Measuring Information Frictions in Migration Decisions: A Revealed-Preference Approach*

Thomas Fujiwara[1], Eduardo Morales[2], and Charly Porcher[3]

[1]Princeton University
[2]Princeton University
[3]Georgetown University

March 2024

Abstract

We investigate the role of information frictions in migration decisions. We show that even without data on what migrants know, most migration models are partially identified under rich patterns of incomplete information. We develop new estimators for key preference parameters based on moment inequalities that allow workers' information sets, migration costs, amenities, and price levels to vary flexibly and remain unobserved by the researcher. We use our estimators to infer the structure of workers' information consistent with observed migration decisions. Using data on internal migration flows in Brazil from 2002-2011, we find that prior work underestimates the importance of expected wages in workers' migration choices and that workers face information frictions that are both substantial and heterogeneous. Workers in regions with better internet access and higher population density have more precise wage information, and workers have more accurate information about wages in nearby areas. We quantify how migration flows and welfare are affected by changes in information frictions, and how the effects of reducing migration costs are altered by information frictions.

JEL Classifications: C13, J61, R23
Keywords: migration, moment inequalities, information frictions

*We thank Anais Galdin and Mark Bamba for excellent research assistance. All errors are our own. Email: fujiwara@princeton.edu, ecmorale@princeton.edu, charly.porcher@georgetown.edu.

1 Introduction

Migration is one of the most effective ways for people to improve their economic outcomes.[1] Yet, even between labor markets in the same country, migration flows tend to be small compared to the significant spatial dispersion in earnings.[2] The migration costs needed to explain this infrequent migration are so high that they have long been understood as a combination of factors, ranging from the monetary and psychological costs of arranging the move to information costs.[3] Disentangling the role of information frictions has proven particularly difficult because little data exists on what migrants know; as a result, migration models often make consequential, but difficult to verify, assumptions about information sets. However, assuming that migrants have more information than they really do leads to serious issues both for interpretation and policy implications. Information frictions can cause migrants to make mistakes, so that their chosen locations should not necessarily be interpreted as optimal. Information frictions also affect how we expect policies like reducing physical and regulatory barriers to mobility to impact migration and welfare.

What do workers know about wages in different local markets? How does the incomplete information they use to make migration decisions affect our estimates of how they value economic opportunities elsewhere and the size of the migration costs they face? How would workers' migration decisions change if their information changed?

In this paper, we propose answers to these questions by making three contributions. First, we develop new estimators based on moment inequalities to show that even without data on the information known to workers, the models of migration used in much of the spatial quantitative literature are partially identified under rich patterns of incomplete information. Second, we establish a method to use observed migration decisions to infer the underlying structure of information consistent with those models. Third, we apply our estimators to internal migration flows in Brazil to recover the key preference parameters. We find that estimates of the migration elasticity to wages are three times larger and that migration costs are 21% lower when accounting for information frictions. Next, we conduct counterfactual exercises to illustrate how migration flows and welfare are affected by changes in information frictions, and how the prediction of counterfactual policies, such as reducing migration costs, are altered by information frictions.

We begin by developing a model of migration with the goal of characterizing workers' uncertainty about economic conditions in different labor markets and quantifying how this uncertainty affects standard estimates of workers' preferences and their location choices. Our model recognizes expected real wages, migration costs, amenities, and idiosyncratic location preferences as the main drivers of workers' migration choices. We assume workers' expectations are rational, but allow their information sets to be both unobserved by the researcher and heterogeneous in an unrestricted way. Our baseline analysis is based on a static model; we discuss how to extend it to a

[1] See Dustmann and Preston (2019), Clemens (2011), Kennan (2013) for international migration. For examples of studies of internal migration, see Bryan et al. (2014), De la Roca and Puga (2017).

[2] See Jia et al. (2023) for a review.

[3] See Sjaastad (1962), Kennan and Walker (2011).

dynamic model in the Appendix. Workers differ in observable types that impact the wages they receive in each location. Types may be exogenous (e.g., determined by demographics) or chosen by the worker simultaneously with their location (e.g., the worker's sector or occupation). Within each type, workers are heterogeneous in both their information on location-specific wages and in their location-specific idiosyncratic preferences. We assume these preferences follow a multinomial logit model, and impose no restriction on the distribution of workers' information sets.[4]

The main challenge is to avoid interpreting as information frictions any departure between our model and observed choices that would be due to an incorrect specification of workers' location-specific utility. We limit this risk by modeling migration costs, amenities, and prices as terms unobserved by the researcher that vary flexibly by origin, destination, and period. Crucially, we assume that those determinants of migration do not vary by worker type. This allows us to account for them through choice-specific fixed effects. This approach, in combination with the large set of possible labor markets to which workers may migrate, implies our model has a large number of parameters (our empirical analysis features 1,000 labor markets, leading to 1 million parameters per year for migration costs alone). Previous work by Dickstein and Morales (2018) shows how to partially identify a three-parameter incomplete information model with moment inequalities estimators, but estimating a high-dimensional parameter vector through moment inequalities is computationally infeasible when standard moment-inequality inference procedures are used, as discussed in Canay et al. (2023). Our first contribution is to show how to estimate confidence intervals on each parameter in discrete-choice models with large choice sets, choice-specific fixed effects, and unrestricted variation in agents' information sets.

Our method relies on deriving a new type of linear inequalities that we call *bounding* inequalities. While much of the literature relies on odds-based inequalities derived from Jensen's inequality, our bounding inequality is based on bounding our concave moment function from above by its tangent at any chosen point, leading to a moment inequality that is linear in all parameters. We use those bounding inequalities to estimate the parameters in two steps. First, we derive an estimator for the migration elasticity to wages. The estimator combines two bounding inequalities for different types of workers. We compare the expected utility in two potential destinations for two workers located in the same origin but of different wage types. Since this comparison is specific to that origin, period, and two destinations, it is not affected by migration costs, amenities, or price levels. Instead, it yields an inequality that depends only on the differences in expected wages between the two destinations for each type, multiplied by the migration elasticity. We prove that this moment inequality partially identifies the migration elasticity as long as it is formed conditional on variables that belong to the worker's information set, and characterize the conditions under which point-identification is achieved. Intuitively, a combination of an origin and two destinations for which workers often choose the destination with the higher expected wage points to a high migration elasticity. If, however, workers are often seen choosing the destination offering the lower expected

[4] In the model with endogenous types, we assume idiosyncratic preferences for each type and location follow a nested logit model, with each type corresponding to a nest. Thus, the cross-location distribution of idiosyncratic preferences within a type follows a multinomial logit model.

wage, high values of the elasticity are rejected. Hence, positive and negative wage differences between locations tend to identify different bounds.

Second, we derive an estimator for each choice-specific element of utility ("amenities", for short); this estimator is conditional on the migration elasticity.[5] It combines a bounding inequality and an odds-based inequality, introduced in Dickstein and Morales (2018) for binary choice settings, which we extend to our multinomial setting. Both are expressed for a chosen origin, destination, and worker type, and infer the attractiveness of the destination from the number of workers who move there, conditional on the expected wage change from the move. Intuitively, a destination to which workers often move despite expecting low wages must offer a high amenity. Conversely, a destination expected to offer high wages but to which workers rarely move must offer a low amenity. We show that valid instruments for workers' expectations must belong to their information sets and that under this condition, our estimator provides valid confidence intervals for amenity parameters. We also derive the stricter conditions under which amenities are point-identified.

