we extend our analysis to models with sunk migration costs and forward-looking workers.

assuming these are rational; as a result, expected wages in any given destination may vary across workers with different information sets, which we treat as potentially heterogeneous across every worker and unobserved to the researcher. Conversely, for each destination, we assume workers' expectations on migration costs, amenities, and prices are common across workers sharing a prior location of residence, but allow them to vary flexibly across these worker groups, and impose no restriction on their rationality. Thus, while wage expectations give rise to individual unobserved heterogeneity, expected migration costs, amenities, and prices are captured by destination-specific effects that may vary across workers depending only on their prior location. A second source of unobserved heterogeneity in workers' choices in our model is the presence of idiosyncratic preferences that we assume independent across workers and destinations and following a type I extreme value distribution.<sup>2</sup>

Our modeling of expected migration costs, amenities, and prices as captured by origin-by-destination fixed effects implies the number of preference parameters increases in the square of the number of locations considered in the analysis. When studying migration decisions, workers' feasible set is in most cases large and, thus, the number of fixed effects to estimate will also be large. Estimating high-dimensional parameter vectors using moment inequalities is computationally challenging when using standard inference procedures. To address this challenge, we introduce a moment inequality procedure that allows to calculate confidence intervals on each preference parameter in multinomial logit models featuring potentially large choice sets, choice-specific fixed effects, and information sets that may be different between any two agents in ways unobserved to the researcher.

Key for our estimation procedure is a new type of moment inequality that we call *bounding* inequality. To derive it, we compare workers' expected utility in any two locations in their choice set, obtaining as a result a conditional moment *equality* that depends on a concave function of the worker's expected utility difference in those two locations. The conditioning set in this moment equality is a covariate vector assumed to belong to the worker's information set. From this equality, we derive moment *inequalities* by bounding the concave function from above by its tangent at any point. Importantly, the resulting inequalities are linear in the worker's expected utility difference between the two locations. We then substitute workers' unobserved wage expectations with the observed realized wages. While this introduces workers' expectational errors in the moment, the rational expectations assumption implies these are mean zero conditional on any variable in the worker's information set. This property of expectational errors, combined with the linearity of the bounding inequality, leaves

---

<sup>2</sup>Generally, our estimation procedure allows for unobserved heterogeneity in the information agents have on choice characteristics whose realized values the researcher observes, but not in the information they have on characteristics the researcher cannot observe. Hence, the different treatment of workers' expectations on wages, which are observed in our data, and location-specific prices or migration costs, which are unobserved.

this one unaffected. The resulting moment inequality thus depends on the difference in the fixed effects corresponding to the two locations being compared, as well as on the utility difference coming from any wage variation between both locations.<sup>3,4</sup>

We use our bounding inequality in a two-step estimation procedure. The first step provides bounds on the wage preference parameter. These are computed by combining the bounding inequality described above for pairs of workers that share the same origin location but have distinct observed wages in any given destination; e.g., because they are employed in different sectors. The resulting inequality does not depend on workers’ expected migration costs, amenities, and price levels, which are differenced out when comparing the utilities of two workers of the same origin. However, it does depend on the wage differences between the two workers. This inequality may thus be used to compute a confidence interval for the wage preference parameter. In the second step, we bound *one at a time* each of the origin-by-destination fixed effects that capture workers’ expected migration costs, amenities, and prices. Thus, instead of estimating a joint confidence set for all fixed effects, which is infeasible in settings with many choices, we estimate separate confidence intervals for each fixed effect. To compute these, we use the bounds on the wage coefficient estimated in the first step, and combine the bounding inequalities described above with the type of odds-based moment inequalities introduced in [Dickstein et al. \(2023\)](#).

We show theoretically that our bounding moment inequalities provide non-trivial bounds on all parameters when the researcher correctly specifies a subset of agents’ information sets, and point identify them when such subset coincides with agents’ true information sets. Thus, point identification is achieved precisely when maximum likelihood estimators are consistent, implying no loss of identification power is incurred in this case when using our inequalities.<sup>5</sup> Furthermore, when the researcher only observes a proper subset of agents’ true information sets, we show in simulations that the maximum likelihood estimates are not only biased, but also often outside of the bounds defined by our inequalities. Our simulations also illustrate that, when the researcher incorrectly includes a variable that does not belong to the worker’s information set in the moment inequality conditioning set, the resulting confidence set may be empty.<sup>6</sup> We use this result in our empirical application to test how precise workers’

---

<sup>3</sup>Prior work has derived similarly linear inequalities when preference shocks have limited heterogeneity (e.g., [Wollman, 2017](#)) or a bounded support (e.g., [Eizenberg, 2014](#)). We do so in a multinomial logit model.

<sup>4</sup>More generally, our inequality will depend on the utility difference coming from any choice characteristic that is both observed by the researcher and not absorbed by the included fixed effects.

<sup>5</sup>A loss of identification power may however arise from having to combine bounding inequalities for worker pairs in the first step of our estimation procedure; see corollaries 1 and 3 in Section 3 for more details.

<sup>6</sup>When the conditioning set includes variables outside of the worker’s information set, the resulting identified set is sometimes non-empty while not including the true parameter value. See [Molinari \(2020\)](#) for a discussion of this phenomenon, and [Andrews and Kwon \(2024\)](#) and [Kaido and Molinari \(2024\)](#) for inference procedures in partially identified models robust to misspecification.

information on location-specific wages is.

We apply our estimator to the study of internal migration in Brazil. We study a population of workers with similar demographic characteristics, and use data from the *Relação Anual de Informações Sociais* (RAIS), which has information on the sector and region of work, and wages, for all formally employed workers.<sup>7</sup> We define a labor market as a sector-region pair, and study the information workers have on market-by-period wage shifters. These shifters account for all demand and supply factors having a common impact on the wages of all workers in our population in a given market and period. To estimate these shifters, we regress observed wages on sector-by-region-by-period fixed effects (which correspond to the shifters of interest) while controlling for worker-by-sector fixed effects (which account for unobserved worker abilities in each sector) and several time-varying observed worker characteristics with sector-specific coefficients. The estimated regression approximates observed wages well, with a cross-sectoral median  $R^2$  of 0.83, suggesting that terms not included in it, such as worker-by-region effects, play a relatively small role in determining workers' wages.

Our analysis yields four conclusions. First, workers face substantial and heterogeneous information frictions. Specifically, we do not reject workers know the wage quartile to which each labor market belonged in the previous year, but reject that they *in general* have more precise information than that. That is, we cannot reject workers can determine whether a labor market is in the top 25% by its previous year's sector-by-region shifter, in the 50-75% bracket, in the 25-50% bracket, or in the bottom 25%, but we reject that they can classify markets according to finer partitions. However, workers in regions with better internet access or larger populations have more precise information overall, and workers in general have more accurate information about wages in nearby areas.

Second, relative to our moment inequalities, estimators common in the migration literature yield smaller estimates of the migration wage elasticity and larger migration cost estimates. Specifically, our approach yields a 95% confidence interval for the elasticity of migration to expected wages centered at 1.5, and does not include the Poisson Pseudo-Maximum Likelihood (PPML) estimate of 0.5. In addition, our migration cost estimates (measured in utility terms) are centered around values 21% lower than the PPML estimates.<sup>8</sup>

Third, we estimate substantial gains from improving workers' information. In particular, the 95% confidence interval for the welfare change that results from improving the precision of workers' information from only discerning quartiles of lagged shifters to knowing their

---

<sup>7</sup>Specifically, we study workers aged 25-64 with at least a high school degree identified as male and white. These are, respectively, the largest education, gender, and race categories in RAIS.

<sup>8</sup>We implement the PPML estimator in [Artuç and McLaren \(2015\)](#), which is consistent if all workers in the same sector have a common information set. The difference between the PPML and moment inequality estimates is thus consistent with our finding that information sets are heterogeneous by worker locations.

current values is  $[2.5\%, 4.1\%]$ . Our confidence intervals also reveal that, while workers with perfect information on current shifters would have migration rates between 9 and 14%, those who only discern quartiles of lagged shifters would have lower rates between 4 and 7%.

Fourth, the information frictions we infer impact the welfare gains from reductions in migration costs. Specifically, the 95% confidence interval for the change in welfare that results from a 10% decline in migration costs is  $[4.2\%, 5.6\%]$  if workers are fully informed about current sector-by-region wage shifters, but only  $[2.2\%, 2.9\%]$  if they only observe the quartiles of the previous year’s shifters.

Our paper is related to three strands of the literature. First, it relates to work studying workers’ mobility within countries. Our baseline static model incorporates worker location-specific idiosyncratic preferences and fixed migration costs as in, e.g., [Tombe and Zhu \(2019\)](#) and [Morten and Oliveira \(2024\)](#). In the dynamic extension to our model, we further allow for forward-looking workers and one-time migration costs, as in, e.g., [Kennan and Walker \(2011\)](#). Our contribution is to show how to estimate static and dynamic migration models without fully specifying workers’ information, and to illustrate the impact on model estimates and counterfactual predictions of misspecifying workers’ information sets.<sup>9</sup>

Second, we contribute to the literature studying information frictions in migration. Recent work has used randomized or natural experiments to evaluate the impact of migrants’ information on their location choices, e.g., [Bryan et al. \(2014\)](#); [Bergman et al. \(2020, 2023\)](#); [Wilson \(2021\)](#); [Baseler \(2023\)](#). In the absence of exogenous variation in information sets, other studies follow a structural approach. [Kaplan and Schulhofer-Wohl \(2017\)](#) introduce a model in which workers acquire information on location characteristics through a Bayesian process. [Porcher \(2022\)](#) extends this approach by endogenizing the information acquisition process of workers who are rationally inattentive. Our contribution is to infer the importance of information frictions while neither observing exogenous shifters of agents’ information sets nor imposing parametric restrictions on the stochastic process determining these sets.

Third, our paper is related to studies using choice data to identify agents’ preferences when their expectations of choice characteristics are rational but unobserved. In the absence of direct measures of agents’ subjective expectations ([Manski, 2004](#)), it is common to assume that the researcher observes agents’ entire information sets (e.g. [Manski, 1991](#)). A recent approach allows the content of agents’ information sets to be partly unobserved to the researcher, but rules out heterogeneity in such content across agents of the same observable type (e.g., [Traiberman, 2019](#)). Building on the moment inequality literature (e.g., [Pakes, 2010](#); [Ho and Pakes, 2014](#); [Pakes et al., 2015](#)), we allow information sets to vary across agents in

---

<sup>9</sup>We study the decision problem of individual workers; see [Fan et al. \(2023\)](#) for a first-order approach that incorporates imperfect information in a general-equilibrium migration model à la [Caliendo et al. \(2019\)](#).

unobservable ways, and use moment inequalities to partially identify agents' preferences in a multinomial logit model, static or dynamic, with large choice sets and choice-specific fixed effects. We combine our new bounding inequalities with the odds-based inequalities in [Dickstein and Morales \(2018\)](#) and [Dickstein et al. \(2023\)](#) to obtain bounds that are informative and simple to compute. When applying our moment inequality estimator to the study of dynamic models, we combine the bounding moment inequalities with the Euler perturbation approach in [Morales et al. \(2019\)](#).

The paper is organized as follows. Section 2 presents a model of workers' location choices. Section 3 describes our moment inequality estimator and Section 4 illustrates its properties using simulated data. Section 5 discusses our empirical application. Section 6 concludes.

## 2 Model of Migration with Incomplete Information

We model the location choice of workers in a population. Workers are partitioned into  $S$  types indexed by  $s$  or  $r$ , and are indexed by  $i$  or  $j$  within a type. Defining a variable  $y_{is}^l$  that equals one if worker  $i$  of type  $s$  chooses location  $l$  (and zero otherwise), we assume

$$y_{is}^l \equiv \mathbb{1}\{l = \operatorname{argmax}_{l'=1,\dots,L} \mathbb{E}[\mathcal{U}_{is}^{l'} | \mathcal{J}_{is}]\} \quad \text{for } l = 1, \dots, L, \quad (1)$$

where  $\mathbb{1}\{A\}$  is an indicator function that equals 1 if  $A$  is true,  $\mathcal{U}_{is}^l \in \mathbb{R}$  denotes the worker's utility of choosing  $l$ ,  $\mathcal{J}_{is} \in \mathbb{R}^{d_{is}}$ , with  $d_{is} \geq 0$ , is the worker's information set, and  $\mathbb{E}[\cdot | \mathcal{J}_{is}]$  is a conditional expectation operator reflecting the worker's beliefs.<sup>10</sup>

We impose five assumptions on workers' expected utilities. First, workers' expectations are rational. That is, for any vector  $\mathcal{X}_{is}$ , denoting by  $F(\cdot | \mathcal{J}_{is})$  the cumulative distribution function of  $\mathcal{X}_{is}$  conditional on  $\mathcal{J}_{is}$ , it holds that

$$\mathbb{E}[\mathcal{X}_{is} | \mathcal{J}_{is}] = \int_x x dF(x | \mathcal{J}_{is}). \quad (2)$$

Second, the utility of choosing location  $l$  for worker  $i$  of type  $s$  is

$$\mathcal{U}_{is}^l = u_{is}^l + \varepsilon_{is}^l, \quad (3a)$$

$$u_{is}^l = \kappa^l + \alpha w_{is}^l, \quad (3b)$$

---

<sup>10</sup>Equation (1) assumes a common choice set for all workers. The moment inequality estimator described in Section 3 can be adapted to allow for heterogeneous choice sets, as long as these are partly observed by the researcher and independent of both workers' information sets and location-specific utilities. See [Barseghyan et al. \(2021\)](#) for a more flexible treatment of agents' choice sets in discrete choice models.

where  $w_{is}^l$  is the natural logarithm of the nominal wage worker  $i$  of type  $s$  would earn if they chose location  $l$  and  $\alpha$  captures the relative importance of wages in workers' utility. The terms  $\kappa^l$  and  $\varepsilon_{is}^l$  are the common and idiosyncratic components of all other determinants of a worker's location choice. For simplicity, we refer to  $\kappa^l$  as location  $l$ 's amenity, although it will also account for location-specific log prices and, depending on the definition of the populations of study, for other determinants of workers' preferences such as migration costs.<sup>11</sup>

Third, defining  $\varepsilon_{is} = (\varepsilon_{is}^1, \dots, \varepsilon_{is}^L)$ , we assume that, for worker  $i$  of type  $s$ ,

$$\varepsilon_{is} \subseteq \mathcal{J}_{is}, \quad (4)$$

where, for vectors  $\mathcal{X}$  and  $\mathcal{X}'$ , we use  $\mathcal{X} \subseteq \mathcal{X}'$  to indicate that the distribution of  $\mathcal{X}$  conditional on  $\mathcal{X}'$  is degenerate. Equation (4) imposes that, when making their location choice, worker  $i$  of type  $s$  knows  $\varepsilon_{is}$ . It does not restrict which other variables belong to  $\mathcal{J}_{is}$ .

Fourth, for worker  $i$  of type  $s$ , worker  $j$  of a type  $r \neq s$ , and locations  $l$  and  $l'$ , it holds

$$\mathbb{E}[\Delta w_{is}^{ll'} | \mathcal{J}_{is}, \mathcal{J}_{jr}] = \mathbb{E}[\Delta w_{is}^{ll'} | \mathcal{J}_{is}] = \mathbb{E}[\Delta w_s^{ll'} | \mathcal{J}_{is}] = \mathbb{E}[\Delta w_s^{ll'} | \mathcal{W}_{is}], \quad (5)$$

where  $\mathcal{W}_{is}$  includes all elements of  $\mathcal{J}_{is}$  other than the idiosyncratic preferences in  $\varepsilon_{is}$ ; i.e.,  $\mathcal{J}_{is} = (\mathcal{W}_{is}, \varepsilon_{is})$ .<sup>12</sup> The first equality in equation (5) imposes that a worker of type  $s$  has at least as much information as any worker of a different type  $r$  about the differences in the wages the type- $s$  worker would earn in different locations. The second equality imposes that the worker's expected wage difference between locations  $l$  and  $l'$  only depends on the expected difference between terms that are common to all workers of the same type. Finally, the third equality imposes that, once we condition on all other elements of the worker's information set, idiosyncratic preferences do not help the worker forecast wages.

The first equality in equation (5) naturally holds if all workers in the population have the same information; i.e., if  $\mathcal{J}_{is} = \mathcal{J}_{jr}$  for any  $i, j, s$ , and  $r$ . However, it also holds if workers of a given type  $s$  know more about their type-specific wage differences than any worker of a different type.<sup>13</sup> Furthermore, this equality leaves the variation in information sets between workers of the same type completely unrestricted.

The second equality in equation (5) is imposed by data limitations. Our moment inequality procedure allows us to flexibly model the information workers have on payoff-relevant

<sup>11</sup>When workers differing in their prior location are classified into different populations, they may differ in the value of  $\kappa^l$  in any  $l$  and, thus, these parameters will account for origin-by-destination migration costs.

<sup>12</sup>Generally, for any variables  $x_{is}^l$  and  $x_{is}^{l'}$ , we define  $\Delta x_{is}^{ll'} \equiv x_{is}^l - x_{is}^{l'}$ .

<sup>13</sup>When types correspond to workers' employment sector, equation (5) imposes that, e.g., real estate workers know more about differences across locations in real estate wages than healthcare workers do, and vice versa. We rely on this assumption to derive the first-step moment inequalities comparing workers in different sectors.

variables whose ex-post (or realized) value the researcher either observes or can consistently estimate. Generally, one cannot estimate, for every worker and location, a wage component that is location- and worker-specific. Hence, workers' wages may contain idiosyncratic location components, but we must impose that workers do not know them when making their location choice. While equation (5) limits workers' information about their idiosyncratic location-specific wage components, it does not restrict the information they may have about their idiosyncratic type-specific wage components or about type-by-location wage shifters. Thus, when types correspond to sectors, it is consistent with [Dix-Carneiro \(2014\)](#), who show that workers choose their sector based on worker-specific wage components.

The third equality in equation (5) is an implication of the exogeneity of idiosyncratic shocks often assumed in discrete choice models.<sup>14</sup>

Fifth, and last, denoting by  $F_\varepsilon(\cdot)$  the cumulative distribution function of  $\varepsilon_{is}$ , it holds

$$F_\varepsilon(\varepsilon_{is}|\mathcal{W}_{is}, \mathcal{J}_{jr}) = F_\varepsilon(\varepsilon_{is}|\mathcal{W}_{is}) = F_\varepsilon(\varepsilon_{is}) = \exp\left(-\sum_{l=1}^L \exp(-\varepsilon_{is}^l)\right). \quad (6)$$

The first equality imposes that a worker's idiosyncratic preferences are independent of all elements of all other workers' information sets, including their idiosyncratic preferences. The second equality imposes that a worker's idiosyncratic preferences are independent of all other elements of their own information sets. The third equality imposes that  $\varepsilon_{is}^l$  is *iid* across all  $l = 1, \dots, L$ , and follows a type I extreme value distribution with location parameter equal to zero and scale parameter equal to one.

Equations (1) to (6) are the only assumptions we impose. Hence, we not only let workers' wage information vary flexibly across workers and locations in unobservable ways, but also leave the data-generating process of wages unrestricted. In contrast, most studies of workers' location choice assume wages are known to agents, or specify both a stochastic process for wages and workers' information sets (e.g., [Kennan and Walker, 2011](#)), or assume these information sets arise from a particular learning process (e.g., [Kaplan and Schulhofer-Wohl, 2017](#); [Porcher, 2022](#)). As workers' information is likely heterogeneous, unobserved to researchers, and difficult to model correctly, the absence of restrictions on agents' information sets is potentially important from an empirical perspective. We show in Section 5 that this is the case in our application.

Equations (2) to (5) imply we can write  $\mathbb{E}[\mathcal{U}_{is}^l|\mathcal{J}_{is}] = \kappa^l + \alpha\mathbb{E}[w_{is}^l|\mathcal{W}_{is}] + \varepsilon_{is}^l$ . Thus, it would be isomorphic if, instead of assuming that workers have perfect information on the

---

<sup>14</sup>Importantly, our model allows for any pattern of correlation between  $\kappa = (\kappa^1, \dots, \kappa^L)$  and the wage shifters  $w_s = (w_s^1, \dots, w_s^L)$  of any type  $s$ . It is thus consistent with the common observation that location-specific amenities and wages are correlated ([Redding and Rossi-Hansberg, 2017](#)).

value of the amenity  $\kappa^l$ , we had considered  $\kappa^l$  as capturing the expectation that all workers in the population have about the amenity level in  $l$ . This formulation would allow workers to have irrational expectations over amenity levels, but would restrict these expectations to be common across workers in a population. Conversely, equation (2) restricts workers' wage expectations to be rational, but our flexible modeling of workers' information sets allows expected wages in any given location to differ across any two workers in the population.

We consider settings in which the researcher observes a random sample of workers of each type. For all workers, the researcher observes the location choice  $y_{is} = (y_{is}^1, \dots, y_{is}^L)$ . Additionally, for all types, the researcher observes the vector of wage shifters  $w_s = (w_s^1, \dots, w_s^L)$  and a vector of covariates  $z_s = (z_s^1, \dots, z_s^L)$ , with  $z_s^l$  a potential predictor of  $w_s^l$ . In practice,  $w_s$  may not be directly observed but consistently estimated. We do not assume the researcher observes  $\mathcal{W}_{is}$  for any sampled worker.

Only differences between the elements of  $\kappa$  are identified. Thus, we impose  $\kappa^1 = 0$ . Subject to this normalization, the goal of estimation is to recover  $(\kappa^2, \dots, \kappa^L)$  and  $\alpha$ , and to learn about the content of workers' information sets. To acquire knowledge about workers' information, we test the null hypothesis that, for a specific subset of locations, a wage predictor  $z_s^l$  belongs to the information set of all workers in a specific group of interest. We denote by  $\theta \equiv (\theta_\alpha, \theta_2, \dots, \theta_L)$  the unknown parameter vector with true value  $\theta^* \equiv (\alpha, \kappa^2, \dots, \kappa^L)$ .

### 3 Estimation Through Moment Inequalities

When the number of choices  $L$  is large,  $\theta$  will be high dimensional. Commonly applied moment inequality inference procedures rely on inverting a test at each point in a grid covering the parameter space, complicating their applicability in settings with a large number of parameters.

We propose a two-step procedure that circumvents these computational challenges and produces a confidence interval for each element of  $\theta$  individually. In the first step, we compute a confidence interval for  $\theta_\alpha$  using inequalities that difference out the parameters in  $(\theta_2, \dots, \theta_L)$ . In the second step, for each  $l = 2, \dots, L$ , we derive inequalities that depend only on  $\theta_\alpha$  and  $\theta_l$ , which we combine with the first-step confidence interval for  $\theta_\alpha$  to obtain a confidence interval for  $\theta_l$ .

The moment inequalities we use in the first step combine those we use in the second step. Thus, for exposition purposes, we first describe the second-step inequalities in Section 3.1. We then describe how we build the first-step inequalities in Section 3.2. Section 3.3 explains how we use these inequalities to estimate confidence intervals for the elements of  $\theta$ .

### 3.1 Second-Step Moment Inequalities

We use two types of inequalities to identify bounds on  $\theta_l$  for each  $l = 2, \dots, L$ . In Section 3.1.1, we introduce a new type of moment inequality that we name *bounding* inequality. In Section 3.1.2, we describe how we apply the odds-based inequality in Dickstein et al. (2023) to our estimation problem.

#### 3.1.1 Bounding Moment Inequalities

Given two locations  $l$  and  $l'$ , we denote by  $\Delta\theta_{ll'} \equiv \theta_l - \theta_{l'}$  the unknown parameter whose true value is  $\Delta\kappa^{ll'} \equiv \kappa^l - \kappa^{l'}$ . Then, for any two locations  $l$  and  $l'$  in the worker's choice set, a random vector  $z_s$ , and a scalar random variable  $e_{is}^{ll'}$ , we define the moment

$$m^{ll'}(z_s, \Delta\theta_{ll'}) \equiv \mathbb{E}[y_{is}^{l'} - y_{is}^l \exp(-e_{is}^{ll'}) (1 + e_{is}^{ll'} - (\Delta\theta_{ll'} + \alpha \Delta w_s^{ll'})) | z_s]. \quad (7)$$

Theorem 1 establishes a property of this moment when evaluated at  $\Delta\theta_{ll'} = \Delta\kappa^{ll'}$ .

**Theorem 1** *Assume equations (1) to (6) hold. Then,  $m^{ll'}(z_s, \Delta\kappa^{ll'}) \geq 0$  if  $e_{is}^{ll'} \subseteq \mathcal{J}_{is}$  and  $z_s \subseteq \mathcal{W}_{is}$ .*

The proof of Theorem 1 is in Appendix A.1. This theorem states that, given equations (1) to (6), the moment in equation (7) is positive when evaluated at  $\Delta\theta_{ll'} = \Delta\kappa^{ll'}$  if, for all workers in the population,  $e_{is}^{ll'}$  belongs to  $\mathcal{J}_{is}$  and  $z_s$  belongs to its subset  $\mathcal{W}_{is}$ . Intuitively, to prove Theorem 1, we first derive a moment *equality* that depends on the ratio of the probability of choosing  $l'$  to the probability of choosing  $l$ , conditional on  $\mathcal{W}_{is}$ . This ratio is concave in  $\Delta\theta_{ll'} + \alpha \mathbb{E}[\Delta w_s^{ll'} | \mathcal{W}_{is}]$ , allowing us to derive a moment *inequality* by bounding it from above by its tangent at any given point  $e_{is}^{ll'}$ . The resulting inequality is thus *linear* in the worker's expected wage difference between  $l$  and  $l'$ , which we subsequently substitute with the corresponding difference in realized wages. This substitution introduces workers' expectational errors in the moment, but its linearity jointly with the conditions listed in Theorem 1 imply that these errors cancel out. Hence, we conclude that the expression in equation (7) must be weakly positive when evaluated at the true parameter values.

Theorem 1 implies that, if its conditions hold, the set of values of  $\Delta\theta_{ll'}$  for which

$$m^{ll'}(z_s, \Delta\theta_{ll'}) \geq 0 \quad (8)$$

will include  $\Delta\kappa^{ll'}$  regardless of the value of  $z_s$ , of the two locations  $l$  and  $l'$  being compared, and of how the approximation points  $e_{is}^{ll'}$  are chosen. However, given locations  $l$  and  $l'$  and a value of  $z_s$ , the values of  $\Delta\theta_{ll'}$  other than  $\Delta\kappa^{ll'}$  that satisfy the inequality in equation (8) will

depend on those approximation points. Appendix B.1 shows that, when restricting ourselves to approximation points that may be written as a function of  $z_s$  and  $\Delta\theta_{ll'}$ , the set of values of the parameter  $\Delta\theta_{ll'}$  that satisfy the inequality in equation (8) is minimized when:

$$e_{is}^{ll'} = \Delta\theta_{ll'} + \alpha \mathbb{E}[\Delta w_s^{ll'} | z_s, y_{is}^l = 1]. \quad (9)$$

To provide intuition on how the moment inequality implied by equations (7) to (9) may be used to partially identify  $\Delta\theta_{ll'}$ , we show in Appendix B.1 this inequality can be written as

$$\frac{\mathbb{E}[y_{is}^l | z_s]}{\mathbb{E}[y_{is}^{l'} | z_s]} \exp(-\alpha \mathbb{E}[\Delta w_s^{ll'} | z_s, y_{is}^l = 1]) \leq \exp(\Delta\theta_{ll'}). \quad (10)$$

We also show in Appendix B.1 that swapping the identity of locations  $l$  and  $l'$  in equations (7) to (9), and rearranging terms, we can obtain the inequality

$$\frac{\mathbb{E}[y_{is}^l | z_s]}{\mathbb{E}[y_{is}^{l'} | z_s]} \exp(-\alpha \mathbb{E}[\Delta w_s^{ll'} | z_s, y_{is}^{l'} = 1]) \geq \exp(\Delta\theta_{ll'}). \quad (11)$$

Equations (10) and (11) identify a lower and an upper bound on  $\Delta\theta_{ll'}$ , respectively. Intuitively, both bounds increase in the relative probability with which workers having access to information on  $z_s$  choose  $l$  over  $l'$ ,  $\mathbb{E}[y_{is}^l | z_s] / \mathbb{E}[y_{is}^{l'} | z_s]$ . The bounds however differ on how they net out the impact of expected wages on that relative probability. While equation (10) nets out the expected wage difference among workers knowing  $z_s$  that choose location  $l$ , equation (11) nets out a similar expected wage difference among workers choosing  $l'$ . Intuitively, both moments rely on  $z_s$  and on the worker's observed choices  $y_{is}^l$  and  $y_{is}^{l'}$  to bound the worker's true wage expectation. That is, the bounds in equations (10) and (11) rely on the property that, for any  $z_s \subseteq \mathcal{W}_{is}$ ,  $\mathbb{E}[\Delta w_s^{ll'} | z_s, y_{is}^l = 1] \leq \mathbb{E}[\Delta w_s^{ll'} | \mathcal{W}_{is}] \leq \mathbb{E}[\Delta w_s^{ll'} | z_s, y_{is}^{l'} = 1]$ . Indeed, for a group of workers having access to information on  $z_s$ , those choosing  $l$  must expect a larger wage difference  $\Delta w_s^{ll'}$  than those choosing  $l'$ . These two inequalities are generally strict and, as a result, the bounds in equations (10) and (11) only partially identify  $\Delta\theta_{ll'}$ . However, the following corollary describes a case where these bounds coincide, implying  $\Delta\theta_{ll'}$  is point identified.

**Corollary 1** *Assume equations (1) to (6) hold. Then, the bounds in equations (10) and (11) imply  $\Delta\theta_{ll'} = \Delta\kappa^{ll'}$  if  $\mathbb{E}[\Delta w_s^{ll'} | z_s] = \mathbb{E}[\Delta w_s^{ll'} | \mathcal{W}_{is}]$ .*

We prove Corollary 1 in Appendix B.2. Corollary 1 strengthens the assumptions in Theorem 1 by requiring that  $\mathbb{E}[\Delta w_s^{ll'} | z_s] = \mathbb{E}[\Delta w_s^{ll'} | \mathcal{W}_{is}]$ . This requirement implies  $z_s$  contains all the information worker  $i$  of type  $s$  has about the wage difference between locations  $l$  and  $l'$ . In this case, equations (10) and (11) jointly hold only when  $\Delta\theta_{ll'}$  equals its true value  $\Delta\kappa^{ll'}$ .

Theorem 1 holds for any two locations  $l$  and  $l'$ . Given the normalization  $\kappa^1 = 0$ , choosing  $l' = 1$  yields  $\Delta\theta_{l1} = \theta_l$ , and equations (10) and (11) identify bounds on  $\theta_l$ , with point identification if Corollary 1 holds. Thus, the bounding inequalities will point identify the amenity in every location in the worker's choice set as long as the researcher observes (and uses in the moment inequalities) all variables belonging to workers' information sets. This is true regardless of how precise workers' wage expectations are.

### 3.1.2 Odds-based Moment Inequalities

For any  $l$  and  $l'$  in the worker's choice set and a random vector  $z_s$ , we define the moment

$$m_o^{ll'}(z_s, \Delta\theta_{ll'}) \equiv \mathbb{E}[y_{is}^l \exp(-(\Delta\theta_{ll'} + \alpha\Delta w_s^{ll'})) - y_{is}^{l'} | z_s]. \quad (12)$$

Theorem 2, from Dickstein et al. (2023), establishes a key property of this moment when evaluated at  $\Delta\theta_{ll'} = \Delta\kappa^{ll'}$ .

**Theorem 2** *Assume equations (1) to (6) hold. Then,  $m_o^{ll'}(z_s, \Delta\kappa^{ll'}) \geq 0$  if  $z_s \subseteq \mathcal{W}_{is}$ .*

The proof of Theorem 2, which follows Dickstein et al. (2023), is in Appendix A.2. This theorem states that, given equations (1) to (6), the moment in equation (12) is positive when evaluated at  $\Delta\theta_{ll'} = \Delta\kappa^{ll'}$  if, for all workers in the population,  $z_s$  belongs to  $\mathcal{W}_{is}$ . Intuitively, the proof of Theorem 2 also uses the property that the ratio of the probability of choosing  $l'$  to the probability of choosing  $l$ , conditional on  $\mathcal{W}_{is}$ , is concave in  $\Delta\theta_{ll'} + \alpha\mathbb{E}[\Delta w_s^{ll'} | \mathcal{W}_{is}]$ . However, instead of first deriving a moment inequality by bounding this ratio from above by its tangent, as the derivation of the bounding inequality in Section 3.1.1 does, it directly substitutes the worker's expected wage difference between  $l$  and  $l'$  with the corresponding difference in realized wages, and applies Jensen's inequality to conclude that the expression in equation (12) must be weakly positive when evaluated at the true parameter values.