We illustrate our estimators' performance relative to other typical estimators with simulations. Given a wage predictor known to workers, we build separate instruments for positive and negative predicted wage differences, which intuitively each identify either the upper or the lower bound on parameters. In settings where both agents and the researcher may have exclusive information about the realized wages, we show that our estimator always contains the true parameter values, but that the typical Maximum Likelihood Estimator (MLE) is biased. This is the case even for the Poisson Pseudo Maximum Likelihood (PPML) estimator, which can accommodate a specific but restrictive type of incomplete information (Artuç and McLaren, 2015).

In the second part of the paper, we apply our estimators to study internal migration across labor markets in Brazil. Our sample is built from linked employer-employee data from the *Relação Anual de Informações Sociais* (RAIS). To hold demographic type constant, we select 10 million individual sector-location decisions made by workers in the largest demographic group: those who have completed high school and identify as male and White. We draw one thousand observations from each of the 10 thousand origin-sector-year cells represented by the 50 largest microregions, the 20 sectors with the largest employment, and the 10 years between 2002 and 2011.

We implement our estimator by defining workers' types as their sector of employment. While our dataset contains the realized wages of each worker in their chosen labor market, it naturally doesn't contain the hypothetical wages they could have obtained in other labor markets. We build a proxy for those wage levels by estimating a wage function that depends on individual worker characteristics (age, education level, sector-specific experience, and unobserved individual ability in their sector) and a labor market-specific shifter. Given our detailed worker panel, this wage function approximates realized wages well, with a median R^2 of 0.84 across sectors. Under our assumption that all worker characteristics are portable across locations, the migration decision only requires predicting the labor market-specific shifter.[6]

[5]We refer to these choice-specific elements as the amenity parameters, even though they also account for migration costs and prices since they vary by origin, destination, and period.

[6]While our wage equation allows for worker-specific comparative advantage across sectors (as in Dix-Carneiro

Our empirical analysis provides three main conclusions. First, workers face substantial and heterogeneous information frictions. Most workers can only identify the wage quartile to which each labor market belonged the previous year. However, workers located in regions with a higher level of internet penetration or high population density have more precise information overall, and workers in general have more accurate information about wages in nearby areas. Second, the assumptions on workers' information imposed in the prior literature yield downward biased estimates of the migration elasticity. Our preferred confidence interval for the migration elasticity is centered on 1.51 and does not include the PPML estimates of 0.5. In addition, our migration cost estimates are centered around values that are 21 percent lower than the PPML estimates. Third, the information frictions we infer also affect the partial-equilibrium welfare gains from counterfactual policies that reduce migration costs.[7] We show this by simulating our model calibrated to the 1,000 largest Brazilian labor markets, making the additional assumption that wages follow an exogenous AR(1) process, which we estimate on wages in our data. Finally, the gains from improving workers' information are substantial, reaching 7 percent when the precision of workers' information improves from only discerning quartiles of the lagged wages to complete information. In contrast to those unambiguous gains, the impact of improving information on overall migration flows is not always positive.

Our paper is related to several strands of the literature. First, it contributes to work studying the observed lack of labor mobility in the face of large variations in income levels across locations. One approach uses static models of labor location built on Roy (1951), which rationalize workers' lack of perfect mobility based on their location-specific idiosyncratic preferences (Borjas, 1987; Dahl, 2002). More recent work incorporates fixed migration costs (Tombe and Zhu, 2019; Morten and Oliveira, 2024). Starting with Kennan and Walker (2011), a different literature has introduced dynamic equilibrium models of migration featuring both one-time migration costs and idiosyncratic worker preferences for locations.[8] In our baseline analysis, we rely on a static model of migration and depart from the previous literature in that we do not specify the content of the workers' information sets. Instead, we use revealed-preference arguments to infer workers' information frictions when deciding where to move. We also show how to extend our moment inequality procedure to a dynamic partial equilibrium model.[9]

Our paper also contributes to the literature studying the relevance of information frictions for migration. Prominent studies have analyzed how experimental variation in access to information

(2014)), it does not include an unobserved (to the researcher) individual effect specific to each region-sector that could be known by workers. See Kennan and Walker (2011) for specifications of wage equations that allow for such a term in the context of dynamic discrete choice models of labor mobility. Allowing for such a component of wages poses estimation challenges when, as is our case, one imposes only a minimum content requirement on agents' information sets.

[7] Examples of such policies include infrastructure improvements as in Morten and Oliveira (2024).

[8] For partial equilibrium models, see, e.g., Bishop and Murphy (2011) and Bayer et al. (2016). For general equilibrium models, see Artuç and McLaren (2015), Caliendo et al. (2019), and Caliendo et al. (2021). Instead of modeling migration costs as a one-time cost, Desmet et al. (2018) model these as a permanent flow utility cost.

[9] The only limitation of our estimator for the dynamic model relative to the static one is that computational feasibility forces us to write migration costs as a function of observable characteristics and small number of parameters.

about payoff-relevant variables affects migration decisions; see, e.g., Bryan et al. (2014), Jensen and Miller (2017), Bergman et al. (2020), Bergman et al. (2023), and Baseler (2023). Other work has illustrated the importance of information frictions in migration through natural experiments (Wilson, 2021). Still a different branch within this literature has followed a structural approach to infer information frictions in migration. Within this third branch, Kaplan and Schulhofer-Wohl (2017) introduce a model in which individuals have imperfect information about amenities in each location, and may learn about through a Bayesian learning process with specific parametric assumptions on the distribution of priors and signals. Porcher (2022) extends this approach by developing a dynamic general equilibrium model of migration in which agents are rationally inattentive and make migration decisions with incomplete information. Our contribution relative to this literature is to infer the importance of information frictions for migration choices without observing exogenous determinants of agents' information sets nor imposing any parametric restriction on the stochastic process determining these information sets.

Finally, our paper contributes to the literature studying the use of choice data to identify agents' preferences when their expectations of the choice characteristics are unobserved. Some studies deal with this identification problem by assuming the researcher observes agents' information sets without error, and can thus construct a perfect proxy of agents' expectations, with noted limitations (Manski, 1991). Others intend to directly measure agents' subjective expectations (Manski, 2004; van der Klauuw, 2012). An alternative approach estimates discrete-choice models under the assumption that agents' expectations are rational and that their information sets are identical across agents of the same observable type (Scott, 2013; Traiberman, 2019). We follow a different approach which, in the wake of Pakes (2010), Ho and Pakes (2014), and Pakes et al. (2015), also assumes agents have rational expectations but relies on moment inequalities to allow information sets to vary across all agents in unobservable ways. We introduce agents' choice-specific idiosyncratic preferences as a second source of error in the model and extend the prior work in Dickstein and Morales (2018) by adding a novel type of bounding inequalities and showing how they apply to settings with large choice sets and choice-specific fixed effects.

The paper is organized as follows. In Section 2, we develop a general framework of location choice with incomplete information. Section 3 derives the moment inequalities. Section 4 illustrates the identification strategy on simulated data. Section 5 presents the empirical application.

2 Model of Location Choice with Incomplete Information

We model the static location choice of workers that belong to a population of interest. Workers are classified into S worker types indexed by s or r, and are indexed by i or j within each type.[10] While the model in this section assumes the worker's type is exogenous, we introduce in Appendix A a model in which this type is a choice variable determined jointly with the worker's location.

[10] In our empirical application (see Section 5), we focus on populations of interest defined by the worker's demographic characteristics and prior location, and identify the worker's type with the sector of employment.

Define a variable y_{is}^{l} that equals one if worker i of type s chooses location l (and zero otherwise). We assume that

$$y_{is}^{l} \equiv \mathbb{1}\{l = \operatorname*{argmax}_{l'=1,\ldots,L} \mathbb{E}[\mathcal{U}_{is}^{l'}|\mathcal{J}_{is}]\} \qquad \text{for } l = 1, \ldots, L, \tag{1}$$

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, $\mathcal{J}_{is} \in \mathbb{R}^{d_{is}}$ is her information set, and $\mathbb{E}[\cdot|\mathcal{J}_{is}]$ is a conditional expectation operator reflecting her beliefs. The scalar $0 \leqslant d_{is} < \infty$ denotes the dimension of the information set of worker i of type s. We impose the following assumptions on the worker's expected utility of choosing any $l = 1, \ldots, L$.

First, we assume the worker's expectations are rational. That is, for any vector $\mathcal{X}_{is}$, it holds

$$\mathbb{E}[\mathcal{X}_{is}|\mathcal{J}_{is}] = \int_{x} x dF_{x,is}(x|\mathcal{J}_{is}), \tag{2}$$

where $F_{x,is}(x|\mathcal{J}_{is})$ is the distribution of $\mathcal{X}_{is}$ conditional on $\mathcal{J}_{is}$.

Second, the utility of choosing location l for the *ith* worker of type s is:

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

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

where w_{is}^{l} is the natural logarithm of the nominal wage that worker i of type s would earn if she were to choose location l, α captures the relative importance of wages in workers' utility, and κ^{l} and ε_{is}^{l} are the common and idiosyncratic components of all other determinants of a worker's location choice. For simplicity, we refer to κ^{l} as the amenity level and ε_{is}^{l} as the idiosyncratic preferences of worker i of type s. Generally, these two terms may account for other determinants of a worker's location choice, such as migration costs and (log) prices.

Third, defining the vector of idiosyncratic preferences $\varepsilon_{is} = (\varepsilon_{is}^{1}, \ldots, \varepsilon_{is}^{L})$ and the vector of amenity levels $\kappa = (\kappa^{1}, \ldots, \kappa^{L})$, we assume that, for the *ith* worker of type s, it holds that

$$(\varepsilon_{is}, \alpha, \kappa) \subseteq \mathcal{J}_{is}, \tag{4}$$

where, for random vectors $\mathcal{X}$ and $\mathcal{X}'$, we use $\mathcal{X} \subseteq \mathcal{X}'$ to denote that $\mathcal{X}$ is a subvector of $\mathcal{X}'$. Equation (4) imposes that when making her location choice, worker *ith* of type s knows her vector of idiosyncratic preferences ε_{is}, the wage sensitivity α, and the amenity levels κ. It does not restrict which other variables may also belong to the vector $\mathcal{J}_{is}$. A model extension would be to allow for learning by dropping the assumption that workers know κ or α prior to choosing their location; for migration models with learning, see Kaplan and Schulhofer-Wohl (2017) and Porcher (2022).

Fourth, for types s and r, locations l and l', and worker indices i and j, it holds that

$$\mathbb{E}[\Delta w_{is}^{ll'}|\mathcal{J}_{is}, \mathcal{J}_{jr}] = \mathbb{E}[\Delta w_{is}^{ll'}|\mathcal{J}_{is}] = \mathbb{E}[\Delta w_{is}^{ll'}|\mathcal{W}_{is}] = \mathbb{E}[\Delta w_{s}^{ll'}|\mathcal{W}_{is}], \qquad \text{if } s \neq s', \tag{5}$$

where $\Delta w_{is}^{ll'} \equiv w_{is}^{l} - w_{is}^{l'}$ and $\mathcal{W}_{is}$ is a vector that includes all elements of $\mathcal{J}_{is}$ other than the vector of idiosyncratic preferences ε_{is}. The first equality in equation (5) imposes that a worker of a type s has at least as much information as any worker of a different type s' about the differences in the wages she will earn depending on her location choice. The second equality imposes that, once we condition on all other elements of the worker's information set, the vector of idiosyncratic preferences does not provide any additional information that helps the worker forecast her wages. Finally, the third equality imposes that the worker's expected wage difference between any two locations l and l' can be rewritten as the expectation of the difference in two terms that do not vary across workers of the same type s.

The assumption that worker i of type s knows about her location-specific wages as much as any worker of a different type s' has similarities with the assumption of strict exogeneity often imposed in panel data models. In our model, the elements of κ_n play a role analogous to the unobserved individual effects in panel data models, and we deal with them through an approach analogous to first-difference estimators. The assumption imposed by the first equal sign in equation (5) naturally holds if we assume all workers have the same information; i.e., if $\mathcal{J}_{is} = \mathcal{J}_{jr}$ for any i, i', s, and s'. The assumption that information sets are common across agents of the same population of interest has indeed been recently imposed to estimate discrete choice models; e.g., Scott (2013); Traiberman (2019); Humlum (2020); Diamond et al. (2019). We relax the assumption of homogeneous information sets by allowing these to vary across workers in unobservable ways. Specifically, the first equality in equation (5) is consistent workers of a type s knowing more about their type-specific wage differences (i.e., about $\Delta w_{s}^{ll'}$ for any l and l') than any worker of a different type s'. It is also consistent with the information sets of workers of the same type s being heterogeneous in an unrestricted way; e.g., worker i of type s may know more about wages in every location than worker i' of the same type s.

The assumption that, once we condition on all other elements of the worker's information set, the wage difference between any two locations l and l', $\Delta w_{is}^{ll'}$, is mean independent of the vector of idiosyncratic preferences, ε_{is}, is an implication of the exogeneity of individual-specific shocks often assumed in migration models; see Redding and Rossi-Hansberg (2017). More specifically, migration models often assume that (a) agents have perfect information on wages and (b) wages and preference shocks are independent. These two assumptions imply the second equality in equation (5). Importantly, our model allows for any pattern of correlation between location-specific amenities and wages, as we do not restrict the correlation between κ_n and the wage vector $w_{is} = (w_{is}^{1}, \ldots, w_{is}^{L})$.

The assumption that worker-specific wage differences across any two locations l and l' are unpredictable except for a type-specific component $w_{s}^{ll'}$ is imposed by data limitations. Our estimation approach allows us to model flexibly the information workers have about payoff-relevant variables that the researcher either observes or for which she has consistent estimates. Our data, however, does not allow us to either observe or estimate, for every worker and location, a wage component that is location- and worker-specific. Thus, limitations in the data force us to impose the assumption in the last equality in equation (5). A model extension would be to allow workers to select their location based on a location- and worker-specific match effect that is constant over time, as in Kennan and Walker (2011).

Importantly, the assumption imposed by the third equality in equation (5) does not restrict the information workers have about their type-specific comparative advantage. Thus, when types are equated to sectors or occupations of employment, it is consistent with the findings in Dix-Carneiro (2014), Traiberman (2019) or Humlum (2020), who show that workers choosing their sector or occupation do so based on idiosyncratic wage components.

Fifth, we assume that, for types s and s' and worker indices i and i', it holds that

$$F_{\varepsilon}(\varepsilon_{is}|\mathcal{W}_{is}, \mathcal{W}_{jr}) = F_{\varepsilon}(\varepsilon_{is}|\mathcal{W}_{is}) = F_{\varepsilon}(\varepsilon_{is}) = \exp(-\sum_{l=1}^{L}\exp(-\varepsilon_{is}^{l})), \qquad \text{if } s \neq s', \tag{6}$$

where, in an abuse of notation, we use $F_{\varepsilon}(\cdot)$ to denote the cumulative distribution function of ε_{is}, regardless of the conditioning set. The first equality in equation (6) imposes that every worker of type s has at least as much information about her own idiosyncratic preferences as any worker of a different type s'. The second equality indicates that the worker's idiosyncratic preferences are independent of all other elements of her information set. The third equality imposes that ε_{is}^{l} is *iid* across all $l = 1, \ldots, L$, and follows a type I extreme value distribution with location parameter equal to zero and scale parameter equal to one.