Theorem 2 implies that, if its conditions hold, the set of values of  $\Delta\theta_{ll'}$  for which

$$m_o^{ll'}(z_s, \Delta\theta_{ll'}) \geq 0 \quad (13)$$

includes  $\Delta\kappa^{ll'}$  regardless of the vector  $z_s$  and of the locations  $l$  and  $l'$  being compared. Appendix B.3 shows the inequality in equation (13) implies

$$\frac{\mathbb{E}[y_{is}^l | z_s]}{\mathbb{E}[y_{is}^{l'} | z_s]} \mathbb{E}[(\exp(\alpha\Delta w_s^{ll'}))^{-1} | z_s, y_{is}^l = 1] \geq \exp(\Delta\theta_{ll'}). \quad (14)$$

By swapping the identity of locations  $l$  and  $l'$  in equation (14) and rearranging the terms, we

obtain the inequality

$$\frac{\mathbb{E}[y_{is}^l | z_s]}{\mathbb{E}[y_{is}^{l'} | z_s]} (\mathbb{E}[\exp(\alpha \Delta w_s^{ll'}) | z_s, y_{is}^{l'} = 1])^{-1} \leq \exp(\Delta \theta_{ll'}). \quad (15)$$

Equations (14) and (15) identify bounds on  $\Delta \theta_{ll'}$ . To gain intuition on these bounds, consider the case in which  $z_s = \mathcal{W}_{is}$ , allowing to rewrite equations (14) and (15) as

$$\frac{\mathbb{E}[y_{is}^l | z_s]}{\mathbb{E}[y_{is}^{l'} | z_s]} (\mathbb{E}[\exp(\alpha \Delta w_s^{ll'}) | z_s])^{-1} \leq \exp(\Delta \theta_{ll'}) \leq \frac{\mathbb{E}[y_{is}^l | z_s]}{\mathbb{E}[y_{is}^{l'} | z_s]} \mathbb{E}[(\exp(\alpha \Delta w_s^{ll'}))^{-1} | z_s], \quad (16)$$

where the lower bound is weakly smaller than the upper bound due to Jensen's inequality. The two inequalities in equation (16) are generally strict and, as a result, these bounds only partially identify  $\Delta \theta_{ll'}$ . However, the following corollary describes a case where these bounds coincide, implying  $\Delta \theta_{ll'}$  is point identified.

**Corollary 2** *Assume equations (1) to (6) hold. Then, the bounds in equations (14) and (15) imply  $\Delta \theta_{ll'} = \Delta \kappa^{ll'}$  if  $\Delta w_s^{ll'} = \mathbb{E}[\Delta w_s^{ll'} | \mathcal{W}_{is}]$ .*

We prove Corollary 2 in Appendix B.4. Corollary 2 strengthens the assumptions in Theorem 2 by requiring workers to be perfectly informed about  $\Delta w_s^{ll'}$ ,  $\Delta w_s^{ll'} = \mathbb{E}[\Delta w_s^{ll'} | \mathcal{W}_{is}]$ . In this case, the inequality in equation (13) only holds if  $\Delta \theta_{ll'}$  equals its true value  $\Delta \kappa^{ll'}$ .

Theorem 2 and Corollary 2 hold for any locations  $l$  and  $l'$ . Setting  $l' = 1$ , they indicate when the inequality in equation (13) partially or point identifies location  $l$ 's amenity  $\theta_l$ . Thus, the odds-based inequalities will point identify the amenity in every location in the worker's choice set as long as workers forecast wages without error. This is true regardless of the extent to which the researcher observes workers' information sets.

### 3.2 First-Step Moment Inequalities

For any two locations  $l$  and  $l'$  in the worker's choice set, distinct types  $s$  and  $r$ , random vectors  $z_s$  and  $z_r$ , and a scalar random variable  $e_{isjr}^{ll'}$ , we define the moment

$$\begin{aligned} \mathbb{M}^{ll'}(z_s, z_r, \theta_\alpha) \equiv \\ \mathbb{E}[y_{is}^l y_{jr}^l + y_{is}^{l'} y_{jr}^{l'} - y_{is}^l y_{jr}^{l'} \exp(-e_{isjr}^{ll'}) (2 + 2e_{isjr}^{ll'} - \theta_\alpha (\Delta w_s^{ll'} + \Delta w_r^{ll'})) | z_s, z_r]. \end{aligned} \quad (17)$$

A key feature of this moment is that it depends on  $\theta_\alpha$  but not on  $(\theta_2, \dots, \theta_L)$ . Theorem 3 establishes a property of the moment in equation (17) when evaluated at  $\theta_\alpha = \alpha$ .

**Theorem 3** *Assume equations (1) to (6) hold. Then,  $\mathbb{M}^{ll'}(z_s, z_r, \alpha) \geq 0$  if  $e_{isjr}^{ll'} \subseteq \mathcal{J}_{is} \cup \mathcal{J}_{jr}$ ,  $z_s \subseteq \mathcal{W}_{is}$ , and  $z_r \subseteq \mathcal{W}_{jr}$ .*

The proof of Theorem 3 is in Appendix A.3. Given equations (1) to (6), Theorem 3 lists sufficient conditions on  $e_{isjr}^{ll'}$ ,  $z_s$ , and  $z_r$  for the moment in equation (17) to be positive when evaluated at  $\theta_\alpha = \alpha$ . To prove Theorem 3, we start with the following two moments:

$$\mathbb{E}[y_{jr}'(y_{is}' - y_{is}^l \exp(-e_{isjr}^{ll'})(1 + e_{isjr}^{ll'} - (\Delta\theta_{ll'} + \theta_\alpha \Delta w_s^{ll'}))) | z_s, z_r], \quad (18a)$$

$$\mathbb{E}[y_{is}^l(y_{jr}' - y_{jr}^{l'} \exp(-e_{jris}^{l'l'})(1 + e_{jris}^{l'l'} - (\Delta\theta_{l'l} + \theta_\alpha \Delta w_r^{l'l'}))) | z_s, z_r], \quad (18b)$$

which are identical except that the indices of the locations  $l$  and  $l'$ , and of the workers  $is$  and  $jr$ , are swapped. Loosely speaking, the moment function in equation (18a) compares the utility of choosing  $l$  over  $l'$  for a worker  $i$  of type  $s$ , and that in equation (18b) compares the utility of choosing  $l'$  over  $l$  for a worker  $j$  of type  $r$ . If the approximation points  $e_{isjr}^{ll'}$  and  $e_{jris}^{l'l'}$  coincide, the sum of both moments in equation (18) results in the moment in equation (17). This moment does not depend on location-specific amenities as all workers in the population of interest, regardless of their type, have a shared valuation of the amenities in every location. Thus, the amenity differences between locations  $l$  and  $l'$  cancel when adding both moments in equation (18). Conversely, workers of different types may earn different wages in any given location and, thus, the wage differences do not cancel when the moments in equation (18) are added. To prove that the moment in equation (18) is positive when  $\theta_\alpha = \alpha$ , we show that, under the conditions in Theorem 3, both moments in equation (18) are positive when evaluated at the true parameter values; thus, their sum must also be positive.

Theorem 3 implies that, if its conditions hold, the set of values of  $\theta_\alpha$  for which

$$\mathbb{M}^{ll'}(z_s, z_r, \theta_\alpha) \geq 0 \quad (19)$$

includes  $\alpha$  regardless of  $(z_s, z_r)$ , of the two locations  $l$  and  $l'$  being compared, and of how the approximation points  $e_{isjr}^{ll'}$  are chosen. However, the values of  $\theta_\alpha$  other than  $\alpha$  that satisfy the inequality in equation (19) will depend on those approximation points. Appendix B.5 shows that, when restricting ourselves to approximation points that are a function of  $z_s$ ,  $z_r$  and  $\theta_\alpha$ , the set of values of  $\theta_\alpha$  that satisfy the inequality in equation (19) is minimized when:

$$e_{isjr}^{ll'} = \theta_\alpha \mathbb{E}[0.5(\Delta w_s^{ll'} + \Delta w_r^{l'l'}) | z_s, z_r, y_{is}^l y_{jr}' = 1]. \quad (20)$$

As shown in Appendix B.5, equations (19) and (20) jointly imply that

$$\frac{\mathbb{E}[y_{is}^l y_{jr}']}{\mathbb{E}[0.5(y_{is}^l y_{jr}^l + y_{is}^{l'} y_{jr}^{l'}) | z_s, z_r]} \leq \exp(\theta_\alpha \mathbb{E}[0.5(\Delta w_s^{ll'} + \Delta w_r^{l'l'}) | z_s, z_r, y_{is}^l y_{jr}' = 1]). \quad (21)$$

This inequality yields a lower or an upper bound on  $\theta_\alpha$  depending on whether its right-

hand side is increasing or decreasing in  $\theta_\alpha$ , respectively. Intuitively, if type- $s$  workers are likely to choose location  $l$  whereas type- $r$  workers are likely to choose  $l'$ , as represented by a high value of the ratio on the left-hand side of equation (21), and type- $s$  workers expect their wage to be higher in  $l$  than in  $l'$ , and vice versa, as represented by a positive value of  $\mathbb{E}[0.5(\Delta w_s^{ll'} + \Delta w_r^{ll'})|z_s, z_r, y_{is}^l y_{jr}^{l'} = 1]$ , then  $\theta_\alpha$  cannot be too low. Conversely, if both worker types are still likely to choose locations  $l$  and  $l'$ , respectively, but now type- $s$  workers expect their wage to be lower in  $l$  than in  $l'$ , and vice versa, as represented by a negative value of  $\mathbb{E}[0.5(\Delta w_s^{ll'} + \Delta w_r^{ll'})|z_s, z_r, y_{is}^l y_{jr}^{l'} = 1]$ , then  $\theta_\alpha$  cannot be too high. Hence, positive and negative expected wage differences between locations identify opposite bounds.

Generally, the parameter  $\theta_\alpha$  is partially identified. The following corollary describes a case where moment inequalities of the type in equation (21) point identify  $\theta_\alpha$ .

**Corollary 3** *Assume equations (1) to (6) hold,  $\Delta\kappa^{ll'} = 0$ ,  $e_{isjr}^{ll'} \subseteq \mathcal{J}_{is} \cup \mathcal{J}_{fr}$ ,  $z_s \subseteq \mathcal{W}_{is}$ ,  $z_r \subseteq \mathcal{W}_{jr}$ , and  $\mathbb{E}[\Delta w_s^{ll'}|z_s] = \mathbb{E}[\Delta w_s^{ll'}|\mathcal{W}_{is}] = \mathbb{E}[\Delta w_r^{ll'}|z_r] = \mathbb{E}[\Delta w_r^{ll'}|\mathcal{W}_{jr}] = \Delta\bar{w}$ , for a constant  $\Delta\bar{w}$ . Then, moment inequalities of the type in equation (21) with values of  $\Delta\bar{w}$  of different sign jointly imply  $\theta_\alpha = \alpha$ .*

The proof of Corollary 3 is in Appendix B.6. According to Corollary 3, the conditions required to point identify  $\theta_\alpha$  are stronger than those in Corollary 1 to point identify  $\Delta\theta_{ll'}$  for any locations  $l$  and  $l'$ . Intuitively, while we identify  $\Delta\theta_{ll'}$  using moments analogous to those in equation (18) separately, we identify  $\theta_\alpha$  by summing them. Importantly, the resulting moment in (17) is built with the optimal approximation points in equation (20). These, however, may not coincide with the approximation points in equation (9) that are optimal for each moment separately. As a result, the inequality we use to identify  $\theta_\alpha$  combines two inequalities that, due to the use of different approximation points, are generally weaker than those we use to identify  $\Delta\theta_{ll'}$ . The conditions in Corollary 3 precisely imply that the expressions of the approximation points in equations (9) and (20) coincide, ensuring, in this case, no loss of identification power.

### 3.3 Using the Inequalities for Estimation

Given a confidence interval for  $\theta_\alpha$ , we describe here how we use the moment inequalities in Section 3.1 to compute a confidence interval for the amenity term  $\theta_l$  for a location  $l$ . The procedure we follow to compute a confidence interval for  $\theta_\alpha$  using the inequalities in Section 3.2 is analogous; see Appendix B.7. To simplify the exposition, we describe our inference procedure in terms of the bounding inequalities introduced in Section 3.1.1; in practice, we combine these with the odds-based inequalities in Section 3.1.2.

The moment inequality in equation (8) is specific to locations  $l$  and  $l'$  and a conditioning vector  $z_s$ . Given that conditional moment inequality, we implement the following steps to derive a set of  $k = 1, \dots, K$  unconditional moment inequalities. First, we choose a scalar  $\Delta z_s^{ll'} \subseteq z_s$  that is correlated with  $\Delta w_s^{ll'}$  and that we use in all  $K$  inequalities. Second, we choose a subset  $[\underline{z}_k, \bar{z}_k]$  of the support of  $\Delta z_s^{ll'}$ , and an integer exponent  $d_k$ , that are specific to inequality  $k$ . Given these choices, we build the inequality

$$\mathbb{E}[(y_{is}^{l'} - y_{is}^l \exp(-e_{is}^{ll'})(1 + e_{is}^{ll'} - (\Delta\theta_{ll'} + \alpha\Delta w_s^{ll'})))g_k(\Delta z_s^{ll'})] \geq 0, \quad (22)$$

where the term in parenthesis coincides with the moment function in equation (7) and the instrument  $g_k(\Delta z_s^{ll'})$  is built as follows

$$g_k(\Delta z_s^{ll'}) = \mathbb{1}\{\underline{z}_k < \Delta z_s^{ll'} \leq \bar{z}_k\} |\Delta z_s^{ll'}|^{d_k}. \quad (23)$$

If the conditions in Theorem 1 are satisfied, the inequality in equation (22) holds at  $\Delta\theta_{ll'} = \Delta\kappa^{ll'}$  regardless of the choice of predictor  $\Delta z_s^{ll'}$ , interval limits  $\underline{z}_k$  and  $\bar{z}_k$ , and exponent  $d_k$ . Effectively, we choose a number  $q \in \mathbb{N}$  and build inequalities where  $\underline{z}_k$  and  $\bar{z}_k$  correspond to consecutive elements of the vector of  $q$ -quantiles of the distribution of  $\Delta z_s^{ll'}$  across all types and location pairs. Concerning  $d_k$ , we set it to  $-1$ ,  $0$ , or  $1$ . E.g., if  $q = 2$  and  $d_k = 0$  for all  $k$ , we have  $K = 2$  inequalities and the instruments are  $\mathbb{1}\{\Delta z_s^{ll'} \leq \text{med}(\Delta z_s^{ll'})\}$  and  $\mathbb{1}\{\text{med}(\Delta z_s^{ll'}) > \Delta z_s^{ll'}\}$ . Thus, in this example, the instruments simply split all observations depending on whether  $\Delta z_s^{ll'}$  is above or below median. Similarly, if  $q = 4$  and  $d_k = 0$  for all  $k$ , the instruments split all observations depending on which quartile  $\Delta z_s^{ll'}$  belongs to.

Our confidence interval (CI) for  $\theta_l$  for a location  $l$  depends on the first-step CI for  $\theta_\alpha$ . To compute a 95% CI for  $\theta_l$ , we first compute 96% CIs for  $\theta_l$  conditional on each value of  $\theta_\alpha$  in a 99% CI for this parameter. We denote these as  $\hat{\Theta}_{.96}^l(\theta_\alpha)$ . We then compute the 95% CI for  $\theta_l$  as the union of  $\hat{\Theta}_{.96}^l(\theta_\alpha)$  for each value of  $\theta_\alpha$  in its 99% CI. All our CIs are computed following the moment selection procedure in Andrews and Soares (2010); see Appendix B.8.<sup>15</sup>

## 4 Properties of Moment Inequalities: Simulation

This section uses simulations to illustrate the properties of our two-step estimator. Three main insights emerge. First, consistent with theorems 1 to 3, when the conditions in those theorems are satisfied, our estimator yields intervals that contain the true parameter values

<sup>15</sup>A potential alternative to our estimator is to apply to each element of  $(\theta_\alpha, \theta_2, \dots, \theta_L)$  a procedure that yields valid CIs for projections of partially identified parameters; e.g., Bugni et al. (2016); Kaido et al. (2019). Importantly, although our bounding inequalities are linear, their conditional variance depends on all parameters and, thus, do not satisfy the conditions for the validity of the procedure in Andrews et al. (2023).

even when workers imperfectly forecast wages and the researcher imperfectly observes agents' information sets. Second, the maximum likelihood (ML) estimator is downward biased when the researcher misspecifies agents' information sets and, in this case, the ML estimates are often not contained in the confidence set defined by our moment inequalities. Third, our moment inequalities may yield empty confidence sets when the researcher incorrectly assumes that certain covariates belong to agents' information set, demonstrating the potential of our estimator to test for the true content of such sets.

## 4.1 Simulation Set-up

Workers choose between three locations  $l = \{1, 2, 3\}$  according to the model in Section 2. We simulate data for 6,000,000 workers, each of them of a different type, allowing us to index observations by  $s$ .<sup>16</sup> We set the wage coefficient to  $\alpha = 1$  and the location-specific amenities to  $\kappa^1 = \kappa^2 = 0$  and  $\kappa^3 = 1$ . While the estimator described in Section 3 is valid for any stochastic process for wages and any specification of workers' information sets, we need to set out these model aspects to generate the model-implied choice for all sampled individuals.

We assume the wage of worker  $s$  in location  $l$  is determined by

$$w_s^l = z_{1s}^l + z_{2s}^l + z_{3s}^l, \quad (24)$$

with  $z_{ks}^l$  independent across  $l$ ,  $s$ , and  $k$ , and distributed uniformly with mean  $\mu_k^l$  and support of length  $2\sigma_k$ .<sup>17</sup> We assume every worker  $s$  observes  $(z_{1s}^l, z_{2s}^l)$  for every location  $l$ . Thus, for all  $s$  and  $l$ ,  $\mathbb{E}[w_s^l | \mathcal{W}_s] = z_{1s}^l + z_{2s}^l$ , and  $z_{3s}^l$  equals the expectational error worker  $s$  makes when forecasting  $w_s^l$ . Thus, the larger the value of  $\sigma_3$ , the larger the variance of the expectational error workers make when forecasting wages in every location.

We assume the researcher observes  $(y_s^l, w_s^l, z_{2s}^l)$  for every worker  $s$  and every location  $l$ . Therefore, for all  $s$  and  $l$ ,  $z_{1s}^l$  is a variable used by worker  $s$  when forming their expectations about  $w_s^l$  but unobserved by the researcher. Thus, the larger the value of  $\sigma_1$ , the larger the role in workers' expectations of variables that are unobserved to the researcher.

## 4.2 Simulation Results

Table 1 presents the main simulation results. Section 4.2.1 describes CIs for the wage coefficient  $\theta_\alpha$ , displayed in the column labeled *First Step*, computed using moment inequalities

<sup>16</sup>As the predictable part of wages varies by type (but not across workers of the same type) according to equation (5), the large  $S$  limits the impact of noise in our simulation.

<sup>17</sup>We set  $\mu_1^l = \mu_3^l = 0$  for all  $l$  and  $(\mu_2^1, \mu_2^2, \mu_2^3) = (0, -0.5, -1)$ ; thus, mean wages decline in order from  $l = 1$  to  $l = 3$ . In terms of the dispersion, we set  $\sigma_2 = 4$  and present results for different values of  $\sigma_1$  and  $\sigma_3$ .

of the type introduced in Section 3.2. Section 4.2.2 describes CIs for  $\theta_2$  and  $\theta_3$ , displayed in the columns labeled *Second Step*, and computed using both the first-step CI for  $\theta_\alpha$  and moment inequalities of the type introduced in Section 3.1. Appendix C.1 describes in detail the moment inequalities we use in this simulation exercise.

#### 4.2.1 Confidence Intervals for the Wage Coefficient

Cases 1 and 2 share the feature that  $\sigma_1 = 0$  and, thus, the researcher observes the agent's full information set—as  $z_{1s}^l = 0$  for every location  $l$ , the agent's information set only includes  $z_{2s}^l$ , which equals the wage predictor used by the researcher. As a result, the CIs for  $\theta_\alpha$  we obtain are very tight around its true value  $\alpha = 1$ . To understand why we obtain such intervals, one should bear in mind two aspects of our setting. First, we build first-stage inequalities separately for each pair of locations; thus, as  $\kappa^1 = \kappa^2$  in our simulated setting, the inequality with location indices equal  $l = 1$  and  $l' = 2$  (or vice versa) verifies that  $\Delta\kappa^{ll'} = 0$ .<sup>18</sup> Furthermore, when building the inequality corresponding to any location pair  $l$  and  $l'$ , we match each worker  $s$  with a worker  $r$  such that  $\mathbb{E}[\Delta w_s^{ll'} | \Delta z_{2s}^{ll'}]$  is as close to  $\mathbb{E}[\Delta w_r^{ll'} | \Delta z_{2r}^{ll'}]$  as the simulated dataset permits.<sup>19</sup> As a result, consistently Corollary 3, the moment inequalities we use to compute the CIs for  $\theta_\alpha$  reported in Table 1 are very close to point identifying this parameter whenever  $\mathbb{E}[w_s^l | \mathcal{W}_s] = \mathbb{E}[w_s^l | \Delta z_{2s}^{ll'}]$  for all  $l$  and  $s$ . This extra condition is met precisely when  $\sigma_1 = 0$ .

Cases 3 and 4 share the feature that  $\sigma_1 > 0$  and, thus, the researcher only observes part of the agent's information set—the true information set is  $(z_{1s}^l, z_{2s}^l)$  for every  $l$ , but the researcher only observes  $z_{2s}^l$ . By using information on only a subset of the agent's information set, the researcher obtains a wider confidence interval for  $\theta_\alpha$ . In case 5, the researcher wrongly assumes the agent has perfect information on wages. In this case, the confidence interval for  $\theta_\alpha$  only includes one value ( $\theta_\alpha = 0.87$ ), different from the true value  $\alpha = 1$ .<sup>20</sup>

#### 4.2.2 Confidence Intervals for Amenities

Consistently with Corollary 1, the bounding inequalities point identify  $\theta_2$  and  $\theta_3$  when the agent's information set is perfectly observed by the researcher; i.e., when  $\sigma_1 = 0$ , as in cases

<sup>18</sup>In Table C.2 in Appendix C.3, we present analogous results for simulation settings with different values of  $\kappa^1$ ,  $\kappa^2$ , and  $\kappa^3$ . The CI for  $\theta_\alpha$  remains tight as long as  $\Delta\kappa^{ll'} = 0$  for at least two locations  $l$  and  $l'$ , and becomes wider as the differences in  $\kappa^l$  across all locations increase.

<sup>19</sup>Both  $w_s^l$  and  $z_{2s}^l$  are continuous. Thus, for any  $l$  and  $l'$  and worker  $s$ , there generally exists no worker  $r$  for which  $\mathbb{E}[\Delta w_s^{ll'} | z_{2s}^l]$  equals  $\mathbb{E}[\Delta w_r^{ll'} | z_{2r}^l]$ . In Table C.1 in Appendix C.2, we show the CI for  $\theta_\alpha$  becomes wider as the mean difference between  $\mathbb{E}[\Delta w_s^{ll'} | z_{2s}^l]$  and  $\mathbb{E}[\Delta w_r^{ll'} | z_{2r}^l]$  for matched workers  $s$  and  $r$  increases.

<sup>20</sup>Molinari (2020) and Andrews and Kwon (2024) show that misspecified models may yield confidence sets that are tight without including the true parameter value. We show in Table C.3 in Appendix C.4 that, in case 5, the CI for  $\theta_\alpha$  becomes empty as we increase the number of instruments used to form our inequalities.

Table 1: Simulation Results - Moment Inequality Confidence Intervals

| Case | $\sigma_1$ | $\sigma_3$ | $z_s^l$    | First Step      | Second Step                    |                                                 |                                              |
|------|------------|------------|------------|-----------------|--------------------------------|-------------------------------------------------|----------------------------------------------|
|      |            |            |            | $\theta_\alpha$ | Mom. Ineq.                     | $\theta_2$                                      | $\theta_3$                                   |
| 1    | 0          | 0          | $z_{2s}^l$ | [1, 1.02]       | Bounding<br>Odds-based<br>Both | [0, 0]<br>[0, 0]<br>[0, 0]                      | [1, 1]<br>[1, 1]<br>[1, 1]                   |
| 2    | 0          | 1          | $z_{2s}^l$ | [1, 1.01]       | Bounding<br>Odds-based<br>Both | [0, 0]<br>[-0.33, 0.32]<br>[0, 0]               | [1, 1]<br>[0.68, 1.33]<br>[1, 1]             |
| 3    | 1          | 0          | $z_{2s}^l$ | [0.82, 1.29]    | Bounding<br>Odds-based<br>Both | [-0.31, 0.31]<br>[0, 0]<br>[0, 0]               | [0.70, 1.30]<br>[1, 1.01]<br>[1, 1.01]       |
| 4    | 1          | 1          | $z_{2s}^l$ | [0.82, 1.31]    | Bounding<br>Odds-based<br>Both | [-0.31, 0.31]<br>[-0.38, 0.39]<br>[-0.31, 0.31] | [0.69, 1.31]<br>[0.68, 1.45]<br>[0.69, 1.31] |
| 5    | 0          | 1          | $w_s^l$    | [0.87, 0.87]    | Bounding<br>Odds-based<br>Both | [-0.05, -0.01]<br>$\emptyset$<br>$\emptyset$    | [0.85, 0.88]<br>$\emptyset$<br>$\emptyset$   |

The true parameter values are  $\alpha = 1$ ,  $\kappa^2 = 0$ , and  $\kappa^3 = 1$ . The column  $\theta_\alpha$  contains 95% CIs based on the estimator described in Section 3.2. The columns  $\theta_2$  and  $\theta_3$  contain 95% CIs based on the estimators described in Section 3.1. The rows labeled *Bounding* use the inequalities in Section 3.1.1; those labeled *Odds-based* use the inequalities in Section 3.1.2; and those labeled *Both* combine both inequalities. CIs are computed using moment selection procedure in Andrews and Soares (2010). See Appendix C.1 for more details.

1 and 2. When  $\sigma_1$  is positive, as in cases 3 and 4, the confidence intervals built using the bounding inequalities alone include parameter values in addition to the true ones. As a comparison of cases 1 and 2, or cases 3 and 4, illustrates, this holds regardless of how large the agent’s expectational errors are. Intuitively, the bounding inequalities are linear in the expectational errors, which average out to zero regardless of their variance, making the value of  $\sigma_3$  irrelevant for the identified set defined by these inequalities. Conversely, the bounding inequalities are weaker the larger the mismatch between the approximation point in equation (9) and the agent’s true expected utility difference between the two locations  $l$  and  $l'$  being compared, and this mismatch is larger in expectation the larger the importance of those variables the agent knows but the researcher does not observe; i.e., the larger  $\sigma_1$  is.<sup>21</sup>

The length of the confidence intervals on  $\theta_2$  and  $\theta_3$  defined by the odds-based moment inequalities increases in the importance of the worker’s expectational errors; i.e., increases in  $\sigma_3$ , as a comparison of cases 1 and 2, or cases 3 and 4, illustrates. Intuitively, these

<sup>21</sup>As described in Appendix C.1, all CIs in Table 1 are computed approximating the conditional expectations in equations (9) and (20) through linear regressions. In Table C.6 in Appendix C.6, we show the CIs become wider as we use approximation points that depart further from those in equations (9) and (20).

inequalities are convex in the agent’s expectational errors and, thus, Jensen’s inequality implies that the moment at any given parameter value is larger in expectation the larger the value of  $\sigma_3$ , making the resulting moment inequality weaker. When workers make no expectational errors (i.e., when  $\sigma_3 = 0$ ) and the first-step confidence interval for  $\theta_\alpha$  equals its true value  $\alpha$ , the odds-based moment inequalities point identify the amenity parameters  $\theta_2$  and  $\theta_3$ , as predicted by Corollary 2. When  $\sigma_3 = 0$  but the confidence interval for  $\theta_\alpha$  includes values other than  $\alpha$ , the odds-based moment inequalities may still point identify the amenities (as in cases 1 and 3), but will not do so always.<sup>22</sup>

Since the bounding inequalities are insensitive to agents’ expectational errors (i.e., insensitive to  $\sigma_3$ ) and the odds-based inequalities are partially insensitive to agents having information the researcher does not observe (i.e., partially insensitive to  $\sigma_1$ ), there are advantages from combining both types of inequalities in estimation. Cases 2 and 3 show that, when either  $\sigma_1 = 0$  or  $\sigma_3 = 0$ , combining bounding and odds-based inequalities point identifies  $\theta_2$  and  $\theta_3$ , although, when considered in isolation, neither of these two inequalities point identifies these parameters in both cases.

Case 4 is likely the most empirically relevant: the agent’s information set is partly unobserved (i.e.,  $\sigma_1 > 0$ ) and the agent predicts wages with error (i.e.,  $\sigma_3 > 0$ ). Even in this scenario, our estimator yields confidence intervals that contain the true parameter values. In this particular case where  $\sigma_1$  and  $\sigma_3$  are equal, the odds-based inequalities are redundant, as they yield larger intervals than those obtained from the bounding inequalities.

Case 5 shows the bounding inequalities may fail to produce empty confidence intervals when the researcher wrongly assumes workers have complete information—although, as discussed in footnote 20, the confidence interval for  $\theta_\alpha$  is empty when we use more instruments. Conversely, the odds-based inequalities alone, or when used jointly with the bounding inequalities, produce empty confidence intervals for the amenity parameters  $\theta_2$  and  $\theta_3$  even when the confidence interval for  $\theta_\alpha$  is non-empty.

### 4.3 Alternative Estimators

*Maximum Likelihood.* Table 2 reports ML estimates. Specifically, given a choice of wage predictor  $z_s^l$ , we compute the ML estimator of  $(\theta_\alpha, \theta_2, \theta_3)$  assuming that, for every location  $l$ ,  $z_s^l$  is all the information every worker  $s$  has on  $w_s^l$ ; that is,

$$\operatorname{argmax}_{(\theta_\alpha, \theta_2, \theta_3)} \left\{ \sum_{s=1}^S \sum_{l=1}^3 \mathbb{1}\{y_s^l = 1\} \ln \left( \frac{\exp(\theta_l + \theta_\alpha \mathbb{E}[w_s^l | z_s^l])}{\sum_{l'=1}^3 \exp(\theta_{l'} + \theta_\alpha \mathbb{E}[\Delta w_s^{l'} | z_s^{l'}])} \right) \right\} \quad \text{with } \theta_1 = 0. \quad (25)$$

---

<sup>22</sup>For example, in unreported results, we observe that the confidence intervals for  $\theta_2$  and  $\theta_3$  defined by the odds-based inequalities include values other than the true ones when  $\sigma_1 = 2$  and  $\sigma_3 = 0$ .

Table 2: Simulation Results - Maximum Likelihood Estimator

| Case | $\sigma_1$ | $\sigma_3$ | $z_s^l$    | $\alpha$ | $\kappa^2$ | $\kappa^3$ |
|------|------------|------------|------------|----------|------------|------------|
| 1    | 0          | 0          | $z_{2s}^l$ | 1        | 0          | 1          |
| 2    | 0          | 1          | $z_{2s}^l$ | 1        | 0          | 1          |
| 3    | 1          | 0          | $z_{2s}^l$ | 0.91     | 0          | 0.92       |
| 4    | 1          | 1          | $z_{2s}^l$ | 0.91     | 0          | 0.92       |
| 5    | 0          | 1          | $w_s^l$    | 0.87     | -0.03      | 0.87       |

The true parameter values are  $\alpha = 1$ ,  $\kappa^2 = 0$ , and  $\kappa^3 = 1$ . Estimates are computed by solving the maximization problem in equation (25). Given the large sample size  $S = 6,000,000$ , unreported standard errors are always smaller than 0.001.