Equations (1) to (6) imply that, for any two locations l and l', we can write the probability that a worker i chooses location l as

$$\mathbb{E}[y_{is}^{l}|\mathcal{W}_{is}] = \frac{\exp(\Delta\kappa^{ll'} + \alpha\mathbb{E}[\Delta x_s^{ll'}|\mathcal{W}_{is}])}{\sum_{l''=1}^{L}\exp(\Delta\kappa^{l''l'} + \alpha\mathbb{E}[\Delta x_s^{l''l'}|\mathcal{W}_{is}])}. \tag{7}$$

Equations (1) to (6) do not fully specify workers' wage expectations when making their location choice. The previous migration literature has specified these expectations in different ways. First, assuming workers have perfect information; i.e., $\mathbb{E}[w_{is}|\mathcal{W}_{is}] = w_{is}$ for every worker type s and every worker i of type s. The assumption of perfect wage information is common in spatial equilibrium models, see Redding and Rossi-Hansberg (2017). Second, assuming the precise content of workers' information sets as well as restricting the data generating process of wages; e.g., assuming that, for every type s and worker i of type s, w_{is} follows a VAR process known up to a finite parameter vector and the information set $\mathcal{J}_{is}$ only includes past values of w_{is}. This type of assumption is imposed in, e.g., Kennan and Walker (2011). Third, specifying the process through which workers' information sets are determined; e.g., specifying the structure of the signals workers receive and how these signals impact workers' beliefs, as in the learning literature mentioned above. As workers' information is likely heterogeneous in ways that are both unobserved to researchers and hard to model correctly, we opt to impose no restrictions on the data-generating process of wages nor on workers' information sets.

We consider a setting in which the researcher observes a random sample of workers of size I_s of each type s. For the *ith* sampled worker of type s, the researcher observes the location choice,

$y_{is} = (y_{is}^1, \ldots, y_{is}^L)$. Additionally, for every type s, the researcher observes the vector of wage components $w_s = (w_s^1, \ldots, w_s^L)$ and a vector of covariates $z_s = (z_s^1, \ldots, z_s^L)$, with z_s^l a vector that may be used to predict w_s^l. In practice, w_s may not be directly observed, but estimated given information on $w_{is}^l y_{is}^l$ for every location l and every sampled worker i of type s; see Appendix E.1 for details. Crucially, we do not assume the researcher observes $\mathcal{W}_i$ for any worker i.

Only pairwise differences between the elements of the vector κ_n are identified. Thus, without loss of generality, we impose the normalization $\kappa^1 = 0$. Given this normalization, the goal of estimation is to recover the value of the parameters $(\kappa^2, \ldots, \kappa^L)$ and α, and to learn about the content of workers' information sets. To acquire knowledge about workers' information, we test the null hypothesis that, for a subset of locations $\mathcal{L}$, a particular wage predictor belongs to the information set for every worker in a group of interest; for example, the researcher may test the null hypothesis that $\{z_s^l\}_{l \in \mathcal{L}} \subseteq \mathcal{W}_{is}$ for all workers of type s. To simplify notation, we use $\theta \equiv (\theta_\alpha, \theta_2, \ldots, \theta_L)$ to denote the unknown parameter vector whose true value is $\theta^* \equiv (\alpha, \kappa^2, \ldots, \kappa^L)$. We use Θ_α to denote the set of possible values of θ_α and, for each $l = 2, \ldots, L$, we use Θ_l to denote the set of possible values of θ_l. Finally, we use Θ to denote the parameter space; i.e., $\Theta = \Theta_\alpha \times \Theta_2 \times \cdots \times \Theta_L$.

3 Estimation Through Moment Inequalities

We describe here the estimation of θ^*. Relative to prior work that has estimated static discrete choice models with restrictions on agents' expectations similar to those imposed in Section 2 (e.g. Dickstein et al., 2023), the key feature of our setting is the large dimensionality of the choice set, which implies that the parameter vector to estimate is equally large dimensional.

As general inference procedures for moment inequalities are based on inverting hypothesis tests at each point in a grid covering the parameter space, computing a confidence set for a large dimensional parameter vector poses computational challenges. To circumvent these, we implement a two-step procedure. In the first step, we compute a confidence interval for α using inequalities that difference out the parameters $(\kappa^2, \ldots, \kappa^L)$. In the second step, for each $l = 2, \ldots, L$, we derive inequalities that depend exclusively on the parameters α and κ^l, and combine these inequalities with the first-step confidence interval for α to compute a confidence interval for κ^l.

The moment inequalities we use in the first step of our procedure are variations of the inequalities we use in the second step. Thus, for exposition purposes, we first describe the second-step inequalities in Section 3.1. Subsequently, we describe how we build the first-step inequalities in Section 3.2. Section 3.3 explains how we use our moment inequalities to estimate confidence sets for θ^*.

3.1 Second-Step Moment Inequalities: Estimation of Migration Costs

Given knowledge of the wage coefficient α, we use two types of inequalities to identify bounds on the amenity parameter κ^l for every $l = 2, \ldots, L$. In Section 3.1.1, we introduce a new type of moment inequalities that we denote as *bounding* inequalities. In Section 3.1.2, we describe how to apply the odds-based inequalities introduced in Dickstein et al. (2023) to our estimation problem.

3.1.1 Bounding Moment Inequalities

For any two locations l and l', we denote as $\Delta\theta_{ll'} \equiv \theta_l - \theta_{l'}$ the unknown parameter whose true value is $\Delta\kappa^{ll'} \equiv \kappa^l - \kappa^{l'}$, and as $\Theta_{ll'}$ the set of possible values of $\Delta\theta_{ll'}$. For any worker type s, we denote as $\mathcal{Z}_s$ the support of z_s. Then, for any locations l and l', worker i of type s, value $z_s \in \mathcal{Z}_s$, and deterministic function $h_{is}^{ll'}\colon \mathcal{Z}_s \times \Theta_{ll'} \to \mathbb{R}$, we define the moment

$$\begin{gathered}\mathbb{m}_{is}^{ll'}(z_s, \Delta\theta_{ll'}, h_{is}^{ll'}(\cdot)) \equiv \\ \mathbb{E}[y_{is}^{l'} - y_{is}^{l} \exp(-h_{is}^{ll'}(z_s, \Delta\theta_{ll'}))(1 + h_{is}^{ll'}(z_s, \Delta\theta_{ll'}) - (\Delta\theta_{ll'} + \alpha\Delta w_s^{ll'}))|z_s]. \end{gathered} \tag{8}$$

This moment is derived from a comparison of the expected utilities of choosing locations l and l' for a worker i of type s. The location-specific superindices ll' in the moment $\mathbb{m}_{is}^{ll'}(\cdot)$ reflect that, according to the model in Section 2, the expectation conditional on z_s entering the right-hand side of equation (8) may vary across location pairs. More specifically, $\mathbb{m}_{is}^{ll'}(\cdot)$ partly depends on the distribution of the wage difference $\Delta w_s^{ll'}$ conditional on z_s. However, the model in Section 2 does not specify the stochastic process determining wages and, consequently, the distribution of $\Delta w_s^{ll'}$ conditional on z_s may vary across location pairs l and l' for any given worker type s.