If the researcher’s specified wage predictor  $z_s^l$  equals  $z_{2s}^l$  for every location  $l$  and worker  $s$ , the ML estimator is consistent if and only if  $\sigma_1 = 0$ , as in cases 1 and 2, as only then does the worker’s wage expectation coincides with the researcher’s assumed one; i.e., only then  $\mathbb{E}[w_s^l | \mathcal{W}_s] = \mathbb{E}[w_s^l | z_s^l]$  for every  $s$  and  $l$ . Conversely, if  $z_s^l = z_{2s}^l$  and  $\sigma_1 > 0$ , as in cases 3 and 4, the worker’s true expectation and the researcher’s assumed one do not coincide and, as a result, the ML estimator of  $\theta_\alpha$  is biased towards zero; i.e., it underestimates the importance of expected wages in the worker’s utility. In case 5, the researcher assumes workers have perfect information (i.e.,  $z_s^l = w_s^l$  and, thus,  $\mathbb{E}[w_s^l | z_s^l] = w_s^l$ ) but, contrary to that assumption, workers make forecasting errors (i.e.,  $\sigma_3 > 0$ ), and the ML estimator is also biased.<sup>23</sup>

A comparison of the estimates in tables 1 and 2 yields three conclusions. First, when the ML estimator identifies the true parameter values (i.e., when  $\sigma_1 = 0$ , as in cases 1 and 2), the CIs defined by the bounding inequalities either include only the true parameter values (as it is the case for  $\theta_2$  and  $\theta_3$ ) or are very tight around them (as it is the case for  $\theta_1$ ). Thus, very little identification power is lost when using our moment inequality estimator instead of the ML estimator in those cases in which this latter one is valid. Second, when the ML estimator does not identify the true parameter values (i.e., in cases 3 to 5), our moment inequality estimator still yields CIs that contain the true parameter values. Moreover, as illustrated by case 3, the CIs produced by our moment inequality estimator may not include the corresponding ML estimates for some of the parameters; e.g., for  $\theta_2$  and  $\theta_3$ .

*Alternative moment inequality estimator.* In Table C.5 in Appendix C.5, we apply an alternative inference procedure that only relies on the second-step inequalities described in Section 3.1. Specifically, we use the inequalities in equations (8) and (13) to compute confidence sets for the vector  $(\theta_\alpha, \theta_l)$  for each location. In our simulation setting, we obtain

<sup>23</sup>Cases 3 and 5 show that, when the difference between the worker’s true expectation and the researcher’s assumed expectation have identical marginal distribution (i.e., if  $\sigma_1 = \sigma_3$ ), the bias in the ML estimator is smaller if the researcher assigns workers an information set that is too small than if they assign them an information set that is too large. Dickstein and Morales (2018) find a similar pattern in their setting.

in this way a two-dimensional confidence set for  $(\theta_\alpha, \theta_2)$  and another one for  $(\theta_\alpha, \theta_3)$ . After projecting each of these two-dimensional confidence sets on each of their elements, we obtain a CI for each of the amenities  $\theta_2$  and  $\theta_3$ , and two CIs for  $\theta_\alpha$ . We then report a randomly chosen CI for  $\theta_\alpha$  among the two available ones. This alternative procedure yields results similar to those in Table 1 except in case 3, where the fact that this alternative estimator adds information from the odds-based inequalities for the identification of  $\theta_\alpha$  results in a tighter CI for this parameter.<sup>24</sup>

## 5 Empirical Application

In our empirical application, we study internal migration in Brazil. We describe our data in Section 5.1, discuss our estimation approach and estimation results in Section 5.2, and present tests of the content of workers' information sets in Section 5.3. In Section 5.4, we evaluate the effect of counterfactual changes in information sets and migration costs.

### 5.1 Data

Our main data source is the *Relação Anual de Informações Sociais* (RAIS), an administrative dataset that includes information on workers and establishments in the Brazilian formal labor market. We use the establishment's location (microregion, the closest equivalent to a commuting zone in the Brazilian administrative map) and sector (industry) to define labor markets, and we measure workers' wages by annualizing their average monthly earnings.

We restrict our sample to workers with similar demographic characteristics. Specifically, we study workers aged 25-64 with at least a high school degree identified as male and white. These are, respectively, the largest education, gender, and race categories in RAIS. Since RAIS only covers formal employment, workers that are not employed or hold informal jobs are absent from the dataset. Hence, our conclusions are limited to formal workers, and we accordingly restrict our sample to individuals with a persistent attachment to the formal labor market, selecting only those recorded in RAIS for at least seven years during our sample period, which spans between 2002 and 2011.

To ensure we observe a large number of individuals per market, we focus on 1,000 labor markets consisting of all combinations of the 50 microregions (out of 558) and 20 sectors (out

---

<sup>24</sup>Unreported simulation results show that, when the agent's information set is heterogeneous across *choices*, the CI for  $\theta_\alpha$  produced by this alternative moment inequality estimator is sensitive to which of the two available CIs for this parameter is chosen. Alternatively, one may find a way of combining the information contained in the two CIs for  $\theta_\alpha$  that results in a tighter CI than when simply one of them is randomly chosen; we leave the design of such an approach for future work.

of 51) with the largest total employment reported in RAIS.<sup>25</sup> We obtain the data we use in estimation by randomly sampling one million individuals per year among those employed in the 1,000 labor markets of interest. Appendix D provides more details on the RAIS data and the construction of our sample, and reports summary statistics on migration rates.

## 5.2 Estimation of Model Parameters

In Section 5.2.1, we describe how we implement the estimator described in Section 3. In Section 5.2.2, we present our estimates and compare them to those obtained using alternative estimators.

### 5.2.1 Implementation of Moment Inequalities

We estimate the parameters of the model described in Section 2, with the type  $s$  of each worker defined by their sector. While we assume that the wage coefficient  $\alpha$  is constant across all sampled workers, we let the location-specific fixed effects in the vector  $\kappa$  vary by year  $t$  and the worker’s prior location. We thus accommodate for unobserved expectations about migration costs, amenities, and price levels that may vary over time and across workers with different locations of origin. For simplicity, we refer below to these origin-by-destination-by-year unobserved effects as *amenities*.

Key variables entering the moment inequalities introduced in Section 3 are the location-by-sector-by-time wage shifters  $w_{st}^l$ . We estimate these using data on the worker’s observed wages. Specifically, we assume the log wage a worker would have obtained if employed in a particular sector, region, and year depends on individual worker characteristics (age and sectoral experience) with sector-specific coefficients, on region-by-sector-by-year fixed effects, on worker-by-sector fixed effects, and a residual assumed to be unknown to the worker when choosing its labor market of employment; see Appendix E.1 for details. For each sector  $s$  and year  $t$ , the estimates of the region-by-sector-by-year fixed effects correspond to the wage shifters  $w_{st} = (w_{st}^1, \dots, w_{st}^L)$  entering our inequalities. These shifters account for supply and demand factors that impact the wages of all workers in a labor market. Since every other term in the wage equation is either portable across locations (i.e., not location-dependent) or assumed to be unknown to workers’ when making their migration decision, these shifters determine the potentially predictable differences in the wages workers receive in different labor markets. For simplicity, we refer to these below as *wages*.

---

<sup>25</sup>As pointed out in footnote 10, our moment inequality estimator only requires the researcher to observe a subset of the labor markets workers choose from. The informal sector and all microregion-sector pairs not considered in our analysis can thus be thought of as being outside of the worker’s choice set or as belonging to the part of the worker’s choice set not considered in our inequalities.

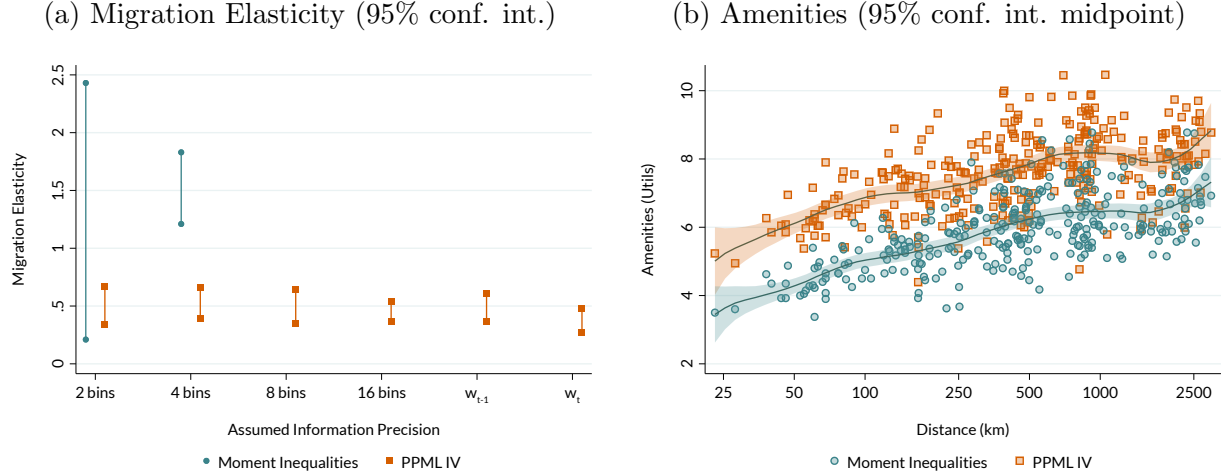
As discussed in Section 3.3, our inequalities require specifying a wage predictor  $z_{st}^l$  for each sector  $s$ , region  $l$ , and year  $t$ . To build the predictors we use in our application, we implement the following procedure. First, given a number  $b \in \mathbb{N}$ , we compute the vector of  $b$ -quantiles of the distribution of  $w_{st-1}^l$  across all sectors and locations. Second, we build as many “bins” as intervals can be constructed using two consecutive elements of this vector of quantiles. Third, we identify the bin to which each labor market belongs. Fourth, for each market, we compute  $z_{st}^l$  as the average wage in year  $t - 1$  across all labor markets that belong to the same bin as the labor market defined by region  $l$  and sector  $s$ . Note that, as  $b$  increases, the wage predictors become closer on average to the  $t - 1$  wages. Thus, a larger  $b$  can be interpreted as workers having more precise information on  $t - 1$  wages. For example, if  $b = 2$ , the wage predictor described above implies assuming workers can determine if lagged wages in any given labor market are above or below the median. If  $b = 4$ , workers can determine the quartile of the  $t - 1$  wage distribution to which each labor market belongs. As  $b \rightarrow \infty$ ,  $z_{st}^l$  and  $w_{st-1}^l$  coincide in every labor market (workers know the entire lagged wage distribution). We provide additional details on the implementation of the inequalities in Appendix E.2.

### 5.2.2 Estimation Results

*First-step estimates: wage coefficient.* Panel (a) in Figure 1 reports 95% CIs for the wage coefficient  $\alpha$  under different informational assumptions. When we set  $b = 2$ , and thus assume workers can at least determine whether lagged wages in any given labor market are above or below the median of the distribution of wages across all labor markets (the “2 bins” case), we obtain a 95% CI that equals  $[0.24, 2.44]$ . The large width of this interval reflects that the dummy variable indicating whether lagged wages are below or above the median is only loosely correlated with current wages. When we increase the assumed precision of workers’ information and impose instead that workers can at least determine the quartile to which lagged market-specific wages belong (“4 bins”), we obtain a tighter interval equal to  $[1.21, 1.83]$ . Assuming workers can classify locations according to more detailed quantiles of the wage distribution, or that they know the actual value of lagged or current wages, yields empty CIs. Thus, we reject that workers know lagged location-specific wages with any level of precision above quartiles. Below, we use  $[1.21, 1.83]$  as our preferred set estimator of  $\alpha$ .

For comparison, we also include 95% CIs computed using the two-step PPML-IV estimator from Artuç and McLaren (2015). As discussed in Appendix E.3, this estimator yields point estimates of the wage coefficient at the expense of assuming that all workers in the same sector in a period  $t$  (regardless of their location of residence) have exactly the same information set, and consequently also the same wage expectations. This is a stronger assumption than the one required for our moment inequalities to bound the true parameter

Figure 1: Migration Elasticity and Amenities from Moment Inequalities vs. PPML-IV



Panel (a) reports 95% confidence intervals of the wage coefficient  $\alpha$  under different estimation procedures and informational assumptions. The blue circles delimit the moment inequality confidence intervals. The absence of circles for cases above 4 bins reflects that the associated intervals are empty. The orange squares mark the PPML-IV confidence intervals. Informational assumptions are stronger as we move towards the right of the horizontal axis. In panel (b), the blue circles indicate the midpoints of the moment inequality 95% confidence interval for parameters  $\kappa_{nt}^l$ , for  $t = 2011$ , and the orange squares indicate the PPML-IV estimates. The fit lines are kernel-weighted local polynomial estimates, with the shaded area representing 95% confidence bands.

value, which requires the researcher to specify a (possibly different) variable that belongs to every worker's information set, but does not restrict the additional information each worker may have, which may vary flexibly across workers and labor markets.<sup>26</sup> The PPML-IV estimator yields confidence intervals for the wage coefficient that generally do not overlap with those generated by our moment inequality estimator: while the PPML-IV estimator yields confidence intervals between 0.3 and 0.6, the lower bound in our preferred moment inequality confidence interval is 1.21. Relative to our moment inequality estimator of the wage coefficient, the PPML-IV estimator thus underestimates the value workers assign to the expected monetary returns of migration relative to other non-income location payoffs.

*Second-step estimates: amenities.* Panel (b) in Figure 1 illustrates the moment inequality estimates of the parameters  $\kappa_{nt}^l$  for the year  $t = 2011$  and all origin  $n$  and destination  $l$  locations in our sample. Specifically, this panel displays midpoints of the 95% moment inequality CI for each amenity term  $\kappa_{nt}^l$ .<sup>27</sup> For comparison, we also display estimates of these

<sup>26</sup>E.g., the moment inequality confidence interval that uses quartiles of lagged wages as the relevant wage predictor is valid if workers, even within the same sector, location, and period, have different information, as long as all workers can at least classify labor markets into quartiles on the basis of lagged market wages.

<sup>27</sup>Appendix E.4 displays the corresponding CIs, revealing they are tight. We report the estimates for the 292 directed pairs of origin-destinations with enough migration events in our sample to obtain an estimate

amenities computed using the PPML-IV estimator. Although we estimate each parameter  $\kappa_{nt}^l$  without imposing any restriction on their variability, our estimated parameters tend to increase in the distance between locations  $n$  and  $l$ , consistently with these parameters accounting for migration costs in our model. The differences in levels between the PPML-IV estimates and the moment inequalities estimates are substantial, with the latter being on average 21% smaller than the former. Moreover, if we convert migration costs into their log-wage equivalents by dividing them by the estimates of the wage coefficient  $\alpha$  produced by each estimation method, we find that the moment inequality estimates are 74% smaller. In sum, our empirical analysis indicates that estimation procedures commonly used in the migration literature yield upward biased estimates of migration costs or, more generally, of the relative importance of non-wage variables in migration decisions.

### 5.3 Tests of Information Heterogeneity

As panel (a) in Figure 1 shows, the moment inequality 95% CI for the wage coefficient  $\alpha$  is non-empty when we assume all workers know the quartile to which lagged wages belong, but becomes empty when we assume *all* workers can classify *all* markets into eight (or more) bins. It is possible, however, that some workers are more informed than others, or that they have more information about some markets than others.

In this section, we explore whether the migration patterns in our sample are consistent with workers having more precise information about wages in some, but not all, markets. We do so by checking whether the 95% CI for  $\alpha$  defined by our moment inequalities becomes empty when we assume that some worker groups have additional wage information about specific groups of labor markets. Specifically, we consider worker groups defined by the population and internet penetration in their prior location of residence, and groups of labor markets defined by their distance and past migration flows from the worker’s location, as well as by the population and internet penetration in the market’s location.<sup>28</sup>

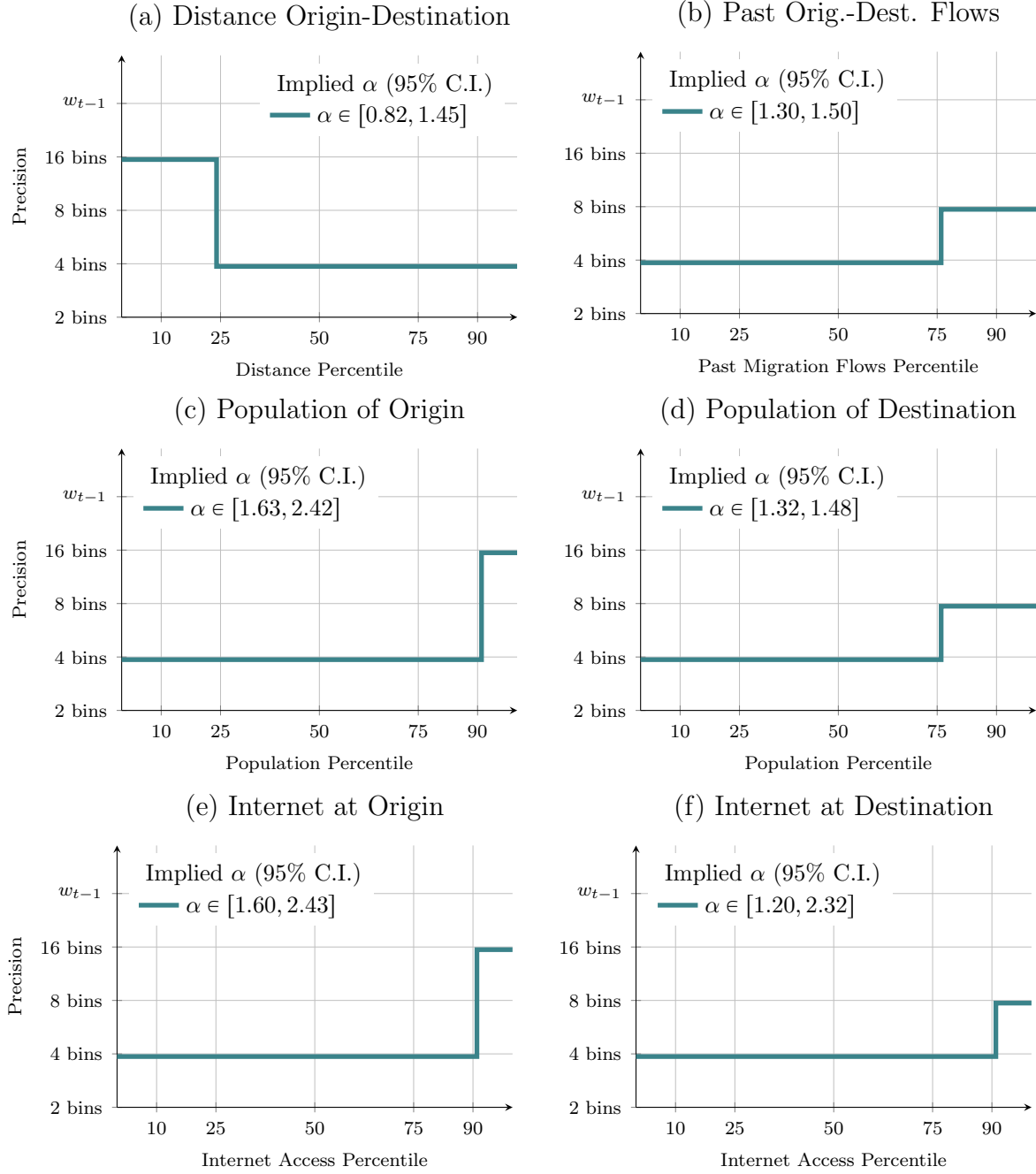
We implement the same testing approach for each dimension of heterogeneity described above. First, we classify workers or markets into six intervals delimited by the 10th, 25th, 50th, 75th, and 90th percentiles of the distribution of workers and markets along the corresponding dimension. Then, we order these intervals according to the direction along which

---

from both the PPML-IV and moment inequalities estimators.

<sup>28</sup>The distance between any two locations equals the geodesic distance between their centroids. Past migration flows are measured as the total number of workers recorded in RAIS as having migrated between any two locations in the three years prior to our sample period (1999-2001). The population of each location is computed as the total employment in RAIS in the period 1999-2001. The share of households with internet access in each location equals the average share of households with broadband internet access between 2007 and 2011, the period for which this information is available (see Appendix D).

Figure 2: Testing for Heterogeneous Information Sets



This figure displays patterns of information precision that cannot be rejected in the data. Each panel studies patterns along a key dimension, including distance, past migration flows, the population of origin and destination, and the share of households with internet access in the origin and destination. Patterns that cannot be rejected, yielded by the testing procedure described in the main text, are shown in solid lines. We test each hypothesis by building an instrument function that defines wage proxies according to the assumed pattern of information precision. These wage proxies reflect the characteristics of the origin and destination labor markets.

we hypothesize information may be more precise. For example, we classify workers into intervals depending on the internet penetration in the location of residence and order these intervals from higher to lower internet penetration. Consistently with our finding in panel (a) in Figure 1, we start from a baseline information set according to which all workers can classify all markets into quartiles, and test whether workers in the first interval can further classify markets into eight bins. This translates into wage predictors with different levels of precision across different combinations of workers and markets, affecting the values of the instruments in equation (23) and, thus, changing our moment inequalities. If the resulting 95% CI for  $\alpha$  is empty, we reject that assumption and end the testing procedure. If it is not empty, we increase the level of precision to 16 bins on that first interval and perform a new test. Calling  $B_j$  the maximum precision level tested and not rejected for the  $j$ th interval, the next iteration maintains  $B_j$  on that interval and searches for the maximum level of precision in the interval  $j + 1$ , up to precision  $B_j$ . This procedure yields a weakly monotone information schedule along each dimension.

Figure 2 displays our results. Panel (a) shows we cannot reject workers are better informed about wages in labor markets within the 25th percentile of distance (383 km) of their location of residence. For those markets, we cannot reject they know  $t - 1$  wages with the precision equivalent to 16 bins. As discussed in Section 5.4.1 below, migration rates in our model increase in workers' information. Thus, the well-known fact that migration rates decrease in distance (Beine et al., 2016) may be due less than previously thought to migration costs increasing in this dimension, and more to the role information frictions play in migration choices.<sup>29</sup> In panel (b), we observe that past migration flows between two locations are positively correlated with the information residents in one location have about wages in the other one. As information frictions in our model decrease migration flows, this finding may explain why workers from a particular origin tend to persistently migrate to the same destinations, providing an explanation for the impact of enclaves on migration flows (Munshi, 2020). Panels (c) and (d) show that workers living in the five largest regions by population are better informed, and that all workers have more information about wages in the top quartile of regions by population. The information premium from living in highly populated areas adds to the benefits of cities discussed in the prior literature (Glaeser and Maré, 2001; De la Roca and Puga, 2017). Finally, panels (e) and (f) provide evidence on a particular mechanism that may explain the findings in panels (c) and (d): workers living in regions with higher internet access are better informed, and all workers have better information about regions with high internet access. This finding is consistent with prior evidence on the informational impact of broadband internet access (Akerman et al., 2022; Grubanov-Boskovic et al., 2022).

---

<sup>29</sup>Appendix D.2 shows migration rates decrease with distance in our context.

## 5.4 Counterfactuals

The findings above indicate migrants face substantial information frictions. To illustrate the quantitative importance of our estimates, we evaluate how changes in workers’ information sets and migration costs affect their migration choices and expected utility. Given our framework, we focus on workers’ individual responses, ignoring the impact that widespread changes in workers’ information and costs may have on migration flows through changes in equilibrium wages and prices (e.g., [Porcher, 2022](#)).

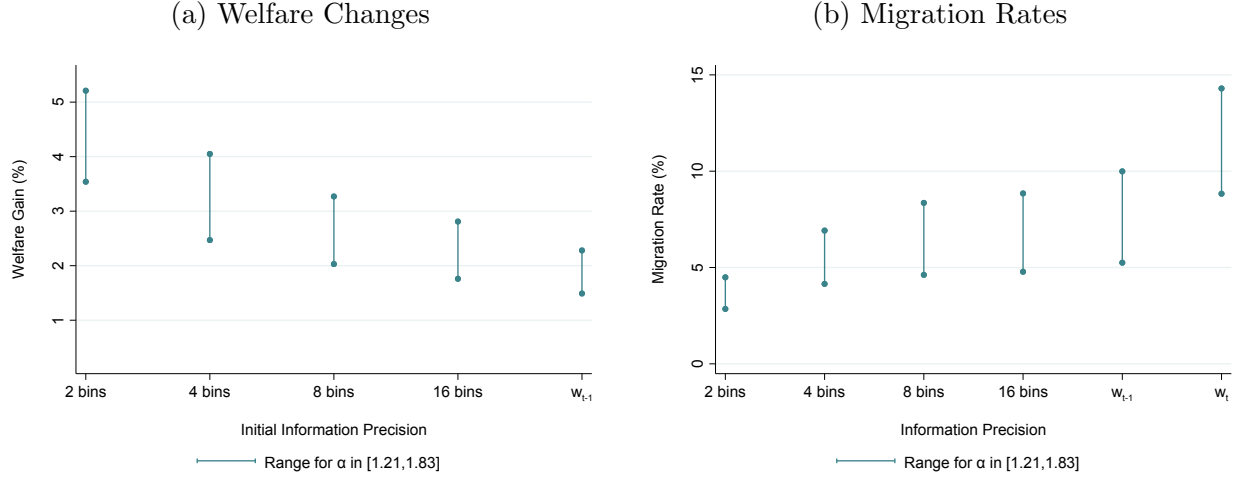
While the results above require no assumption on the stochastic process of wages, specifying one is needed for generating counterfactuals. We assume wages follow an AR(1) process with sector- and location-specific drifts, and estimate this process using observed wages in our sample of 50 regions and 20 sectors over 2002-2011. We find wages are strongly serially correlated (with a persistence estimate of 0.93). Additionally, while our two-step estimation procedure makes it computationally feasible to obtain CIs for all model parameters, computing model predictions that account for uncertainty in all parameter estimates is computationally costly, as it requires building a multidimensional grid that spans all CIs and evaluating our model at each point in that grid. Instead, we consider all points in the CI for the wage coefficient  $\alpha$ , but calibrate amenities by regressing the midpoint of the moment inequality CIs (see Figure [E.1](#) in Appendix [E.4](#)) on a constant and log distance.

We compute all counterfactuals of interest for a set of one million individuals drawn from the 2002 empirical distribution of workers across the 50 locations and 20 sectors in our estimation sample. We simulate the choices of these workers for the period 2002-2011 and for 100 simulation draws of the wage process. For each outcome of interest, we report the average value across all workers and simulation draws.

### 5.4.1 Changes in Information Sets

We first evaluate the impact of giving workers precise information on market wages in all locations. Specifically, we assume all workers have an initial level of information on wages common across all destinations, and focus on the impact that receiving perfect wage information has on workers’ migration probability, measured as the probability a worker changes locations in two consecutive periods, and welfare. We measure welfare as the average utility across simulated workers and periods, including the contribution of idiosyncratic tastes for locations and, importantly, using ex-post wages as the income measure. Hence, workers with perfect wage information choose locations exactly maximizing their utility, while workers with incomplete information maximize expected utility, and may choose locations that do not offer the highest utility ex post.

Figure 3: Effects of Providing Full Information About Wages



For the information sets indicated in the horizontal axis, panel (a) displays changes in welfare as a result of giving workers perfect wage information. Panel (b) displays migration rates for different information sets. The intervals illustrate the range of model predictions consistent with a value of  $\alpha$  in the 95% CI [1.21, 1.83].

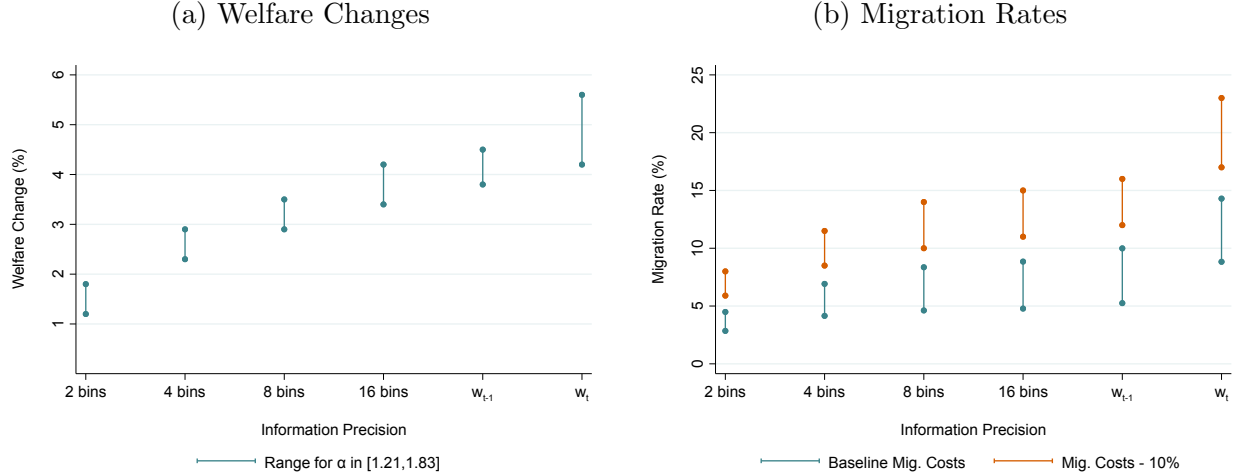
The results are displayed in Figure 3. Panel (a) shows that the largest welfare gains would be obtained by workers whose initial information only allows them to determine whether lagged wages are above or below median. In this case, welfare gains are between 3.5 and 5.2%, with the largest gains in this interval corresponding to the model that sets the wage coefficient  $\alpha$  to the largest value within its 95% CI. The gains decrease but remain sizeable for workers who were initially better informed. Even if all workers observed lagged wages perfectly—a hypothesis we reject in Panel (a) in Figure 1—the gains from observing contemporaneous wages would still range between 1.5 and 2.3%.

Panel (b) reports migration rates for workers with different information. Migration rates increase in the precision of the worker's wage information. They are below 5% for workers with the coarsest information set we consider, and between 9 and 14% when information is complete. When assuming all workers can discern the quartile to which lagged wages belong, which is the strongest informational assumption we tested without rejecting it (see Section 5.2.2), our estimated model predicts a migration rate between 4.2 and 6.9%, which is close to the average migration rate of 6.8% observed in the sample; see Figure D.1a in Appendix D.2.

#### 5.4.2 Reducing Barriers to Mobility

Reducing barriers to geographic mobility within a country is one of the main policy levers available to alleviate spatial misallocation (Morten and Oliveira, 2024). However, the benefits

Figure 4: Effects of Reducing Migration Costs, by Information Level



This figure displays counterfactual changes in welfare (panel (a)) and migration rates (panel (b)) from a 10% reduction in migration costs, depending on workers' information. The intervals correspond to the range of model predictions consistent with a value of  $\alpha$  in the 95% confidence interval  $[1.21, 1.83]$ .

of reducing mobility frictions may depend on whether agents are well-informed about the economic opportunities in the newly accessible regions. In this section, we evaluate how the effects of reductions in mobility frictions depend on the level of information frictions. Specifically, for several information sets, we compute the predictions of our estimated model for a 10% reduction in our calibrated migration costs.

Figure 4 displays the results. Panel (a) reveals that the welfare gains from a 10% reduction in migration costs increase greatly with the precision of workers' wage information. When workers are fully informed about wages in all regions, the welfare gains range from 4.2 to 5.7%, depending on the estimate of  $\alpha$ . When workers can only discern whether lagged wages in a location are above or below its median, the same reduction in migration costs only yields 1.2 to 1.8% welfare gains.

Panel (b) illustrates the increases in migration rates from reducing migration costs at each information level. The migration rates increase significantly for all information levels, and more so in relative terms for workers with a lower level of information precision. However, those larger increases in mobility have a higher rate of mistakes when the information precision is low, leading to the lower welfare gains reported in panel (a).

## 6 Conclusion

We introduce a new moment inequality method to estimate discrete choice models when information sets are unobserved to the researcher and potentially heterogeneous across agents, the choice set is large, and payoffs are parameterized with choice-specific fixed effects.

Using data on internal migration in Brazil, we apply our method to measure the different role migration costs and lack of information about potential destinations play in workers’ migration choices. Importantly, we do so while allowing information sets to vary flexibly across workers and over time, and migration costs to vary flexibly across pairs of origin and destination locations and years. We obtain four main results. First, workers have heterogeneous information on wages in other local markets. In particular, spatial factors play an important role in determining the precision of workers’ information about wages in different labor markets. Second, accounting for this incomplete information alters the mapping from observed location choices and wages to workers’ preferences. Our wage preference estimates are three times larger than those obtained using common estimation procedures, and our migration cost estimates are, on average, 21% lower. Third, our estimated model predicts that providing wage information to workers results in increases in both migration rates and welfare. Fourth, relative to a setting in which workers have perfect information on destination characteristics, policies that reduce migration costs by, for example, improving transportation infrastructure are significantly less effective at inducing mobility and improving migrants’ welfare when workers face information frictions of the magnitude we recover.

Our results emphasize that information frictions affect people’s behavior and, thus, observed choices do not reflect only their preferences. Moreover, even with correct estimates of people’s preferences, information frictions affect the predicted effects of policy proposals, including those that do not target information frictions directly. These insights, as well as our method, apply beyond the context of location choice. Deciding which schools to apply to and which health insurance plan to select are examples of other economically consequential decisions that most people make with limited information.

## References

- Akerman, Anders, Edwin Leuven, and Magne Mogstad**, “Information Frictions, Internet and the Relationship between Distance and Trade,” *American Economic Journal: Applied Economics*, 2022, 14 (1), 133–163. [28]
- Andrews, Donald W. K. and Soonwoo Kwon**, “Misspecified Moment Inequality Models: Inference and Diagnostics,” *Review of Economic Studies*, 2024, 91 (3), 45–76. [3, 18]

- Andrews, Donald W.K. and Gustavo Soares**, “Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection,” *Econometrica*, 2010, 78, 119–157. [16, 19]
- Andrews, Isaiah, Jonathan Roth, and Ariel Pakes**, “Inference for Linear Conditional Moment Inequalities,” *The Review of Economic Studies*, 2023, 90 (6), 2763–2791. [16]
- Artuç, Erhan and John McLaren**, “Trade Policy and Wage Inequality: A Structural Analysis with Occupational and Sectoral Mobility,” *Journal of International Economics*, 2015, 97 (2), 278–294. [4, 24]
- Barseghyan, Levon, Maura Coughlin, Francesca Molinari, and Joshua C. Teitelbaum**, “Heterogeneous Choice Sets and Preferences,” *Econometrica*, 2021, 5 (89), 2015–2048. [6]
- Baseler, Travis**, “Hidden Income and the Perceived Returns to Migration: Experimental Evidence from Kenya,” *American Economic Journal: Applied Economics*, 2023, 15 (4), 321–352. [5]
- Beine, Michel, Simone Bertoli, and Jesús Fernández-Huertas Moraga**, “A Practitioners’ Guide to Gravity Models of International Migration,” *The World Economy*, 2016, 39 (4), 496–512. [28]
- Bergman, Peter, Eric Chan, and Adam Kapor**, “Housing Search Frictions: Evidence from Detailed Search Data and a Field Experiment,” *mimeo*, May 2020. [5]
- , **Raj Chetty, Stefanie DeLuca, Nathaniel Hendren, Lawrence F Katz, and Christopher Palmer**, “Creating Moves to Opportunity: Experimental Evidence on Barriers to Neighborhood Choice,” *mimeo*, August 2023. [5]
- Bryan, Gharad, Shyamal Chowdhury, and Ahmed Mushfiq Mobarak**, “Underinvestment in a Profitable Technology: The Case of Seasonal Migration in Bangladesh,” *Econometrica*, 2014, 82 (5), 1671–1748. [5]
- Bugni, Federico A., Ivan A. Canay, and Xiaoxia Shi**, “Inference for Functions of Partially Identified Parameters in Moment Inequality Models,” *Quantitative Economics*, 2016, 8 (1), 1–38. [16]
- Caliendo, Lorenzo, Maximiliano Dvorkin, and Fernando Parro**, “Trade and Labor Market Dynamics: General Equilibrium Analysis of the China Trade Shock,” *Econometrica*, 2019, 87 (3), 741–835. [5]

- De la Roca, Jorge and Diego Puga**, “Learning by Working in Big Cities,” *Review of Economic Studies*, January 2017, *84* (1), 106–142. [28]
- Dickstein, Michael J. and Eduardo Morales**, “What do Exporters Know?,” *The Quarterly Journal of Economics*, 2018, *133* (4), 1753–1801. [6, 21]
- , **Jihye Jeon, and Eduardo Morales**, “Patient Costs and Physicians’ Information,” *mimeo*, 2023. [36, 3, 6, 10, 12]
- Dix-Carneiro, Rafael**, “Trade Liberalization and Labor Market Dynamics,” *Econometrica*, 2014, *82* (3), 825–885. [8]
- Eizenberg, Alon**, “Upstream Innovation and Product Variety in the U.S. Home PC Market,” *Review of Economic Studies*, 2014, *81*, 1003–1045. [3]
- Fan, Jingting, Sungwan Hong, and Fernando Parro**, “Learning and Expectations in Dynamic Spatial Economies,” *mimeo*, 2023. [5]
- Glaeser, Edward L. and David C. Maré**, “Cities and Skills,” *Journal of Labor Economics*, 2001, *19* (2), 316–342. [28]
- Grubanov-Boskovic, Sara, Sona Kalantaryan, Silvia Migali, and Marco Scipioni**, “The impact of the Internet on migration aspirations and intentions,” *Migration Studies*, 2022, *9* (4), 1807–1822. [28]
- Ho, Katherine and Ariel Pakes**, “Hospital Choices, Hospital Prices, and Financial Incentives to Physicians,” *American Economic Review*, 2014, *104* (12), 3841–3884. [5]
- Kaido, Hiroaki and Francesca Molinari**, “Information Based Inference in Models with Set-Valued Predictions and Misspecification,” *mimeo*, January 2024. [3]
- , —, and **Jorg Stoye**, “Confidence Intervals for Projections of Partially Identified Parameters,” *Econometrica*, 2019, *4* (87), 1397–1432. [16]
- Kaplan, Greg and Sam Schulhofer-Wohl**, “Understanding the Long-run Decline in Interstate Migration,” *International Economic Review*, 2017, *58* (1). [5, 8]
- Kennan, John and James R. Walker**, “The Effect of Expected Income on Individual Migration Decisions,” *Econometrica*, 2011, *79* (1), 211–251. [5, 8]
- Manski, Charles F.**, “Nonparametric Estimation of Expectations in the Analysis of Discrete Choice Under Uncertainty,” in William Barnett, James Powell, and George Tauchen,

- eds., *Nonparametric and Semiparametric Methods in Econometrics and Statistics*, Cambridge: Cambridge University Press, 1991. [5]
- , “Measuring Expectations,” *Econometrica*, 2004, 5 (9), 1329–1376. [5]
- Molinari, Francesca**, “Microeconometrics with Partial Identification,” in Steven N. Durlauf, Lars Peter Hansen, James J. Heckman, and Rosa L. Matzkin, eds., *Handbook of Econometrics, Volume 7A*, Elsevier, 2020, pp. 355–486. [3, 18]
- Morales, Eduardo, Gloria Sheu, and Andrés Zahler**, “Extended Gravity,” *The Review of Economic Studies*, 2019, 86 (6), 2668–2712. [6]
- Morten, Melanie and Jaqueline Oliveira**, “The Effects of Roads on Trade and Migration: Evidence from a Planned Capital City,” *American Economic Journal: Applied Economics*, 2024, 16 (2). [5, 30]
- Munshi, Kaivan**, “Social Networks and Migration,” *Annual Review of Economics*, 2020, 12 (Volume 12, 2020), 503–524. [28]
- Pakes, Ariel**, “Alternative Models for Moment Inequalities,” *Econometrica*, 2010, 78 (6), 1783–1822. [5]
- , **Jack R. Porter, Katherine Ho, and Joy Ishii**, “Moment Inequalities and their Application,” *Econometrica*, 2015, 83 (1), 315–334. [5]
- Porcher, Charly**, “Migration with Costly Information,” *mimeo*, 2022. [5, 8, 29]
- Redding, Stephen J. and Esteban Rossi-Hansberg**, “Quantitative Spatial Economics,” *Annual Review of Economics*, 2017, 9 (1), 21–58. [8]
- Tombe, Trevor and Xiaodong Zhu**, “Trade, Migration and Productivity: A Quantitative Analysis of China,” *American Economic Review*, 2019, *forthcoming*. [5]
- Traiberman, Sharon**, “Occupations and Import Competition,” *American Economic Review*, 2019, 109 (12), 4260–4301. [5]
- Wilson, Riley**, “Moving to Jobs: The Role of Information in Migration Decisions,” *Journal of Labor Economics*, 2021, 39 (4), 1083–1128. [5]
- Wollman, Thomas G.**, “Trucks Without Bailouts: Equilibrium Product Characteristics for Commercial Vehicles,” *American Economic Review*, 2017, 108 (6), 1364–1406. [3]

## A Proofs

Appendix A.1 provides the proof of Theorem 1. Appendix A.2, following Dickstein et al. (2023), provides the proof of Theorem 2. Appendix A.3 contains the proof of Theorem 3.

### A.1 Proof of Theorem 1

Equation (1) implies that, for any worker  $i$  of type  $s$  and locations  $l$  and  $l'$ , it holds that

$$(y_{is}^l + y_{is}^{l'}) (\mathbb{1}\{\mathbb{E}[\mathcal{U}_{is}^l - \mathcal{U}_{is}^{l'} | \mathcal{J}_{is}] \geq 0\} - y_{is}^l) = 0.$$

Equations (2) to (5) imply we can rewrite this equality as

$$(y_{is}^l + y_{is}^{l'}) (\mathbb{1}\{\Delta\kappa^{ll'} + \alpha\mathbb{E}[\Delta w_s^{ll'} | \mathcal{W}_{is}] + \Delta\varepsilon_{is}^{ll'} \geq 0\} - y_{is}^l) = 0. \quad (\text{A.1})$$

Equation (6) implies the expectation of this equality conditional on  $\mathcal{W}_{is}$  and a dummy variable that equals one if worker  $i$  of type  $s$  chooses either location  $l$  or location  $l'$  equals

$$\mathbb{E} \left[ \frac{\exp(\Delta\kappa^{ll'} + \alpha\mathbb{E}[\Delta w_s^{ll'} | \mathcal{W}_{is}])}{1 + \exp(\Delta\kappa^{ll'} + \alpha\mathbb{E}[\Delta w_s^{ll'} | \mathcal{W}_{is}])} - y_{is}^l \middle| \mathcal{W}_{is}, y_{is}^l + y_{is}^{l'} = 1 \right] = 0,$$

which implies the following moment equality

$$\mathbb{E}[1 - y_{is}^l - y_{is}^{l'} \exp(-\Delta\kappa^{ll'} - \alpha\mathbb{E}[\Delta w_s^{ll'} | \mathcal{W}_{is}]) | \mathcal{W}_{is}, y_{is}^l + y_{is}^{l'} = 1] = 0.$$

Given the conditioning on the event  $y_{is}^l + y_{is}^{l'} = 1$ , we can further simplify this equality as

$$\mathbb{E}[y_{is}^{l'} + y_{is}^l (-\exp(-(\Delta\kappa^{ll'} + \alpha\mathbb{E}[\Delta w_s^{ll'} | \mathcal{W}_{is}]))) | \mathcal{W}_{is}] = 0. \quad (\text{A.2})$$

As  $-\exp(-x)$  is concave in  $x \in \mathbb{R}$ , any linear approximation to it bounds it from above. The formula for the linear approximation to  $-\exp(-x)$  at  $x = a$  is  $-\exp(-a)(1 + a - x)$ . Thus, given an approximation point  $e_{is}^{ll'}$  for each worker  $i$  of type  $s$ , we derive the inequality

$$\mathbb{E}[y_{is}^{l'} - y_{is}^l \exp(-e_{is}^{ll'}) (1 + e_{is}^{ll'} - \Delta\kappa^{ll'} - \alpha\mathbb{E}[\Delta w_s^{ll'} | \mathcal{W}_{is}]) | \mathcal{W}_{is}] \geq 0. \quad (\text{A.3})$$

Consider the alternative moment,

$$\mathbb{E}[y_{is}^{l'} - y_{is}^l \exp(-e_{is}^{ll'}) (1 + e_{is}^{ll'} - \Delta\kappa^{ll'} - \alpha\Delta w_s^{ll'}) | \mathcal{W}_{is}]. \quad (\text{A.4})$$

Given  $\nu_{is}^{ll'} \equiv \Delta w_s^{ll'} - \mathbb{E}[\Delta w_s^{ll'} | \mathcal{J}_{is}]$  and equation (5), we can rewrite this moment as

$$\mathbb{E}[y_{is}^{l'} - y_{is}^l \exp(-e_{is}^{ll'}) (1 + e_{is}^{ll'} - \Delta\kappa^{ll'} - \alpha(\mathbb{E}[\Delta w_s^{ll'} | \mathcal{W}_{is}] + \nu_{is}^{ll'})) | \mathcal{W}_{is}].$$

As, by definition,  $\mathcal{W}_{is} \subseteq \mathcal{J}_{is}$ , we can use the Law of Iterated Expectations (LIE) to further write this moment as

$$\mathbb{E}[\mathbb{E}[y_{is}^{l'} - y_{is}^l \exp(-e_{is}^{ll'}) (1 + e_{is}^{ll'} - \Delta\kappa^{ll'} - \alpha(\mathbb{E}[\Delta w_s^{ll'} | \mathcal{W}_{is}] + \nu_{is}^{ll'})) | \mathcal{J}_{is}] | \mathcal{W}_{is}].$$

As  $e_{is}^{ll'} \subseteq \mathcal{J}_{is}$ , we can further write this moment as

$$\mathbb{E}[y_{is}^{l'} - y_{is}^l \exp(-e_{is}^{ll'}) (1 + e_{is}^{ll'} - \Delta\kappa^{ll'} - \alpha(\mathbb{E}[\Delta w_s^{ll'} | \mathcal{W}_{is}] + \mathbb{E}[\nu_{is}^{ll'} | \mathcal{J}_{is}])) | \mathcal{W}_{is}].$$

As equation (5) implies  $\mathbb{E}[\nu_{is}^{ll'} | \mathcal{J}_{is}] = 0$ , the moments in equations (A.3) and (A.4) coincide. Thus, the moment inequality in equation (A.3) implies the following inequality:

$$\mathbb{E}[y_{is}^{l'} - y_{is}^l \exp(-e_{is}^{ll'}) (1 + e_{is}^{ll'} - \Delta\kappa^{ll'} - \alpha\Delta w_s^{ll'}) | \mathcal{W}_{is}] \geq 0.$$

Finally, as  $z_s \subseteq \mathcal{W}_{is}$ , the LIE implies that  $\mathbb{m}^{ll'}(z_s, \Delta\kappa^{ll'}) \geq 0$ , proving Theorem 1. ■

## A.2 Proof of Theorem 2

We start from the moment equality in equation (A.2). Consider the alternative moment

$$\mathbb{E}[y_{is}^{l'} + y_{is}^l (-\exp(-(\Delta\kappa^{ll'} + \alpha\Delta w_s^{ll'}))) | \mathcal{W}_{is}]. \quad (\text{A.5})$$

Given  $\nu_{is}^{ll'} \equiv \Delta w_s^{ll'} - \mathbb{E}[\Delta w_s^{ll'} | \mathcal{J}_{is}]$  and equation (5), we can rewrite this moment as

$$\mathbb{E}[y_{is}^{l'} + y_{is}^l (-\exp(-(\Delta\kappa^{ll'} + \alpha(\mathbb{E}[\Delta w_s^{ll'} | \mathcal{W}_{is}] + \nu_{is}^{ll'})))) | \mathcal{W}_{is}].$$

As, by definition,  $\mathcal{W}_{is} \subseteq \mathcal{J}_{is}$ , we can use the LIE to further write this moment as

$$\mathbb{E}[\mathbb{E}[y_{is}^{l'} + y_{is}^l (-\exp(-(\Delta\kappa^{ll'} + \alpha(\mathbb{E}[\Delta w_s^{ll'} | \mathcal{W}_{is}] + \nu_{is}^{ll'})))) | \mathcal{J}_{is}] | \mathcal{W}_{is}].$$

As  $-\exp(x)$  is concave in  $x \in \mathbb{R}$ , equation (2) and Jensen's inequality imply the inequality

$$\begin{aligned} & \mathbb{E}[\mathbb{E}[y_{is}^{l'} + y_{is}^l (-\exp(-(\Delta\kappa^{ll'} + \alpha(\mathbb{E}[\Delta w_s^{ll'} | \mathcal{W}_{is}] + \nu_{is}^{ll'})))) | \mathcal{J}_{is}] | \mathcal{W}_{is}] \\ & \leq \mathbb{E}[y_{is}^{l'} + y_{is}^l (-\exp(-(\Delta\kappa^{ll'} + \alpha\mathbb{E}[\Delta w_s^{ll'} | \mathcal{W}_{is}]))) | \mathcal{W}_{is}], \end{aligned}$$

where the left-hand side coincides with the moment in equation (A.5), and the right-hand side coincides with the moment in equation (A.2). Thus, given equation (A.2), it holds that

$$\mathbb{E}[y_{is}' + y_{is}^l(-\exp(-(\Delta\kappa^{ll'} + \alpha\Delta w_s^{ll'})))|\mathcal{W}_{is}] \leq 0.$$

After multiplying both sides of this equation by  $-1$ , we obtain

$$\mathbb{E}[y_{is}^l \exp(-(\Delta\kappa^{ll'} + \alpha\Delta w_s^{ll'})) - y_{is}'|\mathcal{W}_{is}] \geq 0.$$

Finally, as  $z_s \subseteq \mathcal{W}_{is}$ , the LIE implies  $m_o^{ll'}(z_s, \Delta\kappa^{ll'}) \geq 0$ , proving Theorem 2. ■

### A.3 Proof of Theorem 3

Equation (A.1) implies that, for any worker  $i$  of type  $s$ , any worker  $j$  of type  $r$ , and any locations  $l$  and  $l'$ , it holds that

$$y_{jr}'(y_{is}^l + y_{is}')(\mathbb{1}\{\Delta\kappa^{ll'} + \alpha\mathbb{E}[\Delta w_s^{ll'}|\mathcal{W}_{is}] + \Delta\varepsilon_{is}^{ll'} \geq 0\} - y_{is}^l) = 0. \quad (\text{A.6})$$

Equation (6) implies the expectation of this equality conditional on  $\mathcal{W}_{is}$ ,  $\mathcal{W}_{jr}$ , and a dummy variable that equals one if worker  $i$  of type  $s$  chooses either location  $l$  or location  $l'$  equals

$$\mathbb{E}\left[y_{jr}' \frac{\exp(\Delta\kappa^{ll'} + \alpha\mathbb{E}[\Delta w_s^{ll'}|\mathcal{W}_{is}])}{1 + \exp(\Delta\kappa^{ll'} + \alpha\mathbb{E}[\Delta w_s^{ll'}|\mathcal{W}_{is}])} - y_{is}^l \middle| \mathcal{W}_{is}, \mathcal{W}_{jr}, y_{is}^l + y_{is}' = 1\right] = 0.$$

which, after some algebra, implies

$$\mathbb{E}[y_{jr}'(1 - y_{is}^l - y_{is}' \exp(-\Delta\kappa^{ll'} - \alpha\mathbb{E}[\Delta w_s^{ll'}|\mathcal{W}_{is}]))|\mathcal{W}_{is}, \mathcal{W}_{jr}, y_{is}^l + y_{is}' = 1] = 0.$$

Given the conditioning on the event  $y_{is}^l + y_{is}' = 1$ , we can further simplify this equality as

$$\mathbb{E}[y_{is}'y_{jr}' + y_{is}^ly_{jr}'(-\exp(-(\Delta\kappa^{ll'} + \alpha\mathbb{E}[\Delta w_s^{ll'}|\mathcal{W}_{is}])))|\mathcal{W}_{is}, \mathcal{W}_{jr}] = 0.$$

As  $-\exp(-x)$  is concave in  $x \in \mathbb{R}$ , any linear approximation to it bounds it from above. Thus, given an approximation point  $e_{isjr}^{ll'}$  for each worker  $i$  of type  $s$ , we derive the inequality

$$\mathbb{E}[y_{is}'y_{jr}' + y_{is}^ly_{jr}' \exp(-e_{isjr}^{ll'})(-(1 + e_{isjr}^{ll'}) + \Delta\kappa^{ll'} + \alpha\mathbb{E}[\Delta w_s^{ll'}|\mathcal{W}_{is}])|\mathcal{W}_{is}, \mathcal{W}_{jr}] \geq 0. \quad (\text{A.7})$$

Consider the alternative moment,

$$\mathbb{E}[y_{is}'y_{jr}' + y_{is}^ly_{jr}' \exp(-e_{isjr}^{ll'})(-(1 + e_{isjr}^{ll'}) + \Delta\kappa^{ll'} + \alpha\Delta w_s^{ll'})|\mathcal{W}_{is}, \mathcal{W}_{jr}]. \quad (\text{A.8})$$

Given  $\nu_{is}^{ll'} \equiv \Delta w_s^{ll'} - \mathbb{E}[\Delta w_s^{ll'} | \mathcal{J}_{is}]$  and equation (5), we can rewrite this moment as

$$\mathbb{E}[y_{is}' y_{jr}' + y_{is}^l y_{jr}' \exp(-e_{isjr}^{ll'}) (- (1 + e_{isjr}^{ll'}) + \Delta \kappa^{ll'} + \alpha(\mathbb{E}[\Delta w_s^{ll'} | \mathcal{W}_{is}] + \nu_{is}^{ll'})) | \mathcal{W}_{is}, \mathcal{W}_{jr}].$$

As, by definition,  $\mathcal{W}_{is} \subseteq \mathcal{J}_{is}$  and  $\mathcal{W}_{jr} \subseteq \mathcal{J}_{jr}$ , we can use the LIE to rewrite this moment as

$$\mathbb{E}[\mathbb{E}[y_{is}' y_{jr}' + y_{is}^l y_{jr}' \exp(-e_{isjr}^{ll'}) \times (- (1 + e_{isjr}^{ll'}) + \Delta \kappa^{ll'} + \alpha(\mathbb{E}[\Delta w_s^{ll'} | \mathcal{W}_{is}] + \nu_{is}^{ll'})) | \mathcal{J}_{is}, \mathcal{J}_{jr}] | \mathcal{W}_{is}, \mathcal{W}_{jr}].$$

As  $e_{isjr}^{ll'} \subseteq \mathcal{J}_{is} \cup \mathcal{J}_{jr}$ , we can further write this moment as

$$\mathbb{E}[y_{is}' y_{jr}' + y_{is}^l y_{jr}' \exp(-e_{isjr}^{ll'}) \times (- (1 + e_{isjr}^{ll'}) + \Delta \kappa^{ll'} + \alpha(\mathbb{E}[\Delta w_s^{ll'} | \mathcal{W}_{is}] + \mathbb{E}[\nu_{is}^{ll'} | \mathcal{J}_{is}, \mathcal{J}_{jr}])) | \mathcal{W}_{is}, \mathcal{W}_{jr}].$$

As equation (5) implies  $\mathbb{E}[\nu_{is}^{ll'} | \mathcal{J}_{is}, \mathcal{J}_{jr}] = 0$ , the moments in equations (A.7) and (A.8) coincide. Thus, the moment inequality in equation (A.7) implies the following inequality:

$$\mathbb{E}[y_{is}' y_{jr}' + y_{is}^l y_{jr}' \exp(-e_{isjr}^{ll'}) (- (1 + e_{isjr}^{ll'}) + \Delta \kappa^{ll'} + \alpha \Delta w_s^{ll'}) | \mathcal{W}_{is}, \mathcal{W}_{jr}] \geq 0. \quad (\text{A.9})$$

This moment inequality is one of the two that we will combine to obtain that in equation (19). To obtain the second moment inequality, we start from

$$y_{is}^l (y_{jr}^l + y_{jr}') (\mathbb{1}\{\Delta \kappa^{ll'} + \alpha \mathbb{E}[\Delta w_r^{ll'} | \mathcal{W}_{jr}] + \Delta \varepsilon_{jr}^{ll'} \geq 0\} - y_{jr}') = 0, \quad (\text{A.10})$$

which is analogous to that in equation (A.6). Following the same steps described above to go from equation (A.6) to equation (A.9), we can derive the following inequality

$$\mathbb{E}[y_{is}^l y_{jr}^l + y_{is}^l y_{jr}' \exp(-e_{jr is}^{ll'}) (- (1 + e_{jr is}^{ll'}) + \Delta \kappa^{ll'} + \alpha \Delta w_r^{ll'}) | \mathcal{W}_{is}, \mathcal{W}_{jr}] \geq 0. \quad (\text{A.11})$$

As the moments in equations (A.9) and (A.11) have the same conditioning set, we can add them. If we further impose that  $e_{isjr}^{ll'} = e_{jr is}^{ll'}$ , we obtain the following moment inequality:

$$\mathbb{E}[y_{is}^l y_{jr}^l + y_{is}' y_{jr}' + y_{is}^l y_{jr}' \exp(-e_{isjr}^{ll'}) (- (1 + e_{isjr}^{ll'}) + \alpha(\Delta w_s^{ll'} + \Delta w_r^{ll'})) | \mathcal{W}_{is}, \mathcal{W}_{jr}] \geq 0.$$

As  $z_s \subseteq \mathcal{W}_{is}$  and  $z_r \subseteq \mathcal{W}_{jr}$ , the LIE implies  $\mathbb{M}^{ll'}(z_s, z_r, \alpha) \geq 0$ , proving Theorem 3. ■

# Online Appendix for “Measuring Information Frictions in Migration Decisions: A Revealed-Preference Approach”

Charly Porcher  
Georgetown University

Eduardo Morales  
Princeton University

Thomas Fujiwara  
Princeton University

## B Additional Derivations

Appendix B.1 derives expressions appearing in Section 3.1.1. Appendix B.2 proves Corollary 1. Appendix B.3 derives expressions appearing in Section 3.1.2. Appendix B.4 proves Corollary 2. Appendix B.5 derives expressions appearing in Section 3.2. Appendix B.6 proves Corollary 3. Appendix B.7 shows how we use inequalities to identify the parameter  $\theta_\alpha$ . Appendix B.8 describes how we implement the procedure in Andrews and Soares (2010).

### B.1 Second-Step Bounding Inequalities: Additional Derivations

*Derivation of equation (9).* Consider points  $e_{is}^{ll'}$  such that  $e_{is}^{ll'} = h^{ll'}(z_s, \Delta\theta_{ll'})$  for some function  $h^{ll'}(\cdot)$ . We compute the function in equation (9) by finding the value of  $h^{ll'}(z_s, \Delta\theta_{ll'})$  that minimizes the moment in equation (7) at each value of  $\Delta\theta_{ll'}$ . Specifically, given  $z_s$  and  $\Delta\theta_{ll'}$ , the first-order condition of the moment in equation (7) with respect to  $h^{ll'}(z_s, \Delta\theta_{ll'})$  is

$$\mathbb{E}[y_{is}^l(h^{ll'}(z_s, \Delta\theta_{ll'}) - (\Delta\theta_{ll'} + \alpha\Delta w_s^{ll'}))|z_s] = 0,$$

or, equivalently,  $\mathbb{E}[h^{ll'}(z_s, \Delta\theta_{ll'}) - (\Delta\theta_{ll'} + \alpha\Delta w_s^{ll'})|z_s, y_{is}^l = 1] = 0$ . Solving for  $h^{ll'}(z_s, \Delta\theta_{ll'})$ , we obtain the expression in equation (9).

*Derivation of equation (10).* Equations (7) to (9) imply the following inequality

$$\mathbb{E}[y_{is}^{l'} - y_{is}^l \exp(-(\Delta\theta_{ll'} + \alpha\mathbb{E}[\Delta w_s^{ll'}|z_s, y_{is}^l = 1]))(1 - \alpha(\Delta w_s^{ll'} - \mathbb{E}[\Delta w_s^{ll'}|z_s, y_{is}^l = 1]))|z_s] \geq 0.$$

We can simplify this inequality as

$$\begin{aligned} \mathbb{E}[y_{is}^{l'}|z_s] &\geq \exp(-(\Delta\theta_{ll'} + \alpha\mathbb{E}[\Delta w_s^{ll'}|z_s, y_{is}^l = 1])) \\ &\quad \times (\mathbb{E}[y_{is}^l|z_s] - \alpha\mathbb{E}[y_{is}^l(\Delta w_s^{ll'} - \mathbb{E}[\Delta w_s^{ll'}|z_s, y_{is}^l = 1])|z_s]), \end{aligned}$$

or equivalently,

$$\mathbb{E}[y_{is}^{l'}|z_s] \geq \exp(-(\Delta\theta_{ll'} + \alpha\mathbb{E}[\Delta w_s^{ll'}|z_s, y_{is}^l = 1]))$$

$$\times (\mathbb{E}[y_{is}^l | z_s] - \alpha(\mathbb{E}[\Delta w_s^{ll'} | z_s, y_{is}^l = 1] \mathbb{E}[y_{is}^l | z_s] - \mathbb{E}[\Delta w_s^{ll'} | z_s, y_{is}^l = 1] \mathbb{E}[y_{is}^l | z_s])).$$

Eliminating terms that cancel each other, we obtain the inequality in equation (10); i.e.,

$$\frac{\mathbb{E}[y_{is}^l | z_s]}{\mathbb{E}[y_{is}^{l'} | z_s]} \exp(-\alpha \mathbb{E}[\Delta w_s^{ll'} | z_s, y_{is}^l = 1]) \leq \exp(\Delta \theta_{ll'}).$$

*Derivation of equation (11).* Swapping the indices  $l$  and  $l'$  in equation (10) we obtain

$$\frac{\mathbb{E}[y_{is}^{l'} | z_s]}{\mathbb{E}[y_{is}^l | z_s]} \exp(-\alpha \mathbb{E}[\Delta w_s^{ll'} | z_s, y_{is}^{l'} = 1]) \leq \exp(\Delta \theta_{ll'}).$$

Rearranging terms, we obtain the inequality in equation (11).