Similarly, the worker subindices is in the moment $\mathbb{m}_{is}^{ll'}(\cdot)$ reveal that, according to the model in Section 2, the expectation conditional on z_s entering the right-hand side of equation (8) may vary across workers, even within a worker type. More specifically, $\mathbb{m}_{is}^{ll'}(\cdot)$ partly depends on the distribution of the choice variables y_{is}^{l} and $y_{is}^{l'}$ conditional on z_s. As equation (1) makes clear, these choice variables are functions of the information set of worker i of type s, $\mathcal{J}_{is}$. However, the model in Section 2 does not fully specify the content of $\mathcal{J}_{is}$ for any worker is, allowing the list of variables included in it to be heterogeneous across workers in unobserved ways. This potential heterogeneity in the content of $\mathcal{J}_{is}$ implies that its distribution conditional on z_s may vary across workers and, consequently, the distribution of y_{is}^{l} and $y_{is}^{l'}$ conditional on z_s may also vary across workers.

Theorem 1 establishes a property of the moment in equation (8) when evaluated at $\Delta\theta_{ll'} = \Delta\kappa^{ll'}$. This property holds for any two locations l and l' and for every worker i of any type s.

Theorem 1 *Assume equations (1) to (6) hold and $z_s \subset \mathcal{J}_{is}$. Then, $\mathbb{m}_{is}^{ll'}(z_s, \Delta\kappa^{ll'}, h_{is}^{ll'}(\cdot)) \geqslant 0$ for any locations l and l', worker i of type s, $z_s \in \mathcal{Z}_s$, and deterministic function $h_{is}^{ll'}\colon \mathcal{Z}_s \times \Theta_{ll'} \to \mathbb{R}$.*

We prove Theorem 1 in Appendix B.1. Theorem 1 states that, given equations (1) to (6) and a vector z_s that belongs to the information set of worker i of type s, the moment in equation (8) is positive when evaluated at $\Delta\theta_{ll'} = \Delta\kappa^{ll'}$. We thus may compute the set of values of $\Delta\theta_{ll'}$ for which

$$\mathbb{m}_{is}^{ll'}(z_s, \Delta\theta_{ll'}, h_{is}^{ll'}(\cdot)) \geqslant 0, \tag{9}$$

and, if equations (1) to (6) hold and $z_s \subset \mathcal{J}_{is}$, $\Delta\kappa^{ll'}$ will belong to this set regardless of the two locations l and l' being compared, of the worker i of type s being considered, and of the value of

z_s on which the moment conditions. Given locations l and l', a worker i of type s, and a value of z_s, the set of values of $\Delta\theta_{ll'}$ other than $\Delta\kappa^{ll'}$ that satisfy the inequality in equation (9) depend on the function $h_{is}^{ll'}(\cdot)$ entering the moment $m_{is}^{ll'}(\cdot)$. As we show in Appendix B.2, for any $z_s \in \mathcal{Z}_s$, the function $h_{is}^{ll'}(z_s, \Delta\theta_{ll'})$ that minimizes the set of values of $\Delta\theta_{ll'}$ that satisfy the inequality in equation (9) is:

$$h_{is}^{ll'}(z_s, \Delta\theta_{ll'}) = \Delta\theta_{ll'} + \alpha\mathbb{E}[\Delta w_s^{ll'}|z_s, y_{is}^l = 1]. \tag{10}$$

The subindices is in this function reflect that, for any two locations l and l', the expectation of $\Delta w_s^{ll'}$ conditional on z_s and y_{is}^l may vary across workers, with this variation being due to potential cross-worker heterogeneity in the content of their information sets.

As we show in Appendix B.2, the inequality implied by equations (8) to (10) can be written as

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

By swapping the identity of locations l and l' in equations (8) to (10), we obtain the inequality

$$\frac{\mathbb{E}[y_{is}^l|z_s]}{\mathbb{E}[y_{is}^{l'}|z_s]}\exp(-\alpha\mathbb{E}[\Delta w_s^{ll'}|z_s, y_{is}^{l'} = 1]) \geqslant \exp(\Delta\theta_{ll'}). \tag{12}$$

Equations (11) and (12) identify lower and upper bounds on $\Delta\theta_{ll'}$. One may derive similar bounds for any $z_s \in \mathcal{Z}_s$ and any worker i of any type s. As these equations show, the parameter $\Delta\theta_{ll'}$ will generally be partially identified. However, the following corollary describes a case where the lower and upper bounds for $\Delta\theta_{ll'}$ in equations (11) and (12) coincide, implying that this parameter is point-identified.

Corollary 1 *Assume equations (1) to (6) hold, $z_s \subset \mathcal{J}_{is}$, and $\mathbb{E}[\Delta w_s^{ll'}|z_s] = \mathbb{E}[\Delta w_s^{ll'}|\mathcal{J}_{is}]$. Then, the bounds in equations (11) and (12) imply $\Delta\theta_{ll'} = \Delta\kappa^{ll'}$.*

We prove Corollary 1 in Appendix B.3. For locations l and l', a vector of covariates z_s, and worker i of type s, Corollary 1 strengthens the assumptions in Theorem 1 by requiring not just that $z_s \subset \mathcal{J}_{is}$ but further requiring that $\mathbb{E}[\Delta w_s^{ll'}|z_s] = \mathbb{E}[\Delta w_s^{ll'}|\mathcal{J}_{is}]$. This additional requirement implies that the vector z_s contains all the information that worker i of type s has about the wage difference between locations l and l'. When this extra assumption holds, the moment inequality defined by equations (8) to (10) only holds when the unknown parameter $\Delta\theta_{ll'}$ equals its true value $\Delta\kappa^{ll'}$.

Corollary 1 thus illustrates that there are data generating processes for which, given knowledge of α, the inequalities in equations (8) to (10) point-identify $\Delta\theta_{ll'}$ even if the researcher never imposes that workers' information sets are observed. If the extra assumption in Corollary 1 does not hold, Theorem 1 still implies that equations (11) and (12) yield an interval for $\Delta\theta_{ll'}$ that contains its true value $\Delta\kappa^{ll'}$. Furthermore, given equations (11) and (12), the length of this interval equals:

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

To gain intuition on this expression, assume $\alpha > 0$. Then, the range of values of $\Delta\theta_{ll'}$ consistent with equations (11) and (12) increases in α as well as in the importance for the worker's choice of the information the worker has about $\Delta w_s^{ll'}$ in addition to that included in z_s.

The results described in this section hold for any two locations l and l'. Given the normalization $\kappa^1 = 0$, it is the case that $\Delta\theta_{l1} = \theta_l$. Thus, when the index l' in equations (11) and (12) is set equal to 1, these equations yield the following bounds on the amenity component of location l; i.e.,

$$\frac{\mathbb{E}[y_{is}^l|z_s]}{\mathbb{E}[y_{is}^1|z_s]}\exp(-\alpha\mathbb{E}[\Delta w_s^{l1}|z_s, y_{is}^l = 1]) \leqslant \exp(\theta_l) \leqslant \frac{\mathbb{E}[y_{is}^l|z_s]}{\mathbb{E}[y_{is}^1|z_s]}\exp(-\alpha\mathbb{E}[\Delta w_s^{l1}|z_s, y_{is}^1 = 1]),$$

and θ_l is point identified if the assumptions in Corollary 1 hold for $l' = 1$.

3.1.2 Odds-based Moment Inequalities

For any locations l and l', worker i of type s, and value $z_s \in \mathcal{Z}_s$, we define the moment

$$\mathbb{m}_{is,o}^{ll'}(z_s, \Delta\theta_{ll'}) \equiv \mathbb{E}[y_{is}^l \exp(-(\Delta\theta_{ll'} + \alpha\Delta w_s^{ll'})) - y_{is}^{l'}|z_s], \tag{13}$$

where the subindex o differentiates this odds-based moment from the bounding moment in equation (13). Theorem 2 establishes a property of this moment when evaluated at $\Delta\theta_{ll'} = \Delta\kappa^{ll'}$.