## B.2 Second-Step Bounding Inequalities: Proof of Corollary 1

Equations (7) to (9) imply the following moment inequality:

$$\mathbb{E}[y_{is}^l \exp(-(\Delta \theta_{ll'} + \alpha \mathbb{E}[\Delta w_s^{ll'} | z_s, y_{is}^l = 1])) - y_{is}^{l'} | z_s] \geq 0. \quad (\text{B.1})$$

Assuming  $z_s \subseteq \mathcal{W}_{is}$  and using the LIE, we can write

$$\mathbb{E}[\mathbb{E}[y_{is}^l \exp(-(\Delta \theta_{ll'} + \alpha \mathbb{E}[\Delta w_s^{ll'} | z_s, y_{is}^l = 1])) - y_{is}^{l'} | \mathcal{W}_{is}] | z_s] \geq 0.$$

Given that  $z_s \subseteq \mathcal{W}_{is}$ , we can further rewrite

$$\mathbb{E}[\mathbb{E}[y_{is}^l | \mathcal{W}_{is}] \exp(-(\Delta \theta_{ll'} + \alpha \mathbb{E}[\Delta w_s^{ll'} | z_s, y_{is}^l = 1])) - \mathbb{E}[y_{is}^{l'} | \mathcal{W}_{is}] | z_s] \geq 0. \quad (\text{B.2})$$

As  $y_{is}^l$  is a function of  $(\mathcal{W}_{is}, \varepsilon_{is})$ , equation (5) implies  $\mathbb{E}[\Delta w_s^{ll'} | \mathcal{W}_{is}, y_{is}^l] = \mathbb{E}[\Delta w_s^{ll'} | \mathcal{W}_{is}]$ . Since  $\mathbb{E}[\Delta w_s^{ll'} | z_s] = \mathbb{E}[\Delta w_s^{ll'} | \mathcal{W}_{is}]$  according to Corollary 1, this corollary implies we can rewrite equation (B.2) as

$$\mathbb{E}[\mathbb{E}[y_{is}^l | \mathcal{W}_{is}] \exp(-(\Delta \theta_{ll'} + \alpha \mathbb{E}[\Delta w_s^{ll'} | z_s])) - \mathbb{E}[y_{is}^{l'} | \mathcal{W}_{is}] | z_s] \geq 0.$$

Given equation (6), we can further rewrite

$$\begin{aligned} \mathbb{E} \left[ \exp(\Delta \kappa^{ll'} + \alpha \mathbb{E}[\Delta w_s^{ll'} | \mathcal{W}_{is}]) \left( \sum_{l''=1}^L \exp(\Delta \kappa^{ll''} + \alpha \mathbb{E}[\Delta w_s^{ll''} | \mathcal{W}_{is}]) \right)^{-1} \times \right. \\ \left. \exp(-(\Delta \theta_{ll'} + \alpha \mathbb{E}[\Delta w_s^{ll'} | z_s])) - \mathbb{E}[y_{is}^{l'} | \mathcal{W}_{is}] | z_s \right] \geq 0. \end{aligned}$$

Using a similar expression for the probability of choosing  $l'$  conditional on  $\mathcal{W}_{is}$ , we derive

$$\mathbb{E}\left[\left(\exp(\Delta\kappa^{ll'} - \Delta\theta_{ll'}) - 1\right)\left(\sum_{l''=1}^L \exp(\Delta\kappa^{l''l'} + \alpha\mathbb{E}[\Delta w_s^{l''l'}|\mathcal{W}_{is}])\right)^{-1}\middle|z_s\right] \geq 0,$$

where we have used that  $\mathbb{E}[\Delta w_s^{ll'}|z_s] = \mathbb{E}[\Delta w_s^{ll'}|\mathcal{W}_{is}]$  according to Corollary 1. Then,

$$(\exp(\Delta\kappa^{ll'} - \Delta\theta_{ll'}) - 1)\mathbb{E}\left[\left(\sum_{l''=1}^L \exp(\Delta\kappa^{l''l'} + \alpha\mathbb{E}[\Delta w_s^{l''l'}|\mathcal{W}_{is}])\right)^{-1}\middle|z_s\right] \geq 0.$$

The expectation in this inequality is always strictly positive. Thus, the inequality implies

$$\exp(\Delta\kappa^{ll'} - \Delta\theta_{ll'}) - 1 \geq 0 \quad \Leftrightarrow \quad \Delta\kappa^{ll'} \geq \Delta\theta_{ll'}. \quad (\text{B.3})$$

This inequality holds for any locations  $l$  and  $l'$ . Swapping the indices  $l$  and  $l'$ , we obtain:

$$\exp(\Delta\kappa^{l'l} - \Delta\theta_{l'l}) - 1 \geq 0 \quad \Leftrightarrow \quad \Delta\kappa^{l'l} \leq \Delta\theta_{l'l}. \quad (\text{B.4})$$

Equations (B.3) and (B.4) imply  $\Delta\kappa^{ll'} = \Delta\theta_{ll'}$ , proving Corollary 1. ■

### B.3 Second-Step Odds-Based Inequalities: Additional Derivations

*Derivation of equation (14).* Equations (12) and (13) imply the following inequality

$$\mathbb{E}[y_{is}^l \exp(-(\Delta\theta_{ll'} + \alpha\Delta w_s^{ll'}))|z_s] \geq \mathbb{E}[y_{is}^{l'}|z_s].$$

We can rewrite this inequality as

$$\mathbb{E}[y_{is}^l|z_s] \exp(-\Delta\theta_{ll'}) \mathbb{E}[(\exp(\alpha\Delta w_s^{ll'}))^{-1}|z_s, y_{is}^l = 1] \geq \mathbb{E}[y_{is}^{l'}|z_s].$$

Rearranging terms, we obtain the expression in equation (14); i.e.,

$$\frac{\mathbb{E}[y_{is}^l|z_s]}{\mathbb{E}[y_{is}^{l'}|z_s]} \mathbb{E}[(\exp(\alpha\Delta w_s^{ll'}))^{-1}|z_s, y_{is}^l = 1] \geq \exp(\Delta\theta_{ll'}).$$

*Derivation of equation (15).* Swapping the indices  $l$  and  $l'$  in equation (14), we obtain

$$\frac{\mathbb{E}[y_{is}^{l'}|z_s]}{\mathbb{E}[y_{is}^l|z_s]} \mathbb{E}[\exp(-\alpha\Delta w_s^{l'l})|z_s, y_{is}^{l'} = 1] \geq \exp(\Delta\theta_{l'l}).$$

Rearranging terms, we immediately obtain the inequality in equation (15).

## B.4 Second-Step Odds-Based Inequalities: Proof of Corollary 2

Equations (12) and (13) imply the following moment inequality:

$$\mathbb{E}[y_{is}^l \exp(-(\Delta\theta_{ll'} + \alpha\Delta w_s^{ll'})) - y_{is}^{l'} | z_s] \geq 0. \quad (\text{B.5})$$

Assuming  $z_s \subseteq \mathcal{W}_{is}$ , the LIE implies we can write this inequality as

$$\mathbb{E}[\mathbb{E}[y_{is}^l \exp(-(\Delta\theta_{ll'} + \alpha\Delta w_s^{ll'})) - y_{is}^{l'} | \mathcal{W}_{is}] | z_s] \geq 0.$$

Since  $\Delta w_s^{ll'} = \mathbb{E}[\Delta w_s^{ll'} | \mathcal{W}_{is}]$  according to Corollary 2, we can rewrite this inequality as

$$\mathbb{E}[\mathbb{E}[y_{is}^l | \mathcal{W}_{is}] \exp(-(\Delta\theta_{ll'} + \alpha\mathbb{E}[\Delta w_s^{ll'} | \mathcal{W}_{is}])) - \mathbb{E}[y_{is}^{l'} | \mathcal{W}_{is}] | z_s] \geq 0.$$

Given the expression for the probability of choosing  $l$  conditional on  $\mathcal{W}_{is}$ , we rewrite

$$\begin{aligned} \mathbb{E} \left[ \exp(\Delta\kappa^{ll'} + \alpha\mathbb{E}[\Delta w_s^{ll'} | \mathcal{W}_{is}]) \left( \sum_{l''=1}^L \exp(\Delta\kappa^{l''l'} + \alpha\mathbb{E}[\Delta w_s^{l''l'} | \mathcal{W}_{is}]) \right)^{-1} \times \right. \\ \left. \exp(-(\Delta\theta_{ll'} + \alpha\mathbb{E}[\Delta w_s^{ll'} | \mathcal{W}_{is}])) - \mathbb{E}[y_{is}^{l'} | \mathcal{W}_{is}] \right] | z_s \geq 0. \end{aligned}$$

Using a similar expression for the probability of choosing  $l'$  conditional on  $\mathcal{W}_{is}$ , we derive

$$(\exp(\Delta\kappa^{ll'} - \Delta\theta_{ll'}) - 1) \mathbb{E} \left[ \left( \sum_{l''=1}^L \exp(\Delta\kappa^{l''l'} + \alpha\mathbb{E}[\Delta w_s^{l''l'} | \mathcal{W}_{is}]) \right)^{-1} \right] | z_s \geq 0. \quad (\text{B.6})$$

The expectation in this inequality is always strictly positive. Thus, the inequality implies

$$\exp(\Delta\kappa^{ll'} - \Delta\theta_{ll'}) - 1 \geq 0 \quad \Leftrightarrow \quad \Delta\kappa^{ll'} \geq \Delta\theta_{ll'}. \quad (\text{B.7})$$

This inequality holds for any locations  $l$  and  $l'$ . Swapping the indices  $l$  and  $l'$ , we obtain:

$$\exp(\Delta\kappa^{l'l} - \Delta\theta_{l'l}) - 1 \geq 0 \quad \Leftrightarrow \quad \Delta\kappa^{l'l} \leq \Delta\theta_{l'l}. \quad (\text{B.8})$$

Equations (B.7) and (B.8) imply  $\Delta\kappa^{ll'} = \Delta\theta_{ll'}$  proving Corollary 2. ■

## B.5 First-Step Moment Inequalities: Additional Derivations

*Derivation of equation (20).* Consider points  $e_{isjr}^{ll'}$  such that  $e_{isjr}^{ll'} = h^{ll'}(z_s, z_r, \theta_\alpha)$  for some function  $h^{ll'}(\cdot)$ . We compute the function in equation (20) by finding the value of  $h^{ll'}(z_s, z_r, \theta_\alpha)$  that minimizes the moment in equation (17) at each value of  $\theta_\alpha$ . Specifically, given  $z_s, z_r$ ,

and  $\theta_\alpha$ , the first-order condition of the moment in equation (17) with respect to  $h^{ll'}(z_s, z_r, \theta_\alpha)$  is

$$\mathbb{E}[y_{is}^l y_{jr}^{l'} (2h^{ll'}(z_s, z_r, \theta_\alpha) - \theta_\alpha(\Delta w_s^{ll'} + \Delta w_r^{l'l})) | z_s, z_r] = 0,$$

or, equivalently,  $\mathbb{E}[2h^{ll'}(z_s, z_r, \theta_\alpha) - \theta_\alpha(\Delta w_s^{ll'} + \Delta w_r^{l'l}) | z_s, z_r, y_{is}^l y_{jr}^{l'} = 1] = 0$ . Solving for  $h^{ll'}(z_s, z_r, \theta_\alpha)$ , we obtain the solution in equation (20).

*Derivation of equation (21).* Equations (17) to (20) imply the following inequality

$$\begin{aligned} & \mathbb{E}[y_{is}^l y_{jr}^l + y_{is}^{l'} y_{jr}^{l'} - y_{is}^l y_{jr}^{l'} \exp(-\theta_\alpha \mathbb{E}[0.5(\Delta w_s^{ll'} + \Delta w_r^{l'l}) | z_s, z_r, y_{is}^l y_{jr}^{l'} = 1]) \times \\ & (2 - \theta_\alpha((\Delta w_s^{ll'} + \Delta w_r^{l'l}) - \mathbb{E}[\Delta w_s^{ll'} + \Delta w_r^{l'l} | z_s, z_r, y_{is}^l y_{jr}^{l'} = 1])) | z_s, z_r] \geq 0, \end{aligned}$$

or, equivalently,

$$\begin{aligned} & \mathbb{E}[y_{is}^l y_{jr}^l + y_{is}^{l'} y_{jr}^{l'} | z_s, z_r] \geq \mathbb{E}[y_{is}^l y_{jr}^{l'} \exp(-\theta_\alpha \mathbb{E}[0.5(\Delta w_s^{ll'} + \Delta w_r^{l'l}) | z_s, z_r, y_{is}^l y_{jr}^{l'} = 1]) \times \\ & (2 - \theta_\alpha((\Delta w_s^{ll'} + \Delta w_r^{l'l}) - \mathbb{E}[\Delta w_s^{ll'} + \Delta w_r^{l'l} | z_s, z_r, y_{is}^l y_{jr}^{l'} = 1])) | z_s, z_r]. \end{aligned}$$

Using the LIE, we can rewrite this inequality as

$$\begin{aligned} & \mathbb{E}[y_{is}^l y_{jr}^l + y_{is}^{l'} y_{jr}^{l'} | z_s, z_r] \geq \mathbb{E}[\exp(-\theta_\alpha \mathbb{E}[0.5(\Delta w_s^{ll'} + \Delta w_r^{l'l}) | z_s, z_r, y_{is}^l y_{jr}^{l'} = 1]) \times \\ & (2 - \theta_\alpha((\Delta w_s^{ll'} + \Delta w_r^{l'l}) - \mathbb{E}[\Delta w_s^{ll'} + \Delta w_r^{l'l} | z_s, z_r, y_{is}^l y_{jr}^{l'} = 1])) | z_s, z_r, y_{is}^l y_{jr}^{l'} = 1] \times \\ & \mathbb{E}[y_{is}^l y_{jr}^{l'} | z_s, z_r] \geq 0. \end{aligned}$$

Simplifying this expression, and rearranging, we obtain the expression in equation (21).

## B.6 First-Step Moment Inequalities: Proof of Corollary 3

Since  $z_s \subseteq \mathcal{W}_{is}$  and  $z_r \subseteq \mathcal{W}_{jr}$  according to Corollary 3, we rewrite equation (21) as

$$\begin{aligned} & \frac{\mathbb{E}[\mathbb{E}[y_{is}^l y_{jr}^{l'} | \mathcal{W}_{is}, \mathcal{W}_{jr}] | z_s, z_r]}{\mathbb{E}[0.5(\mathbb{E}[y_{is}^l y_{jr}^l | \mathcal{W}_{is}, \mathcal{W}_{jr}] + \mathbb{E}[y_{is}^{l'} y_{jr}^{l'} | \mathcal{W}_{is}, \mathcal{W}_{jr}]) | z_s, z_r]} \leq \\ & \exp(\theta_\alpha \mathbb{E}[0.5(\Delta w_s^{ll'} + \Delta w_r^{l'l}) | z_s, z_r, y_{is}^l y_{jr}^{l'} = 1]). \end{aligned}$$

Equations (5) and (6) further imply that we can rewrite this inequality as

$$\begin{aligned} & \frac{\mathbb{E}[\mathbb{E}[y_{is}^l | \mathcal{W}_{is}] \mathbb{E}[y_{jr}^{l'} | \mathcal{W}_{jr}] | z_s, z_r]}{\mathbb{E}[0.5(\mathbb{E}[y_{is}^l | \mathcal{W}_{is}] \mathbb{E}[y_{jr}^l | \mathcal{W}_{jr}] + \mathbb{E}[y_{is}^{l'} | \mathcal{W}_{is}] \mathbb{E}[y_{jr}^{l'} | \mathcal{W}_{jr}]) | z_s, z_r]} \leq \\ & \exp(\theta_\alpha \mathbb{E}[0.5(\Delta w_s^{ll'} + \Delta w_r^{l'l}) | z_s, z_r, y_{is}^l y_{jr}^{l'} = 1]). \end{aligned} \tag{B.9}$$

Given equations (1) to (6), it holds that, for any  $l_1 = 1, \dots, L$  and  $l_2 = 1, \dots, L$ , we can write

$$\mathbb{E}[y_{is}^{l_1} | \mathcal{W}_{is}] = \frac{\exp(\Delta\kappa^{l_1 l_2} + \alpha \mathbb{E}[\Delta w_s^{l_1 l_2} | \mathcal{W}_{is}])}{\sum_{l''=1}^L \exp(\Delta\kappa^{l'' l_2} + \alpha \mathbb{E}[\Delta w_s^{l'' l_2} | \mathcal{W}_{is}])},$$

and similarly for worker  $j$  of type  $r$ . We then rewrite the inequality in equation (B.9) as

$$\begin{aligned} & \frac{\mathbb{E}[\exp(\Delta\kappa^{ll'} + \alpha \mathbb{E}[\Delta w_s^{ll'} | \mathcal{W}_{is}]) \exp(\Delta\kappa^{ll'} + \alpha \mathbb{E}[\Delta w_r^{ll'} | \mathcal{W}_{jr}]) | z_s, z_r]}{\mathbb{E}[0.5(\exp(\Delta\kappa^{ll'} + \alpha \mathbb{E}[\Delta w_s^{ll'} | \mathcal{W}_{is}]) + \exp(\Delta\kappa^{ll'} + \alpha \mathbb{E}[\Delta w_r^{ll'} | \mathcal{W}_{jr}])) | z_s, z_r]} \\ & \leq \exp(\theta_\alpha \mathbb{E}[0.5(\Delta w_s^{ll'} + \Delta w_r^{ll'}) | z_s, z_r, y_{is}^l y_{jr}^{l'} = 1]). \end{aligned}$$

Simplifying this expression, we obtain

$$\begin{aligned} & \frac{\mathbb{E}[\exp(\alpha \mathbb{E}[\Delta w_s^{ll'} | \mathcal{W}_{is}]) \exp(\alpha \mathbb{E}[\Delta w_r^{ll'} | \mathcal{W}_{jr}]) | z_s, z_r]}{\mathbb{E}[0.5(\exp(\Delta\kappa^{ll'} + \alpha \mathbb{E}[\Delta w_s^{ll'} | \mathcal{W}_{is}]) + \exp(\Delta\kappa^{ll'} + \alpha \mathbb{E}[\Delta w_r^{ll'} | \mathcal{W}_{jr}])) | z_s, z_r]} \\ & \leq \exp(\theta_\alpha \mathbb{E}[0.5(\Delta w_s^{ll'} + \Delta w_r^{ll'}) | z_s, z_r, y_{is}^l y_{jr}^{l'} = 1]). \end{aligned}$$

Since  $\mathbb{E}[\Delta w_s^{ll'} | z_s] = \mathbb{E}[\Delta w_s^{ll'} | \mathcal{W}_{is}] = \mathbb{E}[\Delta w_r^{ll'} | z_r] = \mathbb{E}[\Delta w_r^{ll'} | \mathcal{W}_{jr}] = \Delta\bar{w}$ , for a common constant  $\Delta\bar{w} \in \mathbb{R}$ , according to Corollary 3, this inequality becomes

$$\frac{\exp(\alpha \Delta\bar{w}) \exp(\alpha \Delta\bar{w})}{0.5(\exp(\Delta\kappa^{ll'} + \alpha \Delta\bar{w}) + \exp(\Delta\kappa^{ll'} + \alpha \Delta\bar{w}))} \leq \exp(\theta_\alpha \Delta\bar{w}).$$

If  $\Delta\kappa^{ll'} = 0$ , then it becomes

$$\frac{\exp(\alpha \Delta\bar{w}) \exp(\alpha \Delta\bar{w})}{0.5(\exp(\alpha \Delta\bar{w}) + \exp(\alpha \Delta\bar{w}))} \leq \exp(\theta_\alpha \Delta\bar{w}) \quad \Leftrightarrow \quad \exp(\alpha \Delta\bar{w}) \leq \exp(\theta_\alpha \Delta\bar{w}).$$

Thus, two inequalities of this type, one with  $\Delta\bar{w} > 0$  and the other one with  $\Delta\bar{w} < 0$ , will only be satisfied if  $\theta_\alpha = \alpha$ , proving in this way Corollary 3.  $\blacksquare$

## B.7 Using Inequalities for Estimation of the Wage Parameter

We describe here how we use the inequalities in Section 3.2 to compute a CI for  $\theta_\alpha$ . The inequality in equation (17) is specific to locations  $l$  and  $l'$  and conditions on the vectors  $z_s$  and  $z_r$ . Given that conditional moment inequality, we implement the following steps to derive  $k = 1, \dots, K$  unconditional moment inequalities. First, we choose scalars  $\Delta z_s^{ll'} \subseteq z_s$  and  $\Delta z_r^{ll'} \subseteq z_r$  that are correlated with  $\Delta w_s^{ll'}$  and  $\Delta w_r^{ll'}$ , and that we use in all  $K$  inequalities. Second, for each  $k$ , we choose a subset  $[z_{ks}, \bar{z}_{ks}]$  of the support of  $\Delta z_s^{ll'}$ , a subset  $[z_{kr}, \bar{z}_{kr}]$  of the support of  $\Delta z_r^{ll'}$ , and an exponent  $d_k \in \mathbb{Z}$ . Given these choices, we build the inequality

$$\mathbb{E}[(y_{is}^l y_{jr}^l + y_{is}^{l'} y_{jr}^{l'} - y_{is}^l y_{jr}^{l'} \exp(-e_{isjr}^{ll'})(2 + 2e_{isjr}^{ll'} - \theta_\alpha(\Delta w_s^{ll'} + \Delta w_r^{ll'}))) \times g_k(\Delta z_s^{ll'}, \Delta z_r^{ll'})] \geq 0, \quad (\text{B.10})$$

where the term in parenthesis coincides with the moment function in equation (17) and

$$g_k(\Delta z_s^{ll'}, \Delta z_r^{ll'}) = \mathbb{1}\{z_{ks} < \Delta z_s^{ll'} \leq \bar{z}_{ks}\} \mathbb{1}\{z_{kr} < \Delta z_r^{ll'} \leq \bar{z}_{kr}\} (|\Delta z_s^{ll'}| |\Delta z_r^{ll'}|)^{d_k}.$$

Effectively, for a pre-specified  $q \in \mathbb{N}$ , we choose limits  $z_{ks}$  and  $\bar{z}_{ks}$  that correspond to consecutive elements of the vector of  $q$ -quantiles of the distribution of  $\Delta z_s^{ll'}$  across all types and location pairs. We do the same for the limits  $z_{kr}$  and  $\bar{z}_{kr}$ . Concerning the exponent  $d_k$ , we set it to either  $-1$ ,  $0$ , or  $1$ . For example, when  $q = 2$  and  $d_k = 0$  for all  $k$ , the number of inequalities of the type in equation (22) is  $K = 4$ , and the corresponding instruments are

$$g_k(\Delta z_s^{ll'}, \Delta z_r^{ll'}) = \begin{cases} \mathbb{1}\{\Delta z_s^{ll'} \leq \text{med}(\Delta z_s^{ll'})\} \mathbb{1}\{\Delta z_r^{ll'} \leq \text{med}(\Delta z_r^{ll'})\} & \text{if } k = 1, \\ \mathbb{1}\{\Delta z_s^{ll'} \leq \text{med}(\Delta z_s^{ll'})\} \mathbb{1}\{\Delta z_r^{ll'} > \text{med}(\Delta z_r^{ll'})\} & \text{if } k = 2, \\ \mathbb{1}\{\Delta z_s^{ll'} > \text{med}(\Delta z_s^{ll'})\} \mathbb{1}\{\Delta z_r^{ll'} \leq \text{med}(\Delta z_r^{ll'})\} & \text{if } k = 3, \\ \mathbb{1}\{\Delta z_s^{ll'} > \text{med}(\Delta z_s^{ll'})\} \mathbb{1}\{\Delta z_r^{ll'} > \text{med}(\Delta z_r^{ll'})\} & \text{if } k = 4. \end{cases}$$

To compute the sample analogue of the moment inequality in equation (B.10), we average across worker types and workers within each type. If the cardinality  $L$  of the worker's choice set is large, it may be convenient to further average across all possible location pairs  $(l, l')$ .

## B.8 Inference Procedure: Andrews and Soares (2010)

We describe here our implementation of the asymptotic version of the Generalized Moment Selection (GMS) test described on page 135 of Andrews and Soares (2010). The content of this section follows closely that of Appendix A.7 in Dickstein and Morales (2018).

We base inference on the modified method of moments (MMM) statistic. Denote by  $\gamma$  a generic parameter with unknown true value  $\gamma^*$  for which we want to compute a  $1 - \delta$  confidence set. In our context,  $\gamma$  equals either  $\theta_\alpha$  or  $\theta_l$  for some  $l$ . Suppose we use  $k = 1, \dots, K$  inequalities to compute a confidence interval for  $\gamma$ . Denote each of these inequalities by  $\bar{m}_k(\gamma) \geq 0$ , where, for each  $k = 1, \dots, K$ ,

$$\bar{m}_k(\gamma) \equiv N^{-1} \sum_{c=1}^C \sum_{n=1}^{N_c} m_k(x_{nc}, \gamma), \quad (\text{B.11})$$

and where observations are grouped into clusters  $c = 1, \dots, C$  and indexed by  $n = 1, \dots, N_c$  within each cluster  $c$ . The variable  $x_{nc}$  is a generic vector of observed covariates. E.g., in the context of the moment in equation (22), each of the clusters  $c$  in equation (B.11) corresponds to a sector  $s$ , and each observation  $n$  within a cluster  $c$  corresponds to a worker  $i$  within a sector  $s$ . In the moment in equation (B.10), each cluster corresponds to a pair of sectors  $(s, r)$  and each observation corresponds to a pair of workers  $(i, j)$ . The MMM statistic is

$$T(\gamma) = \sum_{k=1}^K \left( \min \left\{ \sqrt{N} \frac{\bar{m}_k(\gamma)}{\hat{\sigma}_k(\gamma)}, 0 \right\} \right)^2, \quad (\text{B.12})$$

where  $\hat{\sigma}_k(\gamma) = \sqrt{\hat{\sigma}_k^2(\gamma)}$  and

$$\hat{\sigma}_k^2(\gamma) = \frac{1}{N} \sum_{c=1}^C \left( \sum_{n=1}^{N_c} (m_k(x_{nc}, \gamma) - \bar{m}_k(\gamma)) \right)^2.$$

Given a grid  $\Gamma_g$  covering the parameter space of  $\gamma$ , we implement the following steps:

**Step 1: choose a point**  $\gamma_p \in \Gamma_g$ . Steps 2 to 8 test:  $H_0 : \gamma^* = \gamma_p$  vs.  $H_1 : \gamma^* \neq \gamma_p$ .

**Step 2: evaluate the MMM test statistic at  $\gamma_p$ :**

$$T(\gamma_p) = \sum_{k=1}^K \left( \min \left\{ \sqrt{N} \frac{\bar{m}_k(\gamma_p)}{\hat{\sigma}_k(\gamma_p)}, 0 \right\} \right)^2. \quad (\text{B.13})$$

**Step 3: compute the correlation matrix of the moments evaluated at  $\gamma_p$ :**

$$\hat{\Omega}(\gamma_p) = \text{Diag}^{-\frac{1}{2}}(\hat{\Sigma}(\gamma_p)) \hat{\Sigma}(\gamma_p) \text{Diag}^{-\frac{1}{2}}(\hat{\Sigma}(\gamma_p)), \quad (\text{B.14})$$

with  $\text{Diag}(\hat{\Sigma}(\gamma_p))$  the  $L \times L$  diagonal matrix with diagonal elements equal to those of  $\hat{\Sigma}(\gamma_p)$ ,  $\text{Diag}^{-\frac{1}{2}}(\hat{\Sigma}(\gamma_p))$  a matrix such that  $\text{Diag}^{-\frac{1}{2}}(\hat{\Sigma}(\gamma_p)) \text{Diag}^{-\frac{1}{2}}(\hat{\Sigma}(\gamma_p)) = \text{Diag}^{-1}(\hat{\Sigma}(\gamma_p))$  and

$$\hat{\Sigma}(\gamma_p) = \frac{1}{N} \sum_{c=1}^C \left( \sum_{n=1}^{N_c} (m(x_{nc}, \gamma_p) - \bar{m}(\gamma_p)) \right) \left( \sum_{i=1}^{N_c} (m(x_{ic}, \gamma_p) - \bar{m}(\gamma_p)) \right)', \quad (\text{B.15})$$

where

$$m(x_{nc}, \gamma_p) = (m_1(x_{nc}, \gamma_p), \dots, m_K(x_{nc}, \gamma_p))', \quad (\text{B.16})$$

$$\bar{m}(\gamma_p) = (\bar{m}_1(\gamma_p), \dots, \bar{m}_K(\gamma_p))'. \quad (\text{B.17})$$

**Step 4: simulate the asymptotic distribution of  $T(\gamma_p)$ .** Take  $D$  draws from the distribution  $N(0_K, I_K)$  where  $0_K$  is a vector of 0s of dimension  $K$  and  $I_K$  is the identity

matrix of dimension  $L$ . Denote each draw as  $\zeta_d$ . Define the criterion function  $T_d^{AA}(\gamma_p)$  as

$$T_d^{AA}(\gamma_p) = \sum_{k=1}^K \left\{ (\min\{\hat{\Omega}^{\frac{1}{2}}(\gamma_p)\zeta_d\}_k, 0\})^2 \times \mathbb{1}\left\{\sqrt{N}\frac{\bar{m}_k(\gamma_p)}{\hat{\sigma}_k(\gamma_p)} \leq \sqrt{\ln N}\right\} \right\}$$

where  $[\hat{\Omega}^{\frac{1}{2}}(\gamma_p)\zeta_d]_k$  is the  $k$ -th element of the vector  $\hat{\Omega}^{\frac{1}{2}}(\gamma_p)\zeta_d$ .

**Step 5: compute critical value.** The critical value  $\hat{c}^{AA}(\gamma_p, 1 - \delta)$  is the  $(1 - \delta)$ -quantile of the distribution of  $T_d^{AA}(\gamma_p)$  across the  $D$  draws taken in the previous step.

**Step 6: accept/reject  $\gamma_p$ .** Include  $\gamma_p$  in confidence set,  $\hat{\Gamma}^{1-\delta}$ , if  $T(\gamma_p) \leq \hat{c}^{AA}(\gamma_p, 1 - \delta)$ .

**Step 7: repeat steps 2 to 6 for every  $\gamma_p$  in the grid  $\Gamma_g$ .**

**Step 8: compare  $\hat{\Gamma}^{1-\alpha}$  to  $\Gamma_g$ .** If no point in  $\hat{\Gamma}^{1-\alpha}$  is at the boundary of  $\Gamma_g$ , define  $\hat{\Gamma}^{1-\alpha}$  as the  $1 - \delta$  confidence set for  $\gamma^*$ . Otherwise, expand the limits of  $\Gamma_g$  and repeat steps 1 to 8.

## C Additional Simulation Results

In Appendix C.1, we describe the inequalities we use to obtain the results in Table 1. In Appendix C.2, we explore alternative ways of building the first-step inequalities. In Appendix C.4, we present estimates analogous to those in Table 1, but using a larger set of instruments. In Appendix C.3, we explore the robustness of the results in Table 1 to different values of amenities. In Appendix C.5, we compare our two-step estimator to other moment inequality estimators. In Appendix C.6, we compare the estimates in Table 1 to alternative estimates computed using approximations points other than those in equations (9) and (20).

### C.1 Inequalities Used in Computing the CIs in Table 1

*First step.* Given a predictor  $z_s^l$  of the wage level  $w_s^l$  in every location  $l$ , a pair of locations  $l$  and  $l'$ , and a pair of indices  $k$  and  $k'$  determining the instruments we use, we compute the CIs for  $\theta_\alpha$  displayed in Table 1 using moment inequalities of the type:

$$\sum_{s=1}^S (y_s^l y_{r(s)}^l + y_s^{l'} y_{r(s)}^{l'} - y_s^l y_{r(s)}^{l'} \exp(-e_s^{l'})) \\ (2 + 2e_s^{l'} - \theta_\alpha(\Delta w_s^{l'} + \Delta w_{r(s)}^{l'}))g_k(\Delta z_s^{l'})g_{k'}(\Delta z_{r(s)}^{l'}) \geq 0, \quad (\text{C.1})$$

where  $r(s)$  indexes the type matched with type  $s$ . In practice, we combine this type of moment inequality for every pair of locations in the set  $\{(l, l'); l \in \{1, 2, 3\}, l' \in \{1, 2, 3\}, l \neq l'\}$  and every pair of instrument indices in the set  $\{(k, k'); k \in \{1, 2\}, k' \in \{1, 2\}\}$ . Thus, we use 24 inequalities to identify  $\theta_\alpha$ . We now describe in more detail how we choose the type  $r(s)$

to match with each type  $s = 1, \dots, S$ ; how we define the approximation points  $e_s^{ll'}$  for each  $s = 1, \dots, S$ ; and how we define the instrument function  $g_k(\cdot)$  for  $k \in \{1, 2\}$ .

First, for each  $s = 1, \dots, S$ , we select  $r(s)$  randomly among those that satisfy

$$|\hat{\mathbb{E}}[\Delta w_s^{ll'} | \Delta z_s^{ll'}, y_s^l = 1] - \hat{\mathbb{E}}[\Delta w_{r(s)}^{ll'} | \Delta z_{r(s)}^{ll'}, y_{r(s)}^{ll'} = 1]| \leq \tau, \quad (\text{C.2})$$

with  $\tau = 0.002$ . In this equation, for example,  $\hat{\mathbb{E}}[\Delta w_s^{ll'} | \Delta z_s^{ll'}, y_s^l = 1]$  is the predicted value of  $\Delta w_s^{ll'}$  computed using a linear regression of  $\Delta w_s^{ll'}$  on  $\Delta z_s^{ll'}$  estimated on the subset of observations with  $y_s^l = 1$ . To understand why we impose the restriction in equation (C.2) when selecting the type  $r(s)$  to match with each given type  $s$ , one should note that, as  $S$  goes to infinity, the moment inequality in equation (C.1) will be satisfied at the true value of the parameter  $\theta_\alpha$  regardless of how the type  $r(s)$  for every  $s = 1, \dots, S$  is chosen. However, Corollary 3 indicates that a condition for this inequality to point identify  $\theta_\alpha$  in certain cases is that, given wage predictors  $z_s^l$  and  $z_r^l$  for the types  $s$  and  $r$  combined in the inequality, these wage predictors and types satisfy

$$\mathbb{E}[\Delta w_s^{ll'} | \mathcal{W}_s] = \mathbb{E}[\Delta w_s^{ll'} | z_s] = \mathbb{E}[\Delta w_r^{ll'} | \mathcal{W}_r] = \mathbb{E}[\Delta w_r^{ll'} | z_r]. \quad (\text{C.3})$$

By selecting the type  $r$  that we match with each type  $s$  randomly among those satisfying the restriction in equation (C.2), we aim to approximate the condition in equation (C.3) while taking into account the limitation that the sets  $\mathcal{W}_s$  and  $\mathcal{W}_r$  are generally not observed.

Second, in terms of the approximation point  $e_s^{ll'}$  for each type  $s$  used in equation (C.1), we build on equation (20) and impose

$$e_s^{ll'} = \theta_\alpha 0.5 (\hat{\mathbb{E}}[\Delta w_s^{ll'} | \Delta z_s^{ll'}, y_s^l = 1] + \hat{\mathbb{E}}[\Delta w_{r(s)}^{ll'} | \Delta z_{r(s)}^{ll'}, y_{r(s)}^{ll'} = 1]), \quad (\text{C.4})$$

where, as indicated above, e.g.,  $\hat{\mathbb{E}}[\Delta w_s^{ll'} | \Delta z_s^{ll'}, y_s^l = 1]$  is the predicted value of  $\Delta w_s^{ll'}$  computed using a linear regression of  $\Delta w_s^{ll'}$  on  $\Delta z_s^{ll'}$  estimated on the subset of observations with  $y_s^l = 1$ . Note that equation (20) indicates that the optimal approximation point  $e_s^{ll'}$  is a function of, for example, the conditional expectation of  $\Delta w_s^{ll'}$  given  $\Delta z_s^{ll'}$  and  $y_s^l = 1$ ; in equation (C.4), we compute instead a linear regression. In unreported results, we have instead approximated nonparametrically the conditional expectation of  $\Delta w_s^{ll'}$  given  $\Delta z_s^{ll'}$  and  $y_s^l = 1$ , obtaining very similar results to those attained when using linear regressions instead.

Third, and finally, in terms of the function  $g_k(\cdot)$  for  $k = 1, 2$ , we impose:

$$g_k(\Delta z_s^{ll'}) = \begin{cases} \mathbb{1}\{\Delta z_s^{ll'} \leq 0\} & \text{for } k = 1, \\ \mathbb{1}\{0 < \Delta z_s^{ll'}\} & \text{for } k = 2. \end{cases} \quad (\text{C.5})$$

This instrument function corresponds to that in Section 3.3 with  $q = 2$  and  $d = 0$ .

*Second step.* Given a location  $l$ , a predictor  $z_s^l$  of  $w_s^l$ , an index  $k$  determining the instruments we use, and a value  $\check{\theta}_\alpha$  that belongs to the CI for  $\theta_\alpha$  computed in the first-step of our inference procedure, we compute CIs for  $\theta_l$  using either the following bounding inequalities

$$\sum_{s=1}^S (y_s^1 - y_s^l \exp(-e_s^{l1})(1 + e_s^{l1} - (\theta_l + \check{\theta}_\alpha \Delta w_s^{l1}))) g_k(\Delta z_s^{l1}) \geq 0, \quad (\text{C.6a})$$

$$\sum_{s=1}^S (y_s^l - y_s^1 \exp(-e_s^{1l})(1 + e_s^{1l} + (\theta_l + \check{\theta}_\alpha \Delta w_s^{1l}))) g_k(\Delta z_s^{1l}) \geq 0; \quad (\text{C.6b})$$

or the following odds-based moment inequalities

$$\sum_{s=1}^S (y_s^l \exp(-(\theta_l + \check{\theta}_\alpha \Delta w_s^{l1})) - y_s^1) g_k(\Delta z_s^{l1}) \geq 0, \quad (\text{C.7a})$$

$$\sum_{s=1}^S (y_s^1 \exp(\theta_l + \check{\theta}_\alpha \Delta w_s^{1l}) - y_s^l) g_k(\Delta z_s^{1l}) \geq 0; \quad (\text{C.7b})$$

or both the bounding and odds-based moment inequalities jointly. The instrument function  $g_k(\cdot)$  in equations (C.6) and (C.7) is defined as in equation (C.5). Thus, the CIs for  $\theta_l$  are computed use either four bounding inequalities, four odds-based inequalities, or all eight inequalities together. The points  $e_s^{l1}$  and  $e_s^{1l}$  entering the inequalities in equation (C.6) are

$$e_s^{l1} = \theta_l + \check{\theta}_\alpha \hat{\mathbb{E}}[\Delta w_s^{l1} | \Delta z_s^{l1}, y_{is}^l = 1], \quad (\text{C.8a})$$

$$e_s^{1l} = -\theta_l + \check{\theta}_\alpha \hat{\mathbb{E}}[\Delta w_s^{1l} | \Delta z_s^{1l}, y_{is}^1 = 1], \quad (\text{C.8b})$$

where, as indicated when discussing equation (C.2) above, for example,  $\hat{\mathbb{E}}[\Delta w_s^{ll'} | \Delta z_{2s}^{ll'}, y_s^l = 1]$  is the predicted value of  $\Delta w_s^{ll'}$  computed using a linear regression of  $\Delta w_s^{ll'}$  on  $\Delta z_s^{ll'}$  estimated on the subset of observations with  $y_s^l = 1$ .

*Differences across cases.* In cases 1 to 4 in Table 1, the wage predictor  $z_s^l$  introduced in equations (C.1) to (C.8) equals the wage shifter  $z_{2s}^l$  in equation (24). In case 5, the wage predictor  $z_s^l$  equals the wage itself  $w_s^l$ . In this case, equation (C.2) simplifies to  $|\Delta w_s^{ll'} - \Delta w_{r(s)}^{ll'}| \leq \tau$ ; equation (C.4) simplifies to  $e_s^{ll'} = \theta_\alpha 0.5(\Delta w_s^{ll'} + \Delta w_{r(s)}^{ll'})$ ; and equation (C.8) simplifies to  $e_s^{l1} = \theta_l + \check{\theta}_\alpha \Delta w_s^{l1}$  and  $e_s^{1l} = -e_s^{l1}$ .

## C.2 First-step Moment Inequalities with Loose Type Matches

In Table C.1, we present CIs for  $\theta_\alpha$  computed according to equations (C.1) to (C.5), and show how these CIs vary as we change the value of  $\tau$  entering equation (C.2). Table C.1

shows that the 95% CI for  $\theta_\alpha$  becomes wider as we increase the value of  $\tau$ .

Table C.1: Simulation Results - Moment Inequality Confidence Intervals With Loose Matches

| Case | $\sigma_1$ | $\sigma_3$ | $z_s^l$    | $\tau$ | 1st Step        |
|------|------------|------------|------------|--------|-----------------|
|      |            |            |            |        | $\theta_\alpha$ |
| 2    | 0          | 1          | $z_{2s}^l$ | 8      | [0.73, 1.32]    |
| 2    | 0          | 1          | $z_{2s}^l$ | 4      | [0.79, 1.25]    |
| 2    | 0          | 1          | $z_{2s}^l$ | 2      | [0.94, 1.08]    |
| 2    | 0          | 1          | $z_{2s}^l$ | 1      | [0.98, 1.03]    |
| 2    | 0          | 1          | $z_{2s}^l$ | 0.8    | [0.99, 1.03]    |
| 2    | 0          | 1          | $z_{2s}^l$ | 0.08   | [1, 1.02]       |
| 2    | 0          | 1          | $z_{2s}^l$ | 0.008  | [1, 1.02]       |
| 2    | 0          | 1          | $z_{2s}^l$ | 0.002  | [1, 1.01]       |

The column  $\theta_\alpha$  contains 95% CIs computed using the inequalities described in Appendix C.1 and the inference procedure in Andrews and Soares (2010). The CI with  $\tau = 0.002$  corresponds to that in Table 1.

### C.3 Amenity Differences Across All Locations

In Table C.2, we present CIs for  $\theta_\alpha$  computed according to equations (C.1) to (C.5), and show how these CIs vary as we change the value of the amenity parameters ( $\kappa^1, \kappa^2, \kappa^3$ ) used to generate the simulated data (see Section 4 for details). Table C.2 shows that the 95% CI for  $\theta_\alpha$  as the minimum value of  $\Delta\kappa^{ll'}$  between any two locations  $l$  and  $l'$  increases; i.e., as  $\min_{l \in \{1,2,3\}, l' \neq l} \Delta\kappa^{ll'}$  increases.

Table C.2: Simulation Results - Confidence Intervals With Amenity Differences

| Case | $\sigma_1$ | $\sigma_3$ | $z_s^l$    | $(\kappa^1, \kappa^2, \kappa^3)$ | 1st Step        |
|------|------------|------------|------------|----------------------------------|-----------------|
|      |            |            |            |                                  | $\theta_\alpha$ |
| 2    | 0          | 1          | $z_{2s}^l$ | (0, 0, 1)                        | [1, 1.01]       |
| 2    | 0          | 1          | $z_{2s}^l$ | (0, 0.5, 1)                      | [0.92, 1.09]    |
| 2    | 0          | 1          | $z_{2s}^l$ | (0, 0, 2)                        | [1, 1.02]       |
| 2    | 0          | 1          | $z_{2s}^l$ | (0, 1, 2)                        | [0.73, 1.29]    |
| 2    | 0          | 1          | $z_{2s}^l$ | (0, 0, 3)                        | [1, 1.02]       |
| 2    | 0          | 1          | $z_{2s}^l$ | (0, 1.5, 3)                      | [0.55, 1.49]    |

The column  $\theta_\alpha$  contains 95% CIs computed using the inequalities described in Appendix C.1 and the inference procedure in Andrews and Soares (2010). The CI with  $(\kappa^1, \kappa^2, \kappa^3) = (0, 0, 1)$  corresponds to that in Table 1.

### C.4 Two-step Moment Inequalities with Additional Instruments

In Table C.3, we present CIs analogous to those in Table 1 with the only difference that, instead of using the instrument function in equation (C.5), we use the following instrument

function

$$g_k(\Delta z_s^{ll'}) = \begin{cases} \mathbb{1}\{\Delta z_s^{ll'} \leq Q_{25}(\Delta z_s^{ll'})\} & \text{for } k = 1, \\ \mathbb{1}\{Q_{25}(\Delta z_s^{ll'}) < \Delta z_s^{ll'} \leq Q_{50}(\Delta z_s^{ll'})\} & \text{for } k = 2, \\ \mathbb{1}\{Q_{50}(\Delta z_s^{ll'}) < \Delta z_s^{ll'} \leq Q_{75}(\Delta z_s^{ll'})\} & \text{for } k = 3, \\ \mathbb{1}\{Q_{75}(\Delta z_s^{ll'}) < \Delta z_s^{ll'}\} & \text{for } k = 4, \end{cases} \quad (\text{C.9})$$

where  $Q_{25}(\Delta z_s^{ll'})$ ,  $Q_{50}(\Delta z_s^{ll'})$ , and  $Q_{75}(\Delta z_s^{ll'})$  respectively denote the percentiles 25, 50, and 75 of the distribution of  $\Delta z_s^{ll'}$  across all types  $s$  and pairs of locations  $l$  and  $l'$ . As a comparison of the results in tables 1 and C.3 illustrates, the CIs for  $\theta_\alpha$ ,  $\theta_2$ , and  $\theta_3$ , become tighter when we swap the instrument function in equation (C.5) for the more detailed function in equation (C.9). This change in instrument functions results in an increase in the number of moment inequalities we use in our estimation.

Table C.3: Simulation Results - Confidence Intervals With Additional Instruments

| Case | $\sigma_1$ | $\sigma_3$ | $z_s^l$    | First Step      | Second Step |               |              |
|------|------------|------------|------------|-----------------|-------------|---------------|--------------|
|      |            |            |            | $\theta_\alpha$ | Mom. Ineq.  | $\theta_2$    | $\theta_3$   |
| 1    | 0          | 0          | $z_{2s}^l$ | [1, 1.02]       | Bounding    | [0, 0]        | [1, 1]       |
|      |            |            |            |                 | Odds-based  | [0, 0]        | [1, 1]       |
|      |            |            |            |                 | Both        | [0, 0]        | [1, 1]       |
| 2    | 0          | 1          | $z_{2s}^l$ | [1, 1.01]       | Bounding    | [0, 0]        | [1, 1]       |
|      |            |            |            |                 | Odds-based  | [-0.33, 0.32] | [0.68, 1.33] |
|      |            |            |            |                 | Both        | [0, 0]        | [1, 1]       |
| 3    | 1          | 0          | $z_{2s}^l$ | [0.91, 1.15]    | Bounding    | [-0.31, 0.31] | [0.70, 1.30] |
|      |            |            |            |                 | Odds-based  | [0, 0]        | [1, 1.01]    |
|      |            |            |            |                 | Both        | [0, 0]        | [1, 1.01]    |
| 4    | 1          | 1          | $z_{2s}^l$ | [0.91, 1.19]    | Bounding    | [-0.31, 0.31] | [0.70, 1.30] |
|      |            |            |            |                 | Odds-based  | [-0.32, 0.32] | [0.68, 1.33] |
|      |            |            |            |                 | Both        | [-0.31, 0.31] | [0.70, 1.31] |
| 5    | 0          | 1          | $w_s^l$    | $\emptyset$     | Bounding    | $\emptyset$   | $\emptyset$  |
|      |            |            |            |                 | Odds-based  | $\emptyset$   | $\emptyset$  |
|      |            |            |            |                 | Both        | $\emptyset$   | $\emptyset$  |

This table contains 95% CIs computed using the inequalities described in Appendix C.1 with the only exception that the instrument functions  $g_k(\cdot)$  for all  $k = 1, \dots, K$  are not those defined in equation (C.5) but those defined in equation (C.9).

## C.5 Alternative Moment Inequality Estimators

*One-step moment inequality estimator when choice set is small.* The two-step estimator described in Section 3 yields CIs for all parameters that are computationally feasible in

discrete-choice settings with many choices and choice-specific fixed effects. As [Dickstein et al. \(2023\)](#) show, when the choice set is small, as in our simulation setting, the inequalities in equations (8) and (13) can be used to identify all parameters in one step: the wage coefficient is treated as an unknown parameter, and the inequalities in Section 3.1 are used to estimate a confidence set for  $(\theta_\alpha, \theta_2, \theta_3)$ . In Table C.4, we reproduce content in [Dickstein et al. \(2023\)](#) and report the projection on each parameter of the three-dimensional confidence set computed following that procedure. Comparing tables 1 and C.4, we see that our two-step procedure yields similar CIs for  $\theta_2$  and  $\theta_3$ , but larger ones for  $\theta_\alpha$  when agents' information sets are partly unobserved by the researcher (i.e., when  $\sigma_1 > 0$ ).

Table C.4: Simulation Results - Confidence Intervals When Choice Set is Small

| Case | $\sigma_1$ | $\sigma_3$ | $z_s^l$    | Mom. Ineq.                      | $\theta_\alpha$                                    | $\theta_2$                                            | $\theta_3$                                         |
|------|------------|------------|------------|---------------------------------|----------------------------------------------------|-------------------------------------------------------|----------------------------------------------------|
| 1    | 0          | 0          | $z_{2s}^l$ | Bounding<br>Odds-based<br>Both  | $[1, 1]$<br>$[1, 1]$<br>$[1, 1]$                   | $[0, 0]$<br>$[0, 0]$<br>$[0, 0]$                      | $[1, 1]$<br>$[1, 1]$<br>$[1, 1]$                   |
| 2    | 0          | 1          | $z_{2s}^l$ | Bounding<br>Odds-based*<br>Both | $[1, 1]$<br>$[0.92, 1.50]$<br>$[1, 1]$             | $[0, 0]$<br>$[-0.33, 0.33]$<br>$[0, 0]$               | $[1, 1]$<br>$[0.67, 1.33]$<br>$[1, 1]$             |
| 3    | 1          | 0          | $z_{2s}^l$ | Bounding<br>Odds-based*<br>Both | $[0.80, 1.10]$<br>$[1, 1]$<br>$[1, 1]$             | $[-0.30, 0.30]$<br>$[0, 0]$<br>$[0, 0]$               | $[0.70, 1.30]$<br>$[1, 1]$<br>$[1, 1]$             |
| 4    | 1          | 1          | $z_{2s}^l$ | Bounding<br>Odds-based*<br>Both | $[0.80, 1.10]$<br>$[0.92, 1.50]$<br>$[0.92, 1.10]$ | $[-0.30, 0.30]$<br>$[-0.48, 0.50]$<br>$[-0.33, 0.30]$ | $[0.70, 1.30]$<br>$[0.65, 1.50]$<br>$[0.70, 1.30]$ |
| 5    | 0          | 1          | $w_s^l$    | Bounding<br>Odds-based<br>Both  | $[0.87, 0.87]$<br>$\emptyset$<br>$\emptyset$       | $[-0.05, -0.03]$<br>$\emptyset$<br>$\emptyset$        | $[0.85, 0.88]$<br>$\emptyset$<br>$\emptyset$       |

This table contains projections of 95% confidence sets for  $(\theta_\alpha, \theta_2, \theta_3)$  computed using moment inequalities analogous to those in equations (8) and (13), with the only difference that the wage coefficient is treated as an unknown parameter. Confidence sets are computed following the procedure in [Andrews and Soares \(2010\)](#) using a 3-dimensional grid with sides  $[0.5, 1.5]$  (for  $\theta_\alpha$ ),  $[-0.5, 0.5]$  (for  $\theta_2$ ) and  $[0.5, 1.5]$  (for  $\theta_3$ ). We mark with an asterisk when the confidence set includes points at the boundary of the grid. Results are from [Dickstein et al. \(2023\)](#).

*Alternative moment inequality estimator when the choice set is large.* In Table C.5, we use the inequalities in equations (8) and (13) to compute two confidence sets, one for  $(\theta_\alpha, \theta_2)$  and another one for  $(\theta_\alpha, \theta_3)$ . After projecting each of these confidence sets on each of their two dimensions, we obtain a CI for each of the amenities  $\theta_2$  and  $\theta_3$ , and two CIs for  $\theta_\alpha$ . In Table C.5, we report the CIs for  $\theta_2$  and  $\theta_3$ , and a randomly chosen CI for  $\theta_\alpha$  among the two available ones. This inference procedure yields results similar to those in Table 1 except in case 3, where the fact that this alternative estimator adds information from the odds-based

inequalities for the identification of  $\theta_\alpha$  results in a tighter CI for this parameter.

Table C.5: Simulation Results - Alternative Confidence Intervals With Large Choice Set

| Case | $\sigma_1$ | $\sigma_3$ | $z_s^l$    | Mom. Ineq.                      | $\theta_\alpha$                                    | $\theta_2$                                            | $\theta_3$                                         |
|------|------------|------------|------------|---------------------------------|----------------------------------------------------|-------------------------------------------------------|----------------------------------------------------|
| 1    | 0          | 0          | $z_{2s}^l$ | Bounding<br>Odds-based<br>Both  | $[1, 1]$<br>$[1, 1]$<br>$[1, 1]$                   | $[0, 0]$<br>$[0, 0]$<br>$[0, 0]$                      | $[1, 1]$<br>$[1, 1]$<br>$[1, 1]$                   |
| 2    | 0          | 1          | $z_{2s}^l$ | Bounding<br>Odds-based*<br>Both | $[1, 1]$<br>$[0.91, 1.50]$<br>$[1, 1]$             | $[0, 0]$<br>$[-0.33, 0.32]$<br>$[0, 0]$               | $[1, 1]$<br>$[0.68, 1.33]$<br>$[1, 1]$             |
| 3    | 1          | 0          | $z_{2s}^l$ | Bounding<br>Odds-based*<br>Both | $[0.80, 1.10]$<br>$[1, 1]$<br>$[1, 1]$             | $[-0.31, 0.31]$<br>$[0, 0]$<br>$[0, 0]$               | $[0.70, 1.30]$<br>$[1, 1.01]$<br>$[1, 1.01]$       |
| 4    | 1          | 1          | $z_{2s}^l$ | Bounding<br>Odds-based*<br>Both | $[0.79, 1.10]$<br>$[0.92, 1.50]$<br>$[0.92, 1.10]$ | $[-0.31, 0.31]$<br>$[-0.49, 0.50]$<br>$[-0.31, 0.31]$ | $[0.69, 1.30]$<br>$[0.64, 1.50]$<br>$[0.69, 1.31]$ |
| 5    | 0          | 1          | $w_s^l$    | Bounding<br>Odds-based<br>Both  | $[0.87, 0.88]$<br>$\emptyset$<br>$\emptyset$       | $[-0.05, -0.01]$<br>$\emptyset$<br>$\emptyset$        | $[0.86, 0.88]$<br>$\emptyset$<br>$\emptyset$       |

This table contains projections of 95% confidence sets for  $(\theta_\alpha, \theta_2)$  and  $(\theta_\alpha, \theta_3)$  computed using moment inequalities analogous to those in equations (8) and (13), with the only difference that the wage coefficient is treated as an unknown parameter. Confidence sets are computed following the procedure in [Andrews and Soares \(2010\)](#) using a 3-dimensional grid with sides  $[0.5, 1.5]$  (for  $\theta_\alpha$ ),  $[-0.5, 0.5]$  (for  $\theta_2$ ) and  $[0.5, 1.5]$  (for  $\theta_3$ ). We mark with an asterisk when the confidence set includes points at the boundary of the grid.

## C.6 Estimator with Alternative Approximation Points

Panel A in Table C.6 presents CIs computed similarly to those in Table 1. The results in Panel B incorporate two differences in the way these CIs are computed: (a) instead of using the approximation points in equation (C.4) in the first-step moment inequalities, we use

$$e_s^{ll'} = \theta_\alpha 0.5(\Delta z_s^{ll'} + \Delta z_{r(s)}^{ll'}); \quad (\text{C.10})$$

and, instead of using the approximation points in equation (C.8) in the second-step bounding moment inequalities, we use

$$e_s^{l1} = \theta_l + \check{\theta}_\alpha \Delta z_s^{l1}, \quad (\text{C.11a})$$

$$e_s^{1l} = -\theta_l + \check{\theta}_\alpha \Delta z_s^{1l}. \quad (\text{C.11b})$$

Relative to the approximation points in equations (C.4) and (C.8), those in equations (C.10) and (C.11) have the advantage that they do not involve any estimation; in particular, they

Table C.6: Simulation Results - Confidence Intervals With Alternative Approximation Points

| $\sigma_1$                                                                                | $\sigma_3$ | $z_{2s}^l$ | Mom. Ineq. | $\theta_\alpha$ | $\theta_2$    | $\theta_3$   |
|-------------------------------------------------------------------------------------------|------------|------------|------------|-----------------|---------------|--------------|
| <i>Panel A: Confidence Intervals Using Approximation Points in Eqs. (C.4) and (C.8)</i>   |            |            |            |                 |               |              |
| 0                                                                                         | 0          | $z_{2s}^l$ | Bounding   | [1, 1.02]       | [0, 0]        | [1, 1]       |
| 0.25                                                                                      | 0          | $z_{2s}^l$ | Bounding   | [0.98, 1.04]    | [-0.02, 0.02] | [0.98, 1.02] |
| 0.50                                                                                      | 0          | $z_{2s}^l$ | Bounding   | [0.95, 1.06]    | [-0.09, 0.08] | [0.92, 1.08] |
| 0.75                                                                                      | 1          | $z_{2s}^l$ | Bounding   | [0.89, 1.16]    | [-0.18, 0.18] | [0.82, 1.18] |
| 1                                                                                         | 0          | $z_{2s}^l$ | Bounding   | [0.82, 1.29]    | [-0.31, 0.31] | [0.70, 1.30] |
| <i>Panel B: Confidence Intervals Using Approximation Points in Eqs. (C.10) and (C.11)</i> |            |            |            |                 |               |              |
| 0                                                                                         | 0          | $z_{2s}^l$ | Bounding   | [1, 1.02]       | [0, 0]        | [1, 1]       |
| 0.25                                                                                      | 0          | $z_{2s}^l$ | Bounding   | [0.98, 1.04]    | [-0.02, 0.02] | [0.98, 1.02] |
| 0.50                                                                                      | 0          | $z_{2s}^l$ | Bounding   | [0.94, 1.09]    | [-0.09, 0.08] | [0.92, 1.08] |
| 0.75                                                                                      | 1          | $z_{2s}^l$ | Bounding   | [0.88, 1.24]    | [-0.20, 0.20] | [0.81, 1.21] |
| 1                                                                                         | 0          | $z_{2s}^l$ | Bounding   | [0.81, 1.50]    | [-0.41, 0.42] | [0.63, 1.46] |

Panel A contains 95% CIs computed using the inequalities described in Appendix C.1. Panel B contains 95% CIs computed using inequalities analogous to those Appendix C.1, with the only exception that the approximation points in equations (C.10) and (C.11) are used instead of those in equations (C.4) and (C.8). As the choice of approximation points only affects the CIs computed using bounding moment inequalities (i.e., they do not affect the CIs computed using odds-based moment inequalities), we only report those in the table.

do not require computing predicted values from a linear regression. Consequently, they do not affect the validity of standard moment inequality inference procedures. However, as a comparison of both panels in Table C.6 illustrates, the use of the approximation points in equations (C.10) and (C.11) results in CIs that are larger than when the approximations points in equations (C.4) and (C.8) are used. The difference between both sets of CIs increases in the value of  $\sigma_1$ .

## D Data and Summary Statistics

### D.1 Data Sources and Sample Construction

*The RAIS data.* Our main data source is the *Relação Anual de Informações Sociais* (RAIS), an administrative dataset maintained by Brazil’s Ministry of Labor. It includes the universe of formal employment spells in the private and public sectors. Individual workers are iden-

tified by government-issued identification numbers (PIS/PASEP and CPF), allowing us to track them as they change employers. For all spells observed between 1993 and 2011, we use information on their start and end dates, average monthly wage, number of work hours stipulated in the contract, 2-digit sector (according to the *Classificação Nacional de Atividades Econômicas*, CNAE), and information on the worker’s gender, age, race, and education level. All information is reported by the employers.

We have no information on workers without formal jobs. These workers may be employed in the informal sector, self-employed, unemployed, or out of the labor force. Based on the 2010 Census, 51% of the Brazilian labor force was in the formal sector. The implied total number of formal workers in the Census closely matches the number of workers at RAIS.

*Geography and wage definitions.* To determine workers’ location, we use the microregion of the *establishment* at which the worker is employed. Microregions are groups of municipalities that span the entirety of the Brazilian territory. They are defined by the *Instituto Brasileiro de Geografia e Estatística* (IBGE). During our sample period, Brazil had 558 microregions. While RAIS does not contain information on the residence of workers, [Dix-Carneiro and Kovak \(2017\)](#) use 2000 Census data to show that only 3.4% of individuals lived and worked in different microregions. Previous research has used microregions as local labor markets (e.g., [Dix-Carneiro, 2014](#); [Dix-Carneiro and Kovak, 2017, 2019](#); [Felix, 2022](#); [Szerman, 2024](#)).

Workers may hold multiple employment spells (jobs) in a year. To obtain a dataset in which each observation corresponds to a worker and a year, we assign to each worker-year pair the microregion and sector corresponding to the job that the worker held for the most extended period during that year. We compute the wage of a worker in a year by adding the labor income earned in every job this worker had in the corresponding year. We calculate the labor income of a worker in each of their jobs by transforming the average monthly wages reported for that job into a measure of average daily wages, and multiplying this one by the total number of days worked in the job reported in the data. If no start and end date is provided, we assume that these are January 1 and December 31, respectively.

*Sample restrictions and sampling.* We limit our data to workers aged 25 to 64, as these are the workers most likely to have completed their education and not have yet retired. The sample period is 2002-2011. We use 1993-2001 information to measure each worker’s experience in each sector and microregion. To limit our data to workers with a sufficiently close labor relationship with the formal sector, we restrict our sample to workers observed at RAIS for at least seven years in the sample period. We also restrict our sample to workers with similar demographic characteristics; specifically, we focus on workers with at least a high school degree identified as male and white. For computational reasons, we focus on a sample of 10 million worker-year pairs. To ensure we observe a large enough number of

individuals per market, we focus on 1,000 labor markets consisting of all combinations of the 50 microregions (out of 558) and 20 sectors (out of 51) with the largest total employment reported in RAIS. We then obtain our sample by randomly sampling 1 million individuals per sample year among those employed in the 1,000 labor markets of interest.

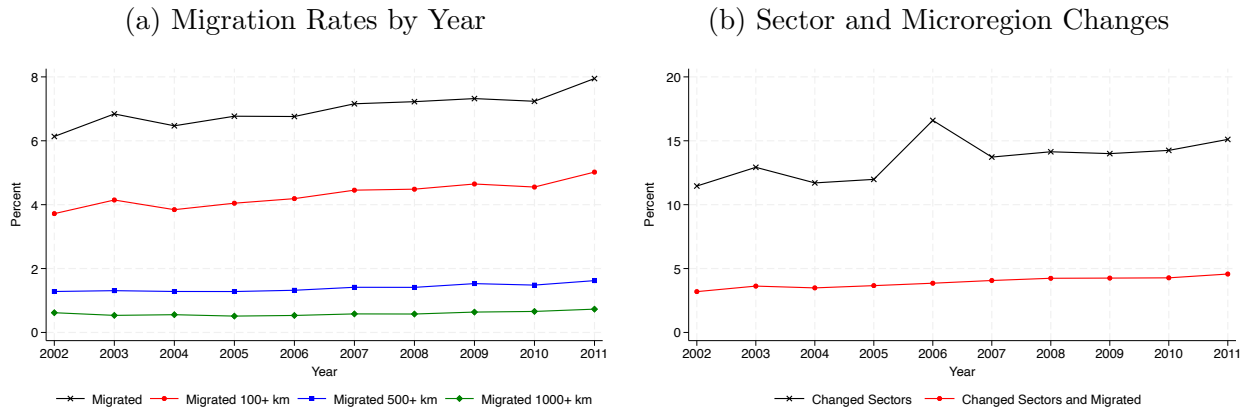
*Additional data sources.* We measure distances between microregions by the geodesic distance between their population centroids and data from the IBGE. Data on internet connections is from the *Agência Nacional de Telecomunicações* (ANATEL), which provides the number of broadband connections by municipality and year from 2007 onwards. We define internet access at the microregion level as the average share of households with broadband internet access in the 2007-2011 period.

## D.2 Summary Statistics: Migration

Consistently with our estimation sample, this section focuses on white male workers with at least a high school education in the 2002-2011 period. However, statistics are computed using information on all workers with those demographic characteristics; i.e., not only those linked to the 50 largest microregions and 20 largest sectors.

Figure D.1a provides yearly migration rates; i.e., share of workers that change microregion of employment between years  $t$  and  $t - 1$ . It shows an upward trend over the sample period, from close to 6% in 2002 to 8% in 2011. Figure D.1a also provides migration rates conditional on the distance between origin and destination microregions: about a third of moves are to microregions within 100 km from the origin, and less than a sixth of moves involve migration

Figure D.1: Migration and Sector Changes, by Year

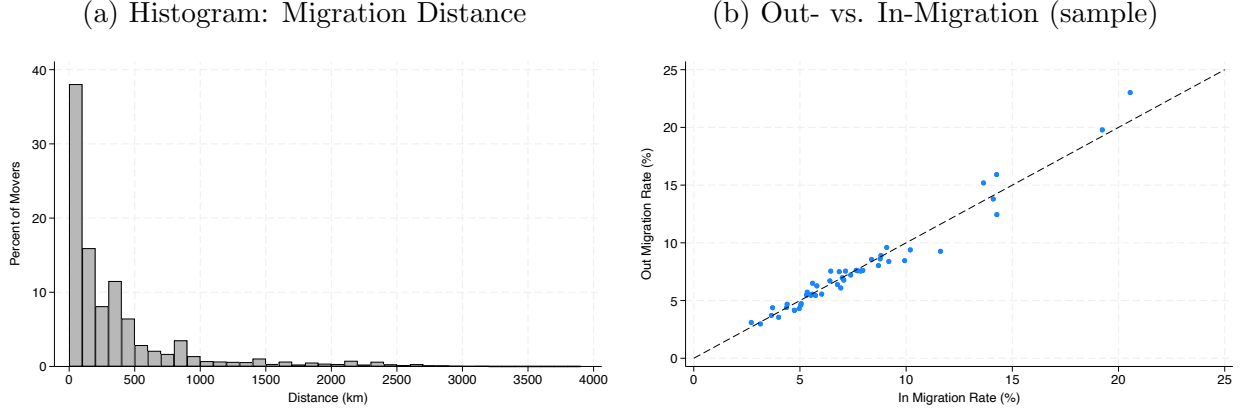


Panel (a) shows migration rates, both aggregate and by distance. Panel (b) shows the share of workers that changed sectors from the previous year (top line) and the share that both changed sectors and migrated (bottom line). Data includes all white male workers with at least a high school education.

over a distance larger than 1,000 km. Figure D.1b provides a similar figure for sectoral changes, which are more common. It also provides the share of workers that change both microregion and sector of employment in a given year, revealing that most changes in the employment sector are not accompanied by migration.