Theorem 2 *Assume equations (1) to (6) hold and $z_s \subset \mathcal{J}_{is}$. Then, $\mathbb{m}_{o,is}^{ll'}(z_s, \Delta\kappa^{ll'}) \geqslant 0$ for any locations l and l', worker i of type s, and $z_s \in \mathcal{Z}_s$.*

We prove Theorem 2 in Appendix B.4. Theorem 2 states that, given equations (1) to (6) and a vector z_s that belongs to the information set of worker i of type s, the moment in equation (13) is positive when evaluated at $\Delta\theta_{ll'} = \Delta\kappa^{ll'}$. We thus may compute the set of values of $\Delta\theta_{ll'}$ for which

$$\mathbb{m}_{is,o}^{ll'}(z_s, \Delta\theta_{ll'}) \geqslant 0 \tag{14}$$

and, if equations (1) to (6) hold and $z_s \subset \mathcal{J}_{is}$, $\Delta\kappa^{ll'}$ will belong to this set regardless of the locations l and l', of the worker i of type s, and of the value of z_s. As we show in Appendix B.5, the inequality represented in equations (13) and (14) can be written as

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

By swapping the identity of locations l and l' in equation (15), we obtain the inequality

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

Equations (15) and (16) identify upper and lower bounds on $\Delta\theta_{ll'}$. As these equations show, $\Delta\theta_{ll'}$ will generally be partially identified. However, the following corollary describes a case where the lower and upper bounds for $\Delta\theta_{ll'}$ in equations (15) and (16) coincide, implying that this parameter is point-identified.

Corollary 2 *Assume equations (1) to (6) hold, $z_s \subset \mathcal{J}_{is}$, and $\Delta w_s^{ll'} = \mathbb{E}[\Delta w_s^{ll'}|\mathcal{W}_{is}]$. Then, the bounds in equations (15) and (16) imply $\Delta\theta_{ll'} = \Delta\kappa^{ll'}$.*

We prove Corollary 2 in Appendix B.6. For a pair of locations l and l', a vector of covariates z_s, and a worker i of type s, Corollary 2 strengthens the assumptions in Theorem 2 by additionally requiring that $\Delta w_s^{ll'} = \mathbb{E}[\Delta w_s^{ll'}|\mathcal{W}_{is}]$. This extra requirement implies worker i of type s has perfect information on $\Delta w_s^{ll'}$. When this extra assumption holds, the moment inequality defined by equations (13) and (14) only holds when the unknown parameter $\Delta\theta_{ll'}$ equals its true value $\Delta\kappa^{ll'}$.

The results described in this section hold for any locations l and l'. When the index l' in equations (15) and (16) is set equal to 1, these equations yield the following bounds on location l's amenity term; i.e.,

$$\frac{\mathbb{E}[y_{is}^l|z_s]}{\mathbb{E}[y_{is}^1|z_s]}(\mathbb{E}[\exp(-\alpha\Delta w_s^{1l})|z_s, y_{is}^1 = 1])^{-1} \leqslant \exp(\theta_l) \leqslant \frac{\mathbb{E}[y_{is}^l|z_s]}{\mathbb{E}[y_{is}^1|z_s]}\mathbb{E}[\exp(-\alpha\Delta w_s^{l1})|z_s, y_{is}^l = 1],$$

and, according to Corollary 2, θ_l is point identified if $\Delta w_s^{l1} \subseteq z_s$.

3.2 First-Step Moment Inequalities: Estimation of Wage Coefficient

For any locations l and l', any worker i of type s and any worker j of type r, any vectors $z_s \in \mathcal{Z}_s$ and $z_r \in \mathcal{Z}_r$, and any deterministic function $g_{ijsr}^{ll'}: \mathcal{Z}_s \times \mathcal{Z}_r \times \Theta_\alpha \to \mathbb{R}$, we define the moment

$$\begin{gathered}\mathbb{M}_{ijsr}^{ll'}(z_s, z_r, \theta_\alpha, g_{ijsr}^{ll'}(\cdot)) \equiv \\ \mathbb{E}[(y_{is}^l y_{jr}^l + y_{is}^{l'} y_{jr}^{l'} - y_{is}^l y_{jr}^{l'} \exp(-g_{ijsr}^{ll'}(z_s, z_r, \theta_\alpha))(2 + 2g_{ijsr}^{ll'}(z_s, z_r, \theta_\alpha) - \theta_\alpha(\Delta w_s^{ll'} + \Delta w_r^{l'l}))|z_s, z_r].\end{gathered} \tag{17}$$

This moment is derived from a comparison of the expected utilities of choosing locations l and l' for a worker i of type s and for a worker j of type r. For reasons analogous to those discussed in Section 3.1 when commenting on the moment in equation (8), the expectation in the right-hand side of equation (17) may vary across the pair of workers is and jr, and across the pair of locations l and l'; hence, the subindices and superindices entering the moment in equation (17). Theorem 3 establishes a key property of this moment when evaluated at $\theta_\alpha = \alpha$. This property holds for any locations l and l', any worker types s and r, any worker i of type s, and any worker j of type r.

Theorem 3 *Assume equations (1) to (6) hold, $z_s \subset \mathcal{J}_{is}$ for worker i of type s, and $z_r \subset \mathcal{J}_{jr}$ for worker j of type r. Then, $\mathbb{M}_{ijsr}^{ll'}(z_s, z_r, \theta_\alpha, g_{ijsr}^{ll'}(\cdot)) \geqslant 0$ for any locations l and l', worker i of type s, worker j of type r, $z_s \in \mathcal{Z}_s$, $z_r \in \mathcal{Z}_r$, and deterministic function $g_{ijsr}^{ll'}: \mathcal{Z}_s \times \mathcal{Z}_r \times \Theta_\alpha \to \mathbb{R}$.*

We prove Theorem 3 in Appendix B.7. Theorem 3 states that, given equations (1) to (6), the assumption that z_s belongs to the information set of worker i of type s, and the assumption that z_r

belongs to the information set of worker j of type r, the moment in equation (17) is positive when evaluated at $\theta_\alpha = \alpha$. Furthermore, this is true regardless of the two locations l and l' we compare, regardless of the workers is and jr we consider, regardless of the vectors z_s and z_r on which we condition, and regardless of the deterministic function $g^{ll'}_{ijsr}(\cdot)$ we use to form the moment. We thus may compute the set of values of θ_α that satisfy

$$\mathbb{M}^{ll'}_{ijsr}(z_s, z_r, \theta_\alpha, g^{ll'}_{ijsr}(\cdot)) \geqslant 0 \tag{18}$$

and, if equations (1) to (6) hold, $z_s \subset \mathcal{J}_{is}$, and $z_r \subset \mathcal{J}_{jr}$, α will belong to this set. The set of values of θ_α other than α that satisfy the inequality in equation (18) depend on the function $g^{ll'}_{ijsr}(\cdot)$ we use to build this inequality. By minimizing the moment in equation (17) with respect to $g^{ll'}_{ijsr}(\cdot)$, we compute the function that minimizes the set of values of θ_α that satisfy the inequality in equation (18). As we show in Appendix B.8, for any $z_s \in \mathcal{Z}_s$ and $z_r \in \mathcal{Z}_r$ the function $g^{ll'}_{ijsr}(z_s, z_r, \Delta\theta_\alpha)$ that minimizes the set of values of θ_α that satisfy the inequality in equation (9) is:

$$g^{ll'}_{ijsr}(z_s, z_r, \theta_\alpha) = \theta_\alpha \mathbb{E}[0.5(\Delta w^{ll'}_s + \Delta w^{l'l}_r)|z_s, z_r, y^l_{is}y^{l'}_{jr} = 1]. \tag{19}$$

As shown in Appendix B.8, the inequality represented in equations (17) to (19) can be written as

$$\frac{\mathbb{E}[y^l_{is}y^{l'}_{jr}|z_s, z_r]}{\mathbb{E}[0.5(y^l_{is}y^l_{jr} + y^{l'}_{is}y^{l'}_{jr})|z_s, z_r]} \leqslant \exp(\theta_\alpha \mathbb{E}[0.5(\Delta w^{ll'}_s + \Delta w^{l'l}_r)|z_s, z_r, y^l_{is}y^{l'}_{jr} = 1]). \tag{20}$$

This inequality yields lower and upper bounds on θ_α when its right-hand side is increasing or decreasing in θ_α, respectively. The wage coefficient θ_α will generally be partially identified. The following corollary describes a case where moment inequalities of the type in equation (20) point identify the parameter θ_l.