Figure D.2a depicts the distribution of distances between origin and destination microregions for those who migrate. Although more than half of moves occur between microregions within 200 km of distance, a sizable share occur at larger distances. Figure D.2b provides a scatter plot depicting the distribution of in-migration and out-migration rates for the 50 microregions in our sample. Each marker represents one of these microregions. In- and out-migration rates are strongly correlated across microregions, varying close to one-to-one. The figure also shows that the bulk of microregions have migration rates in the 3 to 12% range.

Figure D.2: Migration Patterns



Panel (a) provides a histogram of distances between origin and destination microregions for observed migrations. In Panel (b), each marker represents one of the 50 microregions in our sample. The y-axis measures the out-migration rate, while the x-axis measures the in-migration rate. It considers migration with origins and destinations to all microregions (including outside the 50 largest ones). Dashed line represents the 45 degree line. In both panels, data includes all white male workers with at least a high school education.

## E Appendix to Empirical Analysis

### E.1 Wage Proxies

We assume a worker's log wage  $w_{it}^{sl}$  can be expressed according to the following expression:

$$w_{it}^{sl} = w_t^{sl} + \underbrace{\alpha_i^s + \beta_e^s exp_{it}^s + \beta_{ee}^s (exp_{it}^s)^2 + \beta_a^s age_{it} + \beta_{aa}^s age_{it}^2}_{\text{sector-specific skill}} + \nu_{it}^{sl}, \quad (\text{E.1})$$

where  $w_t^{sl}$  (which we label the wage *shifter*) is a labor market-specific term, common to all workers, and  $\nu_{it}^{sl}$  is a worker-specific wage shock unexpected to workers at the time of making their location choices. We model the individual-by-sector-by-year wage component as the sum of an individual-by-sector fixed effect and a flexible function of the worker’s age and experience in the corresponding sector. We measure a worker’s experience in a sector using information on their path of employment from 1993 onwards. Hence, our measure of experience is capped at 8 years at the beginning of our analysis period in 2002 and at 17 years in 2011.

We estimate equation (E.1) on the universe of (white males with at least high school education) workers from the 1,000 labor markets included in our sample. That is, we estimate equation (E.1) before extracting the sample of 1 million individuals per year that we use in our moment inequality estimation. This dataset has 15,313,848 observations. As all coefficients in equation (E.1) are indexed by sector, we estimate them running separate regressions for each of the 20 sectors considered in our analysis. The median  $R^2$  across the 20 sector-specific regressions is 0.83.

## E.2 Implementation of Moment Inequalities

*Approximation points.* Computing the moments in equations (7) and (17) requires specifying the approximation points  $e_{is}^{ll'}$  and  $e_{isjr}^{ll'}$ , respectively. Equations (9) and (20) provide functional forms for these approximation points that, according to corollaries 1 and 3, result in second-step bounding inequalities and first-step inequalities, respectively, that can point identify the parameters of interest. Furthermore, as discussed in Appendix sections B.1 and B.5, the approximation points in equations (9) and (20) yield the tightest identified sets among all approximation points in a family of functions described in detail in those sections. The expressions in equations (9) and (20), however, depend on the expectation of specific wage differences conditional on the wage predictor used to build the corresponding inequality. Since we ignore the true value of those expectations, we must approximate them in some way. In our empirical application, we use approximation points that are simple to compute and that, importantly, do not depend on any regression estimate.<sup>30</sup>

Specifically, we compute the approximation points entering the bounding moment in-

---

<sup>30</sup>This stands in contrast with the approach we follow in our simulation setting when computing the CIs reported in Table 1, where we approximate the conditional expectations entering equations (9) and (20) using linear regressions; see Appendix C.1 for more details. While the large number of observations we use in our simulation setting implies that any noise in the linear regression estimates will have a minimal impact in our estimated CIs, this may not be true in our empirical application. Thus, to simplify the computation of the moment inequality CIs in our application, we restrict ourselves to using approximation points that are not functions of prior estimates.

equalities in equation (7) as

$$e_{is}^{l'} = \psi_0 + \psi_1 \Delta z_{st}^{l'}, \quad (\text{E.2})$$

where  $\psi_0$  and  $\psi_1$  are constants the researcher chooses. In practice, we use simultaneously in estimation several different bounding inequalities of the type in equation (7) computed using the approximation points in equation (E.2) with different values of the constants  $\psi_0$  and  $\psi_1$ . Similarly, we compute the approximation points entering the first-step moment inequalities in equation (17) as

$$e_{is}^{l'} = \psi_2 0.5(\Delta z_{st}^{l'} + \Delta z_{rt}^{l'}), \quad (\text{E.3})$$

where  $\psi_2$  is a constant. In practice, we also use simultaneously in estimation several different first-step moment inequalities of the type in equation (17) computed using the approximation points in equation (E.3) with different values of the constant  $\psi_2$ .<sup>31</sup>

*Aggregating across pairs of locations in the first-step moment inequalities.* The moment inequality in equation (17) holds for any two locations  $l$  and  $l'$ . As we consider 50 possible locations in our empirical application, there is a very large number of potential location pairs. Instead of using a correspondingly large number of different moment inequalities of the type in equation (17), we use a smaller number of inequalities that aggregate across location pairs. Our choice of which location pairs to combine is guided by Corollary 3. This corollary indicates that a requisite for the inequality in equation (17) to point identify the wage parameter  $\theta_\alpha$  is that the locations  $l$  and  $l'$  being compared offer the same amenity level in the population of reference; that is, using the notation in our empirical application,  $\kappa_{nt}^l - \kappa_{nt}^{l'} = 0$ . Enforcing this condition is infeasible as these amenity levels are continuous variables that are only estimated later in our estimation procedure. However, as  $\kappa_{nt}^l$  accounts for migration costs in our setting, it is reasonable to expect it will vary with the distance between locations  $n$  and  $l$ . Thus, we hypothesize that locations  $l$  and  $l'$  that are at a similar distance to an origin  $n$  are more likely to have similar amenity levels from the perspective of workers located in  $n$  and, when combined in the context of the inequality in equation (17), should yield smaller identified sets. Consequently, we form the sample analogue of the moment in equation (17) aggregating only across location pairs  $l$  and  $l'$  for which the

---

<sup>31</sup>The formulas for the approximation points in equations (E.2) and (E.3) are similar to those used to compute the simulation results discussed in Appendix C.6. However, the formulas in equations (E.2) and (E.3) differ from those in equations (C.10) and (C.11) in that they do not depend on the structural parameters to estimate, but on constants chosen by the researcher. The fact that the researcher can explore the identifying power of a large set of possible values for the constants  $(\psi_0, \psi_1, \psi_2)$  has the advantage of generating tighter CIs, and the computational disadvantage that the number of moment inequalities used in estimation increases in the set of values of  $(\psi_0, \psi_1, \psi_2)$  the researcher uses.

difference between the distance from  $n$  to  $l$  and the distance from  $n$  to  $l'$  is in the lower tercile of all pairwise differences in distance to  $n$ . Given such location pairs, we form the moment function in equation (17) by further aggregating across all sector pairs  $s$  and  $r$ , and across worker pairs within those sectors.

*Instrument vectors.* Given a wage predictor  $z_{st}^l$  for every region  $l$ , sector  $s$ , and year  $t$  in the sample built as described in Section 5.2, we construct moment inequalities using instruments of the type in equation (23) for every combination of  $d_k \in \{-1, 1\}$  and  $[z_k, \bar{z}_k] \in \{[-\infty, Q_{25}(\Delta z_{st}^{ll'}), [Q_{25}(\Delta z_{st}^{ll'}), Q_{50}(\Delta z_{st}^{ll'}), [Q_{50}(\Delta z_{st}^{ll'}), Q_{75}(\Delta z_{st}^{ll'}), [Q_{75}(\Delta z_{st}^{ll'}), \infty]\}$ , where  $Q_q(\Delta z_{st}^{ll'})$  denotes the percentile  $q$  of the distribution of  $\Delta z_{st}^{ll'}$ . Thus, we use 8 different instrument vectors.

### E.3 PPML-IV Estimator

We describe here our implementation of the estimator in Artuç and McLaren (2015). To rationalize the implementation of this estimator in our setting, we must assume all workers employed in the same sector  $s$  have the same information set in any given period  $t$ , regardless of their location of residence. Thus,  $\mathcal{J}_{ist} = \mathcal{J}_{i'st}$  for any sector  $s$ , period  $t$ , and any two workers  $i$  and  $i'$  employed in  $s$  at  $t$ . Given this assumption, we can write the model-implied number of sector  $s$  workers that migrate between locations  $n$  and  $l$  at  $t$  as:

$$\begin{aligned} M_{nst}^l &= \frac{\exp(\alpha \mathbb{E}[w_{st}^l | \mathcal{J}_{st}] - \kappa_{nt}^l)}{\sum_k \exp(\alpha \mathbb{E}[w_{st}^k | \mathcal{J}_{st}] - \kappa_{nt}^k)} L_{nst-1} \\ &= \exp(\alpha \mathbb{E}[w_{st}^l | \mathcal{J}_{st}] - \kappa_{nt}^l + \Gamma_{nst}) \\ &= \exp(\alpha(w_{st}^l - \nu_{st}^l) - \kappa_{nt}^l + \Gamma_{nst}) \\ &= \exp(\Lambda_{st}^l + \Psi_{nt}^l + \Gamma_{nst}), \end{aligned} \tag{E.4}$$

where  $L_{nst-1}$  is the total number of workers in location  $n$  and sector  $s$  at period  $t-1$ ;  $\nu_{st}^l \equiv w_{st}^l - \mathbb{E}[w_{st}^l | \mathcal{J}_{st}]$  is these workers' expectational error when predicting wages in sector  $s$ , period  $t$ , and location  $l$ ; and

$$\Lambda_{st}^l \equiv \alpha(w_{st}^l - \nu_{st}^l), \quad \Gamma_{nst} \equiv -\ln\left(\sum_k \exp(\alpha \mathbb{E}[w_{st}^k | \mathcal{J}_{st}] - \kappa_{nt}^k)\right) + \ln L_{nst}, \quad \Psi_{nt}^l \equiv -\kappa_{nt}^l.$$

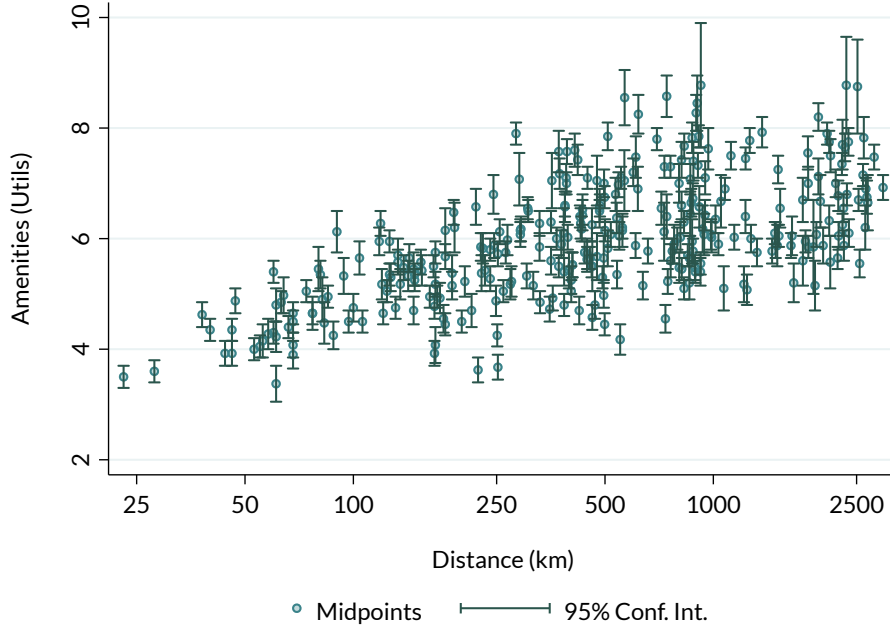
Using information on  $\{M_{nst}^l\}_{n=1, l=1}^{L, L}$  for a period  $t$ , sector  $s$ , and  $L$  origin and destination locations, the procedure in Artuç and McLaren (2015) recovers estimates of  $\alpha$  and  $\{\kappa_{nt}^l\}_{n=1, l=1}^{L, L}$  in three steps. First, it computes PPML estimates of  $\{\Lambda_{st}^l\}_{l=1}^L$ ,  $\{\Gamma_{nst}\}_{n=1}^L$  and  $\{\Psi_{nt}^l\}_{n=1, l=1}^{L, L}$  using the expression in the last line in equation (E.4). Second, under the assumption that a variable  $z_{st}^l$  correlated with  $w_{st}^l$  belongs to the information set  $\mathcal{J}_{st}$ , it computes an IV estimate

of  $\alpha$  by regressing  $\hat{\Lambda}_{st}^l$  on  $w_{st}^l$  using  $z_{st}^l$  as an instrument, with  $\hat{\Lambda}_{st}^l$  the first-step estimate of  $\Lambda_{st}^l$ . Third, it recovers  $\hat{\kappa}_{nt}^l = -\hat{\Psi}_{nt}^l$  for every origin  $n$  and destination  $l$ .

## E.4 Additional Result

Figure E.1 provide the CIs of estimated amenities  $\kappa_{nt}^l$ , described in Section 5.2.2.

Figure E.1: Amenities from Moment Inequalities with Confidence Intervals



Each point shows the midpoint of the 95% confidence interval for a given bilateral migration cost  $\kappa_{nt}^l$  in the year 2011, with the associated 95% confidence interval.

## F Model with Endogenous Worker Types - Sectors

We model workers' choice of market to supply labor. Each labor market is defined by a sector  $s = 1, \dots, S$  and a location  $l = 1, \dots, L$ , and we index each market by the combination of indices  $sl$ . Defining a variable  $y_i^{sl}$  that equals one if worker  $i$  chooses market  $sl$  (and zero otherwise), we assume

$$y_i^{sl} \equiv \mathbb{1}\{l = \underset{\substack{l'=1, \dots, L \\ s'=1, \dots, S}}{\operatorname{argmax}} \mathbb{E}[\mathcal{U}_i^{s'l'} | \mathcal{J}_i]\} \quad \text{for any } l = 1, \dots, L \text{ and } s = 1, \dots, S.$$

Consistently with equation (2), we assume worker expectations are rational. However, instead of equation (3), we assume the utility of choosing market  $sl$  for a worker  $i$  is:

$$\mathcal{U}_i^{sl} = \kappa^l + \tau^s + \alpha w_i^{sl} + \varepsilon_i^{sl},$$

where the new term  $\tau^s$  is a sector-specific unobserved term that accounts for sector-specific amenities as well as for sector-specific switching costs.

The assumption in equation (4) applies directly to the model with endogenous worker types. The assumption in equation (5) extends naturally to the model considered here. Specifically, for any sectors  $s$  and  $r$ , locations  $l$  and  $l'$ , and worker indices  $i$  and  $j$ , it holds:

$$\mathbb{E}[\Delta w_i^{sl'} | \mathcal{J}_i, \mathcal{J}_j] = \mathbb{E}[\Delta w_i^{sl'} | \mathcal{J}_i] = \mathbb{E}[\Delta w_i^{sl'} | \mathcal{W}_i] = \mathbb{E}[\Delta w^{sl'} | \mathcal{W}_i],$$

where  $\Delta w_i^{sl'} = w_i^{sl} - w_i^{sl'}$  and  $\Delta w^{sl'} = w^{sl} - w^{sl'}$ , with  $w^{sl}$  a market-level wage shifter.

Finally, instead of equation (6), we assume that, for any workers  $i$  and  $j$ , it holds that

$$\begin{aligned} F_\varepsilon(\varepsilon_i, \varepsilon_j | \mathcal{W}_i, \mathcal{W}_j) &= F_\varepsilon(\varepsilon_i) F_\varepsilon(\varepsilon_j) \\ &= \exp \left( - \sum_{s=1}^S \left( \sum_{l=1}^L \exp(-\varepsilon_i^{sl}) \right)^\psi - \sum_{s=1}^S \left( \sum_{l=1}^L \exp(-\varepsilon_j^{sl}) \right)^\psi \right), \end{aligned}$$

where  $\psi$  measures the extent to which the type I extreme value idiosyncratic shocks are correlated across locations within a sector. Thus, the model with endogenous sectors is a nested logit model, with each nest defined by a sector  $s = 1, \dots, S$ .

## G Extension: Dynamic Model of Location Choice

We describe here how to extend our estimation method to settings with forward-looking agents facing one-time migration costs. In Appendix G.1, we describe our dynamic migration model. In Appendix G.2, we show how to adapt the procedure in Section 3 to the estimation of the parameters of the dynamic model. Appendix G.3 provides additional details.

### G.1 Theoretical Framework

Defining a variable  $y_{ist}^l$  that equals one if worker  $i$  of type  $s$  chooses location  $l$  at period  $t$ , and zero otherwise, we assume

$$y_{ist}^l \equiv \mathbb{1} \{ l = \underset{l'=1, \dots, L}{\operatorname{argmax}} \mathbb{E}[\mathcal{V}_{ist}^{l'} | \mathcal{J}_{ist}] \} \quad \text{for } l = 1, \dots, L, \quad (\text{G.1})$$

with  $\mathcal{V}_{ist}^l$  the choice-specific value function and  $\mathbb{E}[\cdot]$  defined as in equation (2). We impose:

$$\mathcal{V}_{ist}^l = v_{ist}^l + \varepsilon_{ist}^l, \quad (\text{G.2a})$$

$$v_{ist}^l = \beta x_{nt}^l + \lambda_t^l + \alpha w_{ist}^l + \delta \mathcal{V}_{ist+1}^{(lt)}, \quad (\text{G.2b})$$

where  $n$  indexes the location of worker  $i$  of type  $s$  at period  $t - 1$ , and

$$\mathcal{V}_{ist+1}^{(lt)} \equiv \max_{l'=1, \dots, L} \mathbb{E}[\mathcal{V}_{ist+1}^{(lt)l'} | \mathcal{J}_{ist+1}^{(lt)}]. \quad (\text{G.3})$$

Equation (G.2a) splits the choice-specific value function into the idiosyncratic component  $\varepsilon_{ist}^l$  and a variable  $v_{ist}^l$  that equation (G.2b) defines as the sum of four terms. First, the migration costs between locations  $n$  and  $l$ , modeled as a function of observed covariates  $x_{nt}^l$  and a vector of parameters  $\beta$ . Second, a location- and period-specific term  $\lambda_t^l$ , which captures a location's amenities and (log) price index. Third, the wage component  $\alpha w_{ist}^l$ . Fourth, the product of the discount factor  $\delta$  and a variable  $\mathcal{V}_{ist+1}^{(lt)}$  that, according to equation (G.3), equals the worker's period- $t + 1$  value function conditional on choosing alternative  $l$  at period  $t$ .<sup>32</sup>

Defining  $\lambda_t = (\lambda_t^1, \dots, \lambda_t^L)$  and  $\varepsilon_{ist} = (\varepsilon_{ist}^1, \dots, \varepsilon_{ist}^L)$ , we assume that

$$(\varepsilon_{ist}, \lambda_t) \subseteq \mathcal{J}_{ist}. \quad (\text{G.4})$$

Thus, when making choices at  $t$ , workers know the vectors of contemporaneous idiosyncratic preferences  $\varepsilon_{ist}$  and amenities  $\lambda_t$ . Equation (G.4) does not restrict the information workers have about wages  $w_{ist'} = (w_{ist'}^1, \dots, w_{ist'}^L)$  for any  $t' \geq t$  or amenities  $\lambda_{t'}$  for any  $t' > t$ .

While we do not specify the full content of workers' information sets, we limit the processes that determine them and assume that, for any  $t' > t$ ,

$$\mathcal{J}_{ist'} \perp\!\!\!\perp y_{ist} | \mathcal{J}_{ist}. \quad (\text{G.5})$$

Thus, conditional on the worker's information set at a period  $t$ , the worker's information set in subsequent periods does not depend on the worker's choice at  $t$ . Our framework thus does not allow for endogenous learning, understood as the process through which the worker's information set at  $t$  may depend on the history of locations visited by the worker.

Defining  $\Delta v_{ist}^{ll'} \equiv v_{ist}^l - v_{ist}^{l'}$ , we impose that for any period  $t$ , locations  $l$  and  $l'$ , types  $s$

---

<sup>32</sup>A comparison of equations (3) and (G.2) shows that, at the expense of assuming  $\delta = 0$ , the static model allows for a more flexible specification of migration costs, which may vary freely across locations and periods.

and  $r$ , and workers  $i$  and  $j$  that share a common prior location  $n$ ,

$$\mathbb{E}[\Delta v_{ist}^{ll'} | \mathcal{J}_{ist}, \mathcal{J}_{jrt}] = \mathbb{E}[\Delta v_{ist}^{ll'} | \mathcal{J}_{ist}] = \mathbb{E}[\Delta v_{ist}^{ll'} | \mathcal{W}_{ist}]. \quad (\text{G.6})$$

The first equality imposes that every worker has at least as much information as any other worker of a different type  $r$  with whom it shares prior location  $n$  about differences in their own location-specific value functions. The second equality imposes that, once we condition on all other elements of the information set of worker  $i$  of type  $s$  at period  $t$ , the idiosyncratic preferences in  $\varepsilon_{ist}$  do not contain any information on  $\Delta v_{ist}^{ll'}$  for any two locations  $l$  and  $l'$ .<sup>33</sup>

As in the static model, we restrict the information workers have on location-specific wages. For workers  $i$  and  $j$  of types  $s$  and  $r$ , respectively, and locations  $l$  and  $l'$ , we impose that

$$\mathbb{E}[\Delta w_{ist}^{ll'} | \mathcal{W}_{ist}, \mathcal{W}_{jrt}] = \mathbb{E}[\Delta w_{ist}^{ll'} | \mathcal{W}_{ist}]. \quad (\text{G.7})$$

Thus, the worker's period- $t$  expectation of the contemporaneous wage difference between two locations  $l$  and  $l'$  equals the expectation of terms that do not vary across workers of the same type  $s$ . We do not restrict the information workers have about the difference in type- and location-specific wages  $\Delta w_{st'}^{ll'}$  between any two locations  $l$  and  $l'$  and in any period  $t'$ .

Finally, as in equation (6), we assume that, for workers  $i$  and  $j$  of types  $s$  and  $r$ ,

$$F_\varepsilon(\varepsilon_{ist}, \varepsilon_{jrt} | \mathcal{W}_{ist}, \mathcal{W}_{jrt}) = F_\varepsilon(\varepsilon_{ist})F_\varepsilon(\varepsilon_{jrt}) = \exp\left(-\sum_{l=1}^L (\exp(-\varepsilon_{ist}^l) + \exp(-\varepsilon_{jrt}^l))\right). \quad (\text{G.8})$$

That is, the vectors  $\varepsilon_{ist}$  and  $\varepsilon_{jrt}$  are independent of  $(\mathcal{W}_{ist}, \mathcal{W}_{jrt})$  and of each other, and each of their elements is *iid* according to a type I extreme value distribution.

The elements of  $\lambda_t$  are identified up to a common shifter. We normalize  $\lambda_t^1 = 0$  for all  $t$ ; the model parameters are thus  $(\lambda_t^2, \dots, \lambda_t^L)$  for all  $t$ ,  $\alpha$ , and  $\beta$ .

## G.2 Estimation With Moment Inequalities

We provide a two-step estimator. In the first step, we compute a confidence set for  $(\alpha, \beta)$  using inequalities that difference out the amenity term  $\lambda_t^l$  for all  $l$  and  $t$ . In the second step, for each  $l = 2, \dots, L$  and sample period  $t$ , we derive inequalities that depend only on  $\alpha$ ,  $\beta$ , and  $\lambda_t^l$ , and combine these inequalities with the confidence set for  $(\alpha, \beta)$  to compute a confidence interval for  $\lambda_t^l$ . We denote by  $(\theta_\alpha, \theta_\beta)$  the parameter vector with true value  $(\alpha, \beta)$ , and by  $\Theta_{(\alpha, \beta)}$  the set of possible values of  $(\theta_\alpha, \theta_\beta)$ . We denote by  $\theta_t^l$  the parameter with

---

<sup>33</sup>The variable  $\Delta v_{ist}^{ll'}$  depends on the worker's future choices, which will depend on  $\varepsilon_{ist'}$  for  $t' > t$ ; thus, equation (G.6) will generally not hold unless  $\varepsilon_{ist}$  is independent over time.

true value  $\lambda_t^l$  and by  $\Theta_t^l$  the set of possible values of  $\theta_t^l$ . In Appendix G.2.1, we discuss the estimation of  $\{\theta_t^l\}_{l,t}$ . In Appendix G.2.2, we describe the estimation of  $(\theta_\alpha, \theta_\beta)$ .

### G.2.1 Second-Step: Estimating Location-Specific Amenities

Denote by  $\Delta\theta_t^{ll'} \equiv \theta_t^l - \theta_t^{l'}$  the parameter with true value  $\Delta\lambda_t^{ll'} \equiv \lambda_t^l - \lambda_t^{l'}$ , and by  $\Theta_t^{ll'}$  the set of possible values of  $\Delta\theta_t^{ll'}$ . Then, for any pair of locations  $l$  and  $l'$ ,  $z_{st}$ , and scalar random variable  $e_{ist}^{ll'}$ , we define the moment

$$\tilde{m}^{ll'}(z_{st}, \Delta\theta_t^{ll'}) \equiv \mathbb{E}[y_{ist}^{l'} - y_{ist}^l \exp(-e_{ist}^{ll'}) (1 + e_{ist}^{ll'} - \Delta\tilde{v}_{ist}^{ll'}(\Delta\theta_t^{ll'})) | z_{st}], \quad (\text{G.9})$$

with

$$\Delta\tilde{v}_{ist}^{ll'}(\Delta\theta_t^{ll'}) = \beta\Delta x_{nt}^{ll'} + \Delta\theta_t^{ll'} + \alpha\Delta w_{ist}^{ll'} + \delta\beta \sum_{l''=1}^L y_{ist+1}^{(lt)l''} (x_{lt+1}^{l''} - x_{l't+1}^{l''}). \quad (\text{G.10})$$

Theorem 4 establishes a property of this moment when evaluated at  $\Delta\theta_t^{ll'} = \Delta\lambda_t^{ll'}$ .

**Theorem 4** *Assume equations (G.1) to (G.8) hold. Then,  $\tilde{m}^{ll'}(z_s, \Delta\lambda_t^{ll'}) \geq 0$  if  $e_{ist}^{ll'} \subseteq \mathcal{J}_{ist}$  and  $z_{st} \subseteq \mathcal{W}_{ist}$ .*

We prove this theorem in Appendix G.3.1. Theorem 4 shows that, given knowledge of  $(\alpha, \beta)$ , one may bound the amenity difference  $\Delta\lambda_t^{ll'}$  for any sample period  $t$  and locations  $l$  and  $l'$ , and provide an expression for the optimal scalar  $e_{ist}^{ll'}$  in Appendix G.3.2

To obtain the inequality  $\tilde{m}^{ll'}(z_{st}, \Delta\lambda_t^{ll'}) \geq 0$ , we first follow steps analogous to those taken to derive the static bounding inequality in equation (8).<sup>34</sup> In this way, we obtain

$$\mathbb{E}[y_{ist}^{l'} - y_{ist}^l \exp(-e_{ist}^{ll'}) (1 + e_{ist}^{ll'} - (v_{ist}^l - v_{ist}^{l'})) | z_{st}] \geq 0. \quad (\text{G.11})$$

This inequality cannot be used for estimation as the value function difference  $v_{ist}^l - v_{ist}^{l'}$  is not a function only of observed covariates and parameters. We follow Morales et al. (2019) and implement a discrete analogue of Euler's perturbation method to derive an inequality that can be used for estimation. Specifically, we substitute  $v_{ist}^{l'}$  in equation (G.11) by a function  $\tilde{v}_{ist}^{l'}$ , where  $v_{ist}^{l'}$  and  $\tilde{v}_{ist}^{l'}$  differ in that the latter conditions on the choices that, from period  $t + 1$  onwards, would be optimal for worker  $i$  of type  $s$  if they had chosen alternative  $l$  at  $t$ . As our dynamic model exhibits one-period dependence,  $v_{ist}^l - \tilde{v}_{ist}^{l'}$  is a function exclusively of

<sup>34</sup>While it may be feasible to use the odds-based moment inequalities introduced in Section 3.1.2 in the context of our dynamic model, we have not found the way of doing so.

the difference in static utilities at period  $t$  and the discounted difference in static utilities at  $t + 1$  that are due to whether the worker chooses alternatives  $l$  or  $l'$  at  $t$ . Specifically,

$$v_{ist}^l - \tilde{v}_{ist}^{l'} = u_{ist}^l - u_{ist}^{l'} + \delta\beta \sum_{l''=1}^L y_{ist+1}^{(lt)l''} (x_{lt+1}^{l''} - x_{l't+1}^{l''}), \quad (\text{G.12})$$

where  $y_{ist+1}^{(lt)l''}$  is the optimal choice at  $t+1$  of worker  $i$  of type  $s$  if they were to choose alternative  $l$  at  $t$ . The expression in equation (G.12) is a function of observed covariates and parameters. Moreover,  $v_{ist}^{l'} \geq \tilde{v}_{ist}^{l'}$  for every worker, period, and choices  $l$  and  $l'$ . Thus, the sign of the moment inequality in equation (G.11) is preserved if  $\tilde{v}_{ist}^{l'}$  takes the place of  $v_{ist}^{l'}$ .

### G.2.2 First-Step: Estimating Migration Costs and Wage Coefficient

For any period  $t$ , locations  $l$  and  $l'$ , worker  $i$  of type  $s$  and worker  $j$  of type  $r$ , vectors  $z_{st}$  and  $z_{rt}$ , and scalar random variable  $e_{ijsrt}^{ll'}$ , we define the moment

$$\begin{aligned} \tilde{\mathbb{M}}^{ll'}(z_{st}, z_{rt}, \theta_\alpha, \theta_\beta) \equiv \\ \mathbb{E}[y_{ist}^l y_{jrt}^l + y_{ist}^{l'} y_{jrt}^{l'} - y_{ist}^l y_{jrt}^{l'} \exp(-e_{ijsrt}^{ll'}) (2 + 2e_{ijsrt}^{ll'} - (\Delta \tilde{v}_{ist}^{ll'} + \Delta \tilde{v}_{jrt}^{ll'})) | z_s, z_r]. \end{aligned} \quad (\text{G.13})$$

Theorem 5 establishes a property of this moment when evaluated at  $\theta_\alpha = \alpha$  and  $\theta_\beta = \beta$ .

**Theorem 5** *Assume equations (G.1) to (G.8) hold. Then,  $\mathbb{M}^{ll'}(z_s, z_r, \alpha, \beta) \geq 0$  if  $e_{ijsrt}^{ll'} \subseteq \mathcal{J}_{ist} \cup \mathcal{J}_{jrt}$ ,  $z_{st} \subseteq \mathcal{W}_{ist}$ , and  $z_{rt} \subseteq \mathcal{W}_{jrt}$ .*

We prove Theorem 5 in Appendix G.3.3. Theorem 5 states that, given equations (G.1) to (G.8), the assumption that  $z_{st}$  belongs to the information set of worker  $i$  of type  $s$  at period  $t$ , and the assumption that  $z_{rt}$  belongs to the information set of worker  $j$  of type  $r$  at period  $t$ , the moment in equation (G.13) is positive when evaluated at  $(\theta_\alpha, \theta_\beta) = (\alpha, \beta)$ . Furthermore, this is true regardless of the period  $t$ , the two locations  $l$  and  $l'$  we compare, the workers  $is$  and  $jr$  we consider, the vectors  $z_{st}$  and  $z_{rt}$  on which we condition, and scalar random variables  $e_{ijsrt}^{ll'}$  we use to form the moment. We thus may compute the set of values of  $(\theta_\alpha, \theta_\beta)$  that satisfy

$$\tilde{\mathbb{M}}^{ll'}(z_{st}, z_{rt}, \theta_\alpha, \theta_\beta) \geq 0, \quad (\text{G.14})$$

and, if equations (G.1) to (G.8) hold,  $z_{st} \subseteq \mathcal{J}_{ist}$ ,  $z_{rt} \subseteq \mathcal{J}_{jrt}$ , and  $e_{ijsrt}^{ll'} \subseteq \mathcal{J}_{ist} \cup \mathcal{J}_{jrt}$ ,  $(\alpha, \beta)$  will belong to this set.

### G.3 Proofs and Additional Details

#### G.3.1 Second-Step Bounding Inequalities: Proof of Theorem 4

Equation (G.1) implies that, for any worker  $i$  of type  $s$ , period  $t$ , and locations  $l$  and  $l'$ ,

$$(y_{ist}^l + y_{ist}^{l'}) (\mathbb{1}\{\mathbb{E}[\mathcal{V}_{ist}^l - \mathcal{V}_{ist}^{l'} | \mathcal{J}_{ist}] \geq 0\} - y_{ist}^l) = 0.$$

Equations (G.2a), (G.4), and (G.6) imply we can rewrite this equality as

$$(y_{ist}^l + y_{ist}^{l'}) (\mathbb{1}\{\mathbb{E}[\Delta v_{ist}^{ll'} | \mathcal{W}_{ist}] + \Delta \varepsilon_{ist}^{ll'} \geq 0\} - y_{ist}^l) = 0, \quad (\text{G.15})$$

where  $\Delta v_{ist}^{ll'} = v_{ist}^l - v_{ist}^{l'}$  and  $\Delta \varepsilon_{ist}^{ll'} = \varepsilon_{ist}^l - \varepsilon_{ist}^{l'}$ . This equality holds for any worker  $i$  of any type  $s$ , any period  $t$ , and any two locations  $l$  and  $l'$ . As the next step, we take the expectation of both sides of this equality conditional on a value of  $\mathcal{W}_{ist}$  and a dummy variable that equals one if worker  $i$  of type  $s$  chooses either location  $l$  or location  $l'$  at period  $t$ ; i.e.,

$$\mathbb{E}[\mathbb{1}\{\mathbb{E}[\Delta v_{ist}^{ll'} | \mathcal{W}_{ist}] + \Delta \varepsilon_{ist}^{ll'} \geq 0\} - y_{ist}^l | \mathcal{W}_{ist}, y_{ist}^l + y_{ist}^{l'} = 1] = 0.$$

Equation (G.8) implies we can rewrite this moment equality as

$$\mathbb{E}\left[\frac{\exp(\mathbb{E}[\Delta v_{ist}^{ll'} | \mathcal{W}_{ist}])}{1 + \exp(\mathbb{E}[\Delta v_{ist}^{ll'} | \mathcal{W}_{ist}])} - y_{ist}^l \middle| \mathcal{W}_{ist}, y_{ist}^l + y_{ist}^{l'} = 1\right] = 0,$$

or, equivalently, after multiplying by  $1 + \exp(-\mathbb{E}[\Delta v_{ist}^{ll'} | \mathcal{W}_{ist}])$  and rearranging, we obtain

$$\mathbb{E}[1 - y_{ist}^l - y_{ist}^l \exp(-\mathbb{E}[\Delta v_{ist}^{ll'} | \mathcal{W}_{ist}]) | \mathcal{W}_{ist}, y_{ist}^l + y_{ist}^{l'} = 1] = 0.$$

Given the conditioning on the event  $y_{ist}^l + y_{ist}^{l'} = 1$ , we can further rewrite

$$\mathbb{E}[y_{ist}^{l'} + y_{ist}^l (-\exp(-\mathbb{E}[\Delta v_{ist}^{ll'} | \mathcal{W}_{ist}])) | \mathcal{W}_{ist}] = 0. \quad (\text{G.16})$$

As  $-\exp(-x)$  is concave in  $x \in \mathbb{R}$ , a linear approximation to it at any value  $a \in \mathbb{R}$  will bound it from above. Thus, given any scalar random variable  $e_{ist}^{ll'}$ , we obtain

$$\mathbb{E}[y_{ist}^{l'} + y_{ist}^l \exp(-e_{ist}^{ll'}) (- (1 + e_{ist}^{ll'}) + \mathbb{E}[\Delta v_{ist}^{ll'} | \mathcal{W}_{ist}]) | \mathcal{W}_{ist}] \geq 0. \quad (\text{G.17})$$

Defining as  $\nu_{ist}^{ll'}$  the expectational error worker  $i$  of type  $s$  makes when forecasting at period  $t$  the variable  $\Delta v_{ist}^{ll'}$ , the assumption that expectations are rational in equation (2) implies

$$\nu_{ist}^{ll'} \equiv \Delta v_{ist}^{ll'} - \mathbb{E}[\Delta v_{ist}^{ll'} | \mathcal{W}_{ist}], \quad \implies \quad \mathbb{E}[\nu_{ist}^{ll'} | \mathcal{W}_{ist}] = 0. \quad (\text{G.18})$$

Let's consider the alternative moment

$$\mathbb{E}[y_{ist}' + y_{ist}^l \exp(-e_{ist}^{ll'})(-(1 + e_{ist}^{ll'}) + \Delta v_{ist}^{ll'}) | \mathcal{W}_{ist}]. \quad (\text{G.19})$$

Given that  $\mathcal{W}_{ist} \subseteq \mathcal{J}_{ist}$ , we use the LIE to rewrite this moment as

$$\mathbb{E}[\mathbb{E}[y_{ist}' + y_{ist}^l \exp(-e_{ist}^{ll'})(-(1 + e_{ist}^{ll'}) + (\mathbb{E}[\Delta v_{ist}^{ll'} | \mathcal{W}_{ist}] + \nu_{ist}^{ll'})) | \mathcal{J}_{ist}] | \mathcal{W}_{ist}].$$

Equation (G.1) implies  $\mathbb{E}[y_{ist}^l | \mathcal{J}_{ist}] = y_{ist}^l$ . Consequently, if  $z_{st} \subseteq \mathcal{W}_{ist}$ , then  $z_{st} \subseteq \mathcal{J}_{ist}$  and, using  $\Delta v_{ist}^{ll'} = \mathbb{E}[\Delta v_{ist}^{ll'} | \mathcal{W}_{ist}] + \nu_{ist}^{ll'}$ , equation (G.6) implies we can write equation (G.19) as

$$\mathbb{E}[y_{ist}' + y_{ist}^l \exp(-e_{ist}^{ll'})(-(1 + e_{ist}^{ll'}) + \mathbb{E}[\mathbb{E}[\Delta v_{ist}^{ll'} | \mathcal{W}_{ist}] + \nu_{ist}^{ll'} | \mathcal{W}_{ist}]) | \mathcal{W}_{ist}],$$

and, if  $e_{ist}^{ll'} \subseteq \mathcal{J}_{ist}$ , equation (G.18) implies we can further rewrite it as

$$\mathbb{E}[y_{ist}' + y_{ist}^l \exp(-e_{ist}^{ll'})(-(1 + e_{ist}^{ll'}) + \mathbb{E}[\Delta v_{ist}^{ll'} | \mathcal{W}_{ist}]) | \mathcal{W}_{ist}].$$

This moment is the same as in equation (G.17). Hence, we can conclude that

$$\mathbb{E}[y_{ist}' + y_{ist}^l \exp(-e_{ist}^{ll'})(-(1 + e_{ist}^{ll'}) + \Delta v_{ist}^{ll'}) | \mathcal{W}_{ist}] \geq 0.$$

Finally, if  $z_{st} \subseteq \mathcal{W}_{ist}$ , we can use the LIE and conclude that

$$\mathbb{E}[y_{ist}' + y_{ist}^l \exp(-e_{ist}^{ll'})(-(1 + e_{ist}^{ll'}) + \Delta v_{ist}^{ll'}) | z_{st}] \geq 0. \quad (\text{G.20})$$

This moment is analogous to that in equation (7). The only difference between both moments is that the difference in the static utility between alternatives  $l$  and  $l'$  entering the moment in equation (7) (i.e.,  $\Delta \theta_{ll'} + \alpha \Delta w_s^{ll'}$ ) is substituted by the difference in the corresponding choice-specific value functions (i.e.,  $\Delta v_{ist}^{ll'}$ ).

The inequality in equation (G.20) is not immediately useful for the identification of the parameters of the dynamic model described in Appendix G.1. The difference in the choice-specific value functions  $\Delta v_{ist}^{ll'}$  depends on the optimal choices of the  $i$ th worker of type  $s$  in every period  $t' > t$  both conditional on choosing  $l$  at period  $t$  (which matters for the value of  $v_{ist}^l$ ) and conditional on choosing  $l'$  at period  $t$  (which matters for the value of  $v_{ist}^{l'}$ ). To

derive a moment inequality that may be used for the (partial) identification of the parameters of the model described in Appendix G.1, we follow the approach in Morales et al. (2019). Specifically, we substitute  $\Delta v_{ist}^{ll'}$  by the variable

$$\Delta \tilde{v}_{ist}^{ll'} \equiv v_{ist}^l - \tilde{v}_{ist}^{l'}, \quad (\text{G.21})$$

where  $\tilde{v}_{ist}^{l'}$  is the discounted sum of static utilities from period  $t$  onwards (that is, in every period  $t' \geq t$ ) if worker  $i$  of type  $s$  chooses location  $l'$  at period  $t$  but follows in every subsequent period  $t' > t$  the path of choices that would be optimal if they had instead chosen location  $l$  at  $t$ . To define  $v_{ist}^l$ ,  $v_{ist}^{l'}$ , and  $\tilde{v}_{ist}^{l'}$  formally, denote by

$$y_{ist'}^{(lt)} = (y_{ist'}^{(lt)1}, \dots, y_{ist'}^{(lt)L}) \quad (\text{G.22})$$

the vector of choices of worker  $i$  of type  $s$  at  $t'$  if they were to choose alternative  $l$  at  $t$ . Then, we can write

$$v_{ist}^l = \kappa_{nt}^l + \alpha w_{st}^l + \delta \sum_{l''=1}^L y_{ist+1}^{(lt)l''} (\kappa_{lt+1}^{l''} + \alpha w_{st+1}^{l''} + \varepsilon_{ist+1}^{l''}) \quad (\text{G.23a})$$

$$+ \sum_{t'>t+1} \delta^{t'-t} \sum_{n'=1}^L \sum_{l''=1}^L y_{ist'}^{(lt)n'} y_{ist'+1}^{(lt)l''} (\kappa_{n't'+1}^{l''} + \alpha w_{st'+1}^{l''} + \varepsilon_{ist'+1}^{l''}),$$

$$v_{ist}^{l'} = \kappa_{nt}^{l'} + \alpha w_{st}^{l'} + \delta \sum_{l''=1}^L y_{ist+1}^{(l't)l''} (\kappa_{lt+1}^{l''} + \alpha w_{st+1}^{l''} + \varepsilon_{ist+1}^{l''}) \quad (\text{G.23b})$$

$$+ \sum_{t'>t+1} \delta^{t'-t} \sum_{n'=1}^L \sum_{l''=1}^L y_{ist'}^{(l't)n'} y_{ist'+1}^{(l't)l''} (\kappa_{n't'+1}^{l''} + \alpha w_{st'+1}^{l''} + \varepsilon_{ist'+1}^{l''}),$$

$$\tilde{v}_{ist}^{l'} = \kappa_{nt}^{l'} + \alpha w_{st}^{l'} + \delta \sum_{l''=1}^L y_{ist+1}^{(lt)l''} (\kappa_{lt+1}^{l''} + \alpha w_{st+1}^{l''} + \varepsilon_{ist+1}^{l''}) \quad (\text{G.23c})$$

$$+ \sum_{t'>t+1} \delta^{t'-t} \sum_{n'=1}^L \sum_{l''=1}^L y_{ist'}^{(lt)n'} y_{ist'+1}^{(lt)l''} (\kappa_{n't'+1}^{l''} + \alpha w_{st'+1}^{l''} + \varepsilon_{ist'+1}^{l''}).$$

Equations (G.1) and (G.5) imply that  $\mathbb{E}[v_{ist}^{l'} | \mathcal{W}_{ist}] \geq \mathbb{E}[\tilde{v}_{ist}^{l'} | \mathcal{W}_{ist}]$ , and, consequently,

$$\mathbb{E}[\Delta v_{ist}^{ll'} | \mathcal{W}_{ist}] \geq \mathbb{E}[\Delta \tilde{v}_{ist}^{ll'} | \mathcal{W}_{ist}]. \quad (\text{G.24})$$

Equations (G.20) and (G.24) imply the following moment inequality

$$\mathbb{E}[y_{ist}^{l'} + y_{ist}^l \exp(-h_{ist}^{ll'}(z_{st}, \Delta \lambda_t^{ll'}))(-(1 + h_{ist}^{ll'}(z_{st}, \Delta \lambda_t^{ll'})) + \Delta \tilde{v}_{ist}^{ll'}) | z_{st}] \geq 0. \quad (\text{G.25})$$

Comparing the expressions for  $v_{ist}^l$  and  $\tilde{v}_{ist}^{l'}$  in equations (G.23a) and (G.23c), we can write

$$\begin{aligned}\Delta\tilde{v}_{ist}^{l'} &= v_{ist}^l - \tilde{v}_{ist}^{l'} = (\kappa_{nt}^l - \kappa_{nt}^{l'}) + \alpha(w_{st}^l - w_{st}^{l'}) + \delta \sum_{l''=1}^L y_{ist+1}^{(lt)l''} (\kappa_{lt+1}^{l''} - \kappa_{l't+1}^{l''}) \\ &= \beta(x_{nt}^l - x_{nt}^{l'}) + (\lambda_t^l - \lambda_t^{l'}) + \alpha(w_{st}^l - w_{st}^{l'}) + \delta\beta \sum_{l''=1}^L y_{ist+1}^{(lt)l''} (x_{lt+1}^{l''} - x_{l't+1}^{l''}).\end{aligned}\quad (\text{G.26})$$

Combining equations (G.26) and (G.25), we obtain an inequality whose moment equals that in equation (G.9) for  $\Delta\theta_{l'} = \Delta\lambda^{l'}$ . Equations (G.25) and (G.26) thus imply Theorem 4. ■

### G.3.2 Second-Step Bounding Inequalities: Additional Derivations

*Derivation of optimal approximation points.* Given  $z_{st} \in \mathcal{Z}_{st}$ , consider points  $e_{ist}^{l'}$  such that  $e_{ist}^{l'} = h_{ist}^{l'}(z_{st}, \Delta\theta_{l'})$  that minimizes the moment in equation (G.9) at each value of  $\Delta\theta_{l'}$ . Given  $z_{st}$  and  $\Delta\theta_{l'}$ , the first-order condition of the moment in equation (G.9) with respect to  $h_{ist}^{l'}(z_{st}, \Delta\theta_{l'})$  is

$$\mathbb{E}[y_{ist}^l (h_{ist}^{l'}(z_{st}, \Delta\theta_{l'}) - \Delta\tilde{v}_{ist}^{l'}) | z_{st}] = 0;$$

equivalently,  $\mathbb{E}[h_{ist}^{l'}(z_{st}, \Delta\theta_{l'}) - \Delta\tilde{v}_{ist}^{l'} | z_{st}, y_{ist}^l = 1] = 0$ . Solving for  $h_{ist}^{l'}(z_{st}, \Delta\theta_{l'})$ , we obtain:

$$h_{ist}^{l'}(z_{st}, \Delta\theta_{l'}) = \mathbb{E}[\Delta\tilde{v}_{ist}^{l'} | z_{st}, y_{ist}^l = 1], \quad (\text{G.27})$$

### G.3.3 First-Step Moment Inequalities: Proof of Theorem 5

For choices  $l$  and  $l'$ , worker  $i$  of type  $s$ , worker  $j$  of type  $r$ , and  $t$ , equation (G.15) implies

$$y_{jrt}^{l'} (y_{ist}^l + y_{ist}^{l'}) (\mathbb{1}\{\mathbb{E}[\Delta v_{ist}^{l'} | \mathcal{W}_{ist}] + \Delta\varepsilon_{ist}^{l'} \geq 0\} - y_{ist}^l) = 0. \quad (\text{G.28})$$

Taking the expectation of both sides of this equality conditional on  $\mathcal{W}_{ist}$ ,  $\mathcal{W}_{jrt}$ , and a dummy variable that equals one if worker  $i$  of type  $s$  chooses either  $l$  or  $l'$  at  $t$ , we obtain

$$\mathbb{E}[y_{jrt}^{l'} (y_{ist}^l + y_{ist}^{l'}) (\mathbb{1}\{\mathbb{E}[\Delta v_{ist}^{l'} | \mathcal{W}_{ist}] + \Delta\varepsilon_{ist}^{l'} \geq 0\} - y_{ist}^l) | \mathcal{W}_{ist}, \mathcal{W}_{jrt}, y_{ist}^l + y_{ist}^{l'} = 1] = 0.$$

Given equations (G.1) and (G.8), we can rewrite this moment equality as

$$\mathbb{E}\left[y_{jrt}^{l'} \left( \frac{\exp(\mathbb{E}[\Delta v_{ist}^{l'} | \mathcal{W}_{ist}])}{1 + \exp(\mathbb{E}[\Delta v_{ist}^{l'} | \mathcal{W}_{ist}])} - y_{ist}^l \right) \middle| \mathcal{W}_{ist}, \mathcal{W}_{jrt}, y_{ist}^l + y_{ist}^{l'} = 1 \right] = 0,$$

or, equivalently, after multiplying by  $1 + \exp(-\mathbb{E}[\Delta v_{ist}^{ll'} | \mathcal{W}_{ist}])$ , and rearranging,

$$\mathbb{E}[y_{jrt}'(1 - y_{ist}^l - y_{ist}^l \exp(-\mathbb{E}[\Delta v_{ist}^{ll'} | \mathcal{W}_{ist}])) | \mathcal{W}_{ist}, \mathcal{W}_{jrt}, y_{ist}^l + y_{ist}^{l'} = 1] = 0.$$

Given that this expectation conditions on the event  $y_{ist}^l + y_{ist}^{l'} = 1$ , we can further rewrite

$$\mathbb{E}[y_{ist}' y_{jrt}' + y_{ist}^l y_{jrt}' (-\exp(-\mathbb{E}[\Delta v_{ist}^{ll'} | \mathcal{W}_{ist}])) | \mathcal{W}_{ist}, \mathcal{W}_{jrt}] = 0.$$

Since  $-\exp(-x)$  is concave in  $x \in \mathbb{R}$ , we can derive the following inequality, given any scalar random variable  $e_{ijsrt}^{ll'}$ ,

$$\mathbb{E}[y_{ist}' y_{jrt}' + y_{ist}^l y_{jrt}' \exp(-e_{ijsrt}^{ll'})(-(1 + e_{ijsrt}^{ll'}) + \mathbb{E}[\Delta v_{ist}^{ll'} | \mathcal{W}_{ist}]) | \mathcal{W}_{ist}, \mathcal{W}_{jrt}] \geq 0. \quad (\text{G.29})$$

Let's consider the alternative moment

$$\mathbb{E}[y_{ist}' y_{jrt}' + y_{ist}^l y_{jrt}' \exp(-e_{ijsrt}^{ll'})(-(1 + e_{ijsrt}^{ll'}) + \Delta v_{ist}^{ll'}) | \mathcal{W}_{ist}, \mathcal{W}_{jrt}]. \quad (\text{G.30})$$

Equation (G.6) implies  $\nu_{ist}^{ll'} = \Delta w_{ist}^{ll'} - \mathbb{E}[\Delta w_{st}^{ll'} | \mathcal{W}_{ist}, \mathcal{W}_{jrt}]$  and, thus, we can conclude that

$$\mathbb{E}[\nu_{ist}^{ll'} | \mathcal{W}_{ist}, \mathcal{W}_{jrt}] = 0. \quad (\text{G.31})$$

As  $\mathcal{W}_{ist} \subseteq \mathcal{J}_{ist}$  and  $\mathcal{W}_{jrt} \subseteq \mathcal{J}_{jrt}$ , the LIE allows to write the moment in equation (G.30) as

$$\mathbb{E}[\mathbb{E}[y_{ist}' y_{jrt}' + y_{ist}^l y_{jrt}' \exp(-e_{ijsrt}^{ll'})(-(1 + e_{ijsrt}^{ll'}) + \mathbb{E}[\Delta v_{ist}^{ll'} | \mathcal{W}_{ist}] + \nu_{ist}^{ll'}) | \mathcal{J}_{ist}, \mathcal{J}_{jrt}] | \mathcal{W}_{ist}, \mathcal{W}_{jrt}].$$

Equation (G.1) implies  $\mathbb{E}[y_{ist}^l y_{jrt}' | \mathcal{J}_{ist}, \mathcal{J}_{jrt}] = y_{ist}^l y_{jrt}'$ . Consequently, if  $z_{st} \subseteq \mathcal{W}_{ist}$  and  $z_{rt} \subseteq \mathcal{W}_{jrt}$ , then  $z_{st} \subseteq \mathcal{J}_{ist}$  and  $z_r \subseteq \mathcal{J}_{jrt}$ , and we can rewrite the moment in equation (G.30) as

$$\mathbb{E}[y_{ist}' y_{jrt}' + y_{ist}^l y_{jrt}' \exp(-e_{ijsrt}^{ll'})(-(1 + e_{ijsrt}^{ll'}) + \mathbb{E}[\mathbb{E}[\Delta v_{ist}^{ll'} | \mathcal{W}_{ist}] + \nu_{ist}^{ll'} | \mathcal{W}_{ist}, \mathcal{W}_{jrt}]) | \mathcal{W}_{ist}, \mathcal{W}_{jrt}],$$

and equation (G.31) implies we can rewrite the moment in equation (G.30) as

$$\mathbb{E}[y_{ist}' y_{jrt}' + y_{ist}^l y_{jrt}' \exp(-e_{ijsrt}^{ll'})(-(1 + e_{ijsrt}^{ll'}) + \mathbb{E}[\Delta v_{ist}^{ll'} | \mathcal{W}_{ist}]) | \mathcal{W}_{ist}, \mathcal{W}_{jrt}].$$

This moment is the same in equation (G.29), hence, we have shown:

$$\mathbb{E}[y_{ist}' y_{jrt}' + y_{ist}^l y_{jrt}' \exp(-e_{ijsrt}^{ll'})(-(1 + e_{ijsrt}^{ll'}) + \Delta v_{ist}^{ll'}) | \mathcal{W}_{ist}, \mathcal{W}_{jrt}] \geq 0. \quad (\text{G.32})$$

Following steps analogous to those we follow to derive the inequality in equation (G.25) from that in equation (G.20) (see Appendix G.3.1), we derive the following inequality from that in equation (G.32), where  $\Delta\tilde{v}_{ist}^{ll'}$  is defined as in equation (G.26):

$$\mathbb{E}[y_{ist}^{l'}y_{jrt}^{l'} + y_{ist}^ly_{jrt}^{l'}\exp(-e_{ijsrt}^{ll'})(-(1 + e_{ijsrt}^{ll'}) + \Delta\tilde{v}_{ist}^{ll'})|\mathcal{W}_{ist}, \mathcal{W}_{jrt}] \geq 0. \quad (\text{G.33})$$

This inequality is one of the two we combine to obtain the inequality that we use to bound the parameters  $\theta_\alpha$  and  $\theta_\beta$ . To obtain the one, we start with the following expression

$$y_{ist}^l(y_{jrt}^l + y_{jrt}^{l'}) (\mathbb{1}\{\mathbb{E}[\Delta v_{jrt}^{l'l}|\mathcal{W}_{jrt}] + \Delta\varepsilon_{jrt}^{l'l} \geq 0\} - y_{jrt}^{l'}) = 0. \quad (\text{G.34})$$

Following the same steps we use to go from equation (G.28) to (G.33), we derive from equation (G.34) the following inequality

$$\mathbb{E}[y_{ist}^{l'}y_{jrt}^{l'} + y_{ist}^ly_{jrt}^{l'}\exp(-e_{ijsrt}^{ll'})(-(1 + e_{ijsrt}^{ll'}) + \Delta\tilde{v}_{jrt}^{l'l})|\mathcal{W}_{ist}, \mathcal{W}_{jrt}] \geq 0. \quad (\text{G.35})$$

As the moments in equations (G.32) and (G.35) share the same function  $g_{ijsrt}^{ll'}: \mathcal{Z}_{st} \times \mathcal{Z}_{rt} \times \Theta_{(\alpha,\beta)} \rightarrow \mathbb{R}$  and the same conditioning set, we add them to obtain:

$$\mathbb{E}[y_{ist}^ly_{jrt}^l + y_{ist}^{l'}y_{jrt}^{l'} - y_{ist}^ly_{jrt}^{l'}\exp(-e_{ijsrt}^{ll'})(2 + 2e_{ijsrt}^{ll'} - (\Delta\tilde{v}_{ist}^{ll'} + \Delta\tilde{v}_{jrt}^{l'l}))|\mathcal{W}_{ist}, \mathcal{W}_{jrt}] \geq 0,$$

with

$$\begin{aligned} \Delta\tilde{v}_{ist}^{ll'} &= \beta(x_{nt}^l - x_{nt}^{l'}) + (\lambda_t^l - \lambda_t^{l'}) + \alpha(w_{st}^l - w_{st}^{l'}) + \delta\beta \sum_{l''=1}^L y_{ist+1}^{(lt)l''}(x_{lt+1}^{l''} - x_{l't+1}^{l''}), \\ \Delta\tilde{v}_{jrt}^{l'l} &= \beta(x_{nt}^{l'} - x_{nt}^l) + (\lambda_t^{l'} - \lambda_t^l) + \alpha(w_{rt}^{l'} - w_{rt}^l) + \delta\beta \sum_{l''=1}^L y_{ist+1}^{(lt)l''}(x_{l't+1}^{l''} - x_{lt+1}^{l''}). \end{aligned}$$

Finally, if  $z_{st} \subseteq \mathcal{W}_{ist}$  and  $z_{rt} \subseteq \mathcal{W}_{jrt}$ , we can use the LIE and conclude that

$$\mathbb{E}[y_{ist}^ly_{jrt}^l + y_{ist}^{l'}y_{jrt}^{l'} - y_{ist}^ly_{jrt}^{l'}\exp(-e_{ijsrt}^{ll'})(2 + 2e_{ijsrt}^{ll'} - (\Delta\tilde{v}_{ist}^{ll'} + \Delta\tilde{v}_{jrt}^{l'l}))|z_{st}, z_{rt}] \geq 0, \quad (\text{G.37})$$

with

$$\begin{aligned} \Delta\tilde{v}_{ist}^{ll'} + \Delta\tilde{v}_{jrt}^{l'l} &= \alpha(w_{ist}^l - w_{ist}^{l'} + w_{jrt}^{l'} - w_{jrt}^l) \\ &\quad + \delta\beta \sum_{l''=1}^L y_{ist+1}^{(lt)l''}(x_{lt+1}^{l''} - x_{l't+1}^{l''} + x_{l't+1}^{l''} - x_{lt+1}^{l''}). \end{aligned} \quad (\text{G.38})$$

Plugging equation (G.38) into equation (G.37), we obtain a moment inequality whose moment equals that in equation (G.13) when evaluated at  $(\theta_\alpha, \theta_\beta) = (\alpha, \beta)$ . Equations (G.37) and (G.38) thus imply Theorem 5.  $\blacksquare$

### G.3.4 First-Step Moment Inequalities: Additional Derivations

*Derivation of optimal approximation points.* Consider approximation points expressed as  $e_{ijsrt}^{ll'} = g_{ijsrt}^{ll'}(z_{st}, z_{rt}, \theta_\alpha, \theta_\beta)$ . We find the value of  $g_{ijsrt}^{ll'}(z_{st}, z_{rt}, \theta_\alpha, \theta_\beta)$  that, given  $z_{st} \in \mathcal{Z}_{st}$  and  $z_{rt} \in \mathcal{Z}_{rt}$ , minimizes the moment in equation (G.37) at each value of  $(\theta_\alpha, \theta_\beta)$ . Specifically, given  $z_{st}$ ,  $z_{rt}$ ,  $\theta_\alpha$ , and  $\theta_\beta$ , the first-order condition of the moment in equation (G.37) with respect to the scalar  $g_{ijsrt}^{ll'}(z_{st}, z_{rt}, \theta_\alpha, \theta_\beta)$  is

$$\mathbb{E}[y_{ist}^l y_{jrt}^{l'} (2g_{ijsrt}^{ll'}(z_{st}, z_{rt}, \theta_\alpha, \theta_\beta) - (\Delta \tilde{v}_{ist}^{ll'} + \Delta \tilde{v}_{jrt}^{ll'})) | z_s, z_r] = 0,$$

or, equivalently,  $\mathbb{E}[2g_{ijsrt}^{ll'}(z_{st}, z_{rt}, \theta_\alpha, \theta_\beta) - (\Delta \tilde{v}_{ist}^{ll'} + \Delta \tilde{v}_{jrt}^{ll'}) | z_{st}, z_{rt}, y_{ist}^l y_{jrt}^{l'} = 1] = 0$ . Solving for  $g_{ijsrt}^{ll'}(z_{st}, z_{rt}, \theta_\alpha, \theta_\beta)$ , we obtain:

$$g_{ijsrt}^{ll'}(z_{st}, z_{rt}, \theta_\alpha, \theta_\beta) = \mathbb{E}[\Delta \tilde{v}_{ist}^{ll'} + \Delta \tilde{v}_{jrt}^{ll'} | z_{st}, y_{ist}^l = 1], \quad (\text{G.39})$$

with  $\Delta \tilde{v}_{ist}^{ll'} + \Delta \tilde{v}_{jrt}^{ll'}$  defined as in equation (G.38).

## References

- Andrews, Donald W.K. and Gustavo Soares**, “Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection,” *Econometrica*, 2010, 78, 119–157. [16, 19, 16, 19, 1, 7, 12, 14, 15]
- Artuç, Erhan and John McLaren**, “Trade Policy and Wage Inequality: A Structural Analysis with Occupational and Sectoral Mobility,” *Journal of International Economics*, 2015, 97 (2), 278–294. [4, 24, 4, 24, 22]
- Dickstein, Michael J. and Eduardo Morales**, “What do Exporters Know?,” *The Quarterly Journal of Economics*, 2018, 133 (4), 1753–1801. [6, 21, 6, 21, 7]
- , **Jihye Jeon, and Eduardo Morales**, “Patient Costs and Physicians’ Information,” *mimeo*, 2023. [36, 3, 6, 10, 12, 3, 6, 10, 12, 14]
- Dix-Carneiro, Rafael**, “Trade Liberalization and Labor Market Dynamics,” *Econometrica*, 2014, 82 (3), 825–885. [8, 17]

- **and Brian K. Kovak**, “Trade Liberalization and Regional Dynamics,” *American Economic Review*, 2017, *107* (10), 2908–2946. [17]
- **and —**, “Margins of Labor Market Adjustment to Trade,” *Journal of International Economics*, 2019, *117*, 125–142. [17]
- Felix, Mayara**, “Trade, Labor Market Concentration, and Wages,” *mimeo*, 2022. [17]
- Morales, Eduardo, Gloria Sheu, and Andrés Zahler**, “Extended Gravity,” *The Review of Economic Studies*, 2019, *86* (6), 2668–2712. [6, 27, 31]
- Szerman, Christiane**, “The Labor Market Effects of Disability Hiring Quotas,” *mimeo*, 2024. [17]