Corollary 3 *Assume equations (1) to (6) hold, $\Delta\kappa^{ll'} = 0$, $\mathbb{E}[\Delta w^{ll'}_s|z_s] = \mathbb{E}[\Delta w^{ll'}_s|\mathcal{J}_{is}] = \Delta\bar{w}$, and $\mathbb{E}[\Delta w^{l'l}_r|z_r] = \mathbb{E}[\Delta w^{l'l}_r|\mathcal{J}_{ir}] = \Delta\bar{w}$, for a constant $\Delta\bar{w} \in \mathbb{R}$. Then, the combination of two moment inequalities of the type in equation (20), one with $\Delta\bar{w} > 0$ and the other one with $\Delta\bar{w} < 0$, point identifies θ_α.*

We prove Corollary 3 in Appendix B.9. According to Corollary 3, the conditions that must be satisfied for inequalities of the type in equation (20) to point identify θ_α are much stronger than those required in Corollary 1 to point identify $\Delta\theta_{ll'}$ for locations l and l' given knowledge of α. Intuitively, the moment inequality defined in equations (17) to (19) results from adding two inequalities, one for worker i of type s that compares her expected utility of location l to that of location l', and one for worker j of type r that compares her expected utility of location l' to that of location l; see Appendix B.7 for details. The combination these two inequalities is useful because the resulting inequality does not depend on the amenity parameters κ^l and $\kappa^{l'}$, which cancel due to the fact that these are shared by workers of types s and r. However, the combined inequality is naturally weaker than each of the two inequalities that are combined in it and, as a consequence, the conditions required for it to point identify the parameter of interest are stronger.

3.3 Using Inequalities for Estimation

Given a confidence set for θ_α, we describe here how we use the moment inequalities introduced in Section 3.1 to compute a confidence set for the amenity term θ_l for a location l. The procedure we follow to use the inequalities introduced in Section 3.2 to compute a confidence set for θ_α is analogous; see Appendix B.10.

The moment in equation (8) depends on a generic instrument vector z_s. Given a predictor $\Delta x_s^{ll'}$ of the wage difference $\Delta w_s^{ll'}$ and a vector $(Q_0^q, \ldots, Q_q^q)$ of q-quantiles of the distribution of $\Delta x_s^{ll'}$ across all sectors s and location pairs l and l' (i.e., $\Pr(\Delta x_s^{ll'} \leqslant Q_k^q) = k/q$ for $k \in [0, q]$), we construct the following instrument vector:

$$z_s^{ll'} = (z_{s,1}^{ll'}, \ldots, z_{s,q}^{ll'})' \quad \text{with} \quad z_{s,k}^{ll'} \equiv \mathbb{1}\{Q_{k-1}^q < \Delta x_s^{ll'} \leqslant Q_k^q\} \quad \text{for } k \in [1, q]. \tag{21}$$

In our application, we equate $\Delta x_s^{ll'}$ to the one-year lag of $\Delta w_s^{ll'}$, and build instrument vectors according to the formula in equation (28) for $q \in \{2, 4, 8, 16\}$. For example, if $q = 2$, the instrument vector in equation (28) partitions all triplets of location pairs and sectors (l, l', s) into two subsets depending on whether the one-year lag of the wage difference $\Delta w_s^{ll'}$ is above or below the median.

The instruments introduced in equation (28) are indicator functions and, thus, weakly positive. Consequently, given a vector $z_s^{ll'}$ built according to equation (28), Theorem 1 and the LIE implies that, for any locations l and l', worker i of type s, and deterministic function $h_{is}^{ll'} \colon \mathcal{Z}_s \times \Theta_{ll'} \to \mathbb{R}$, the $q \times 1$ vector of moment inequalities

$$\mathbb{E}[\mathbb{m}_{is}^{ll'}(z_s, \Delta\theta_{ll'}, h_{is}^{ll'}(\cdot)) z_s^{ll'}] \geqslant 0, \tag{22}$$

holds for $\Delta\theta_{ll'} = \Delta\kappa^{ll'}$ if $z_s^{ll'} \subset \mathcal{J}_{is}$, where, as a reminder, $\mathbb{m}_{is}^{ll'}(\cdot)$ is defined in equation (8). The choice of the number of intervals q in which we split the support of $\Delta x_s^{ll'}$ is consequential for the validity of the inequalities in equation (22). If $q = 2$, $z_s^{ll'}$ includes the following two elements:

$$z_{s,1}^{ll'} = \mathbb{1}\{-\infty < \Delta x_s^{ll'} \leqslant \text{med}(\Delta x_s^{ll'})\}, \quad \text{and} \quad z_{s,2}^{ll'} = \mathbb{1}\{\text{med}(\Delta x_s^{ll'}) < \Delta x_s^{ll'} \leqslant \infty\}; \tag{23}$$

thus, the set of values of $\Delta\theta_{ll'}$ consistent with the inequalities in equation (22) includes $\Delta\kappa^{ll'}$ if worker i of type s knows, for locations l and l', whether the realized value of $\Delta x_s^{ll'}$ is above or below its median. Similarly, a choice of $q = 4$ implies the researcher uses four inequalities of the type in equation (22), and these hold for $\Delta\theta_{ll'} = \Delta\kappa^{ll'}$ if worker i of type s knows the quarter to which $\Delta x_s^{ll'}$ belongs. In general, the larger q is, the larger the number of inequalities used in estimation, and the stronger the assumptions imposed on the worker's information set. As $q \to \infty$, the inequalities in equation (22) are guaranteed to hold at the true parameter value only if worker i of type s has perfect information about $\Delta x_s^{ll'}$.

The data setting described in Section 2 includes one observation per worker. The sample analogue of the moment inequality in equation (22) thus averages over only one observation. However, as this inequality is valid for every worker i of every type s, it holds that, for any locations l and l' and function $h_{is}^{ll'}: \mathcal{Z}_s \times \Theta_{ll'} \to \mathbb{R}$, the $q \times 1$ vector of moment inequalities

$$\sum_{s=1}^{S} \sum_{i=1}^{I_s} \mathbb{E}[\mathbb{m}_{is}^{ll'}(z_s, \Delta\theta_{ll'}, h_{is}^{ll'}(\cdot))z_s] \geqslant 0 \tag{24}$$

is satisfied at $\Delta\theta_{ll'} = \Delta\kappa^{ll'}$ if $z_s \subset \mathcal{J}_{is}$ for every worker $i = 1, \ldots, I_s$ of every type $s = 1, \ldots, S$.

Given the normalization of the amenity value in location $l = 1$ (i.e., $\kappa^1 = 0$), the inequality in equation (24) may be used to compute a confidence set for the amenity level corresponding to a location $\check{l}$, $\theta_{\check{l}}$, by fixing the pair of locations (l, l') that index the inequalities in equation (24) to both $(\check{l}, 1)$ and $(1, \check{l})$. The moment inequalities in equation (24) with location indices $(l, l') = (\check{l}, 1)$ identify an upper bound on $\theta_{\check{l}}$; the same inequalities with location indices $(l, l') = (1, \check{l})$ identify a lower bound on $\theta_{\check{l}}$.

To compute a 95% confidence interval for the amenity level corresponding to a location $\check{l}$, we first compute 96% confidence intervals for $\theta_{\check{l}}$ conditional on each value of θ_α in a 99% confidence interval for this parameter. We denote these confidence intervals as $\hat{\Theta}^{\check{l}}_{.96}(\theta_\alpha)$. We compute them by applying the procedure in Andrews and Soares (2010) to the sample analogue of the moment inequalities in equation (24) for the appropriate choice of location indices l and l'; see Appendix B.11 for details. We then compute the 95% confidence interval for $\theta_{\check{l}}$ as the union of $\hat{\Theta}^{\check{l}}_{.96}(\theta_\alpha)$ for each of the values of θ_α in its 99% confidence interval. As discussed in Bei (2024), the resulting confidence interval for $\theta_{\check{l}}$ is valid.[11,12]

4 Properties of Moment Inequalities: Simulation

In this section, we illustrate the properties of the moment inequality estimator described in Section 3. We use simulations to show how our estimator performs in settings where the precision of workers' information varies and where the researchers observe all or only part of workers' information sets. Our simulation results demonstrate and complement the theoretical results in theorems 1 to 3 and corollaries 1 to 3.

4.1 Simulation Set-up

Workers choose between three possible locations $l = \{1, 2, 3\}$ according to the assumptions in equations (1) to (6). We simulate data for 6,000,000 workers, each of them of a different type;

[11] Bei (2024) shows our procedure likely yields unnecessarily wide confidence intervals for any given θ_l. She suggests an alternative procedure that yields smaller confidence intervals that still have the proper coverage.

[12] To obtain a 95% confidence set for the vector $(\theta_1, \ldots, \theta_L)$, we also report $100(1 - 0.05/L)$% confidence intervals for each θ_l for all $l = 1, \ldots, L$. In our application, $L = 50$ and, thus, the Bonferroni confidence interval for θ_l is a $100(1 - 0.05/50) = 100(1 - 0.001) = 99.9$% confidence interval computed as described in the main text.

thus, $S = 6,000,000$, $I_s = 1$ for $s = 1, \ldots, S$. We index each simulated observation by s, set the wage coefficient to $\alpha = 1$, and the location-specific amenity levels to $\kappa^1 = \kappa^2 = 0$ and $\kappa^3 = 1$.[13]

The moment-inequality estimator described in Section 3 requires specifying neither a stochastic process for wages nor the information set of every individual in the sample. However, specifying these aspects of the model is needed to generate the model-implied choice for all sampled individuals.

Concerning the stochastic process for wages, we assume the wage vector $w_s = \{w_s^l\}_{l=1}^3$ is *iid* across individuals $s = 1, \ldots, S$, and each element w_s^l follows a distribution determined by

$$w_s^l = x_{1s}^l + x_{2s}^l + x_{3s}^l, \tag{25}$$

with x_{ks}^l independent across $l = \{1, 2, 3\}$ and $k = \{1, 2, 3\}$, and distributed uniformly with support $[\mu_k^l - \sigma_k, \mu_k^l + \sigma_k]$. In our baseline simulation, we set $(\mu_1^1, \mu_1^2, \mu_1^3) = (\mu_3^1, \mu_3^2, \mu_3^3) = (0, 0, 0)$ and $(\mu_2^1, \mu_2^2, \mu_2^3) = (0, -0.5, -1)$; thus, location-specific mean wages decline in order from $l = 1$ to $l = 3$. Concerning the dispersion of wages across locations and sectors, we set $\sigma_2 = 4$ in our baseline exercise, and present results for different values of σ_1 and σ_3.

Concerning workers' information sets, we assume every worker s observes $x_{1s} = \{x_{1s}^l\}_{l=1}^3$ and $x_{2s} = \{x_{2s}^l\}_{l=1}^3$ and has no information on $x_{3s} = \{x_{3s}^l\}_{l=1}^3$. Thus, $\mathbb{E}[(x_{1s}, x_{2s})|\mathcal{W}_s] = (x_{1s}, x_{2s})$ and $\mathbb{E}[x_{3s}|\mathcal{W}_s] = \mathbb{E}[x_{3s}]$ for $s = 1, \ldots, S$. Equation (25) and this assumption on worker information imply that, for any two locations l and l', it holds

$$\mathbb{E}[\Delta w_s^{ll'}|\mathcal{W}_s] = \Delta x_{1s}^{ll'} + \Delta x_{2s}^{ll'}, \tag{26}$$

where, for $k = \{1, 2, 3\}$, we write $\Delta x_{ks}^{ll'} = x_{ks}^l - x_{ks}^{l'}$. According to equations (25) and (26), $\Delta x_{3s}^{ll'}$ equals the expectational error worker s makes when forecasting $\Delta w_s^{ll'}$. Consequently, by changing σ_3, we evaluate the impact that workers' imperfect information on wages has on the performance of the different estimators we consider.

We assume the researcher observes (y_s^l, w_s^l, x_{2s}^l) for every $s = 1, \ldots, S$. Thus, x_{1s} is a vector that belongs to the worker s's information set but the researcher does not observe. Consequently, by changing σ_1, we evaluate the impact that unobserved (by the researcher) components of a worker's information set have on different estimators.

4.2 Simulation Results

We report the main simulation results in Table 1. We discuss in Section 4.2.1 the 95% confidence intervals for the wage coefficient θ_α displayed in the column labeled *First Step*. Subsequently, we discuss in Section 4.2.2 the 95% confidence intervals for the amenity terms θ_2 and θ_3 displayed in the columns labeled *Second Step*.

[13] Consistently with equation (5), expected wage differences between locations do not vary by worker within a type s. By setting S to a large number, we limit the impact of simulation noise on workers' aggregate simulated choices.

4.2.1 Confidence Intervals for Wage Coefficient

As we describe in detail in Appendix C.1, the confidence intervals for θ_α are computed using sample analogues of the type of inequalities introduced in Section 3.2. More specifically, we build separate inequalities of this type for each pair of locations l and l'. When building each of these inequalities, we combine each observation s with an observation r trying to reproduce the conditions under which, according to Corollary 3, the resulting inequality would point identify θ_α if the restriction $\mathbb{E}[\Delta w_s^{ll'}|z_s] = \mathbb{E}[\Delta w_s^{ll'}|\mathcal{J}_{is}]$ were to hold for all $s = 1, \dots, S$.

Cases 1 and 2 share the feature that $\sigma_1 = 0$ and, thus, $x_{1s}^l = 0$ for every location l and observation s. Consequently, it is indeed the case that $\mathbb{E}[\Delta w_s^{ll'}|x_{2s}] = \mathbb{E}[\Delta w_s^{ll'}|\mathcal{J}_{is}]$ for $s = 1, \dots, S$. That is, in cases 1 and 2, the researcher observes all wage predictors on which the agent bases her decision. Consistently with the prediction of Corollary 3, the resulting 95% confidence interval for θ_α is very tight around its true value α. As a comparison of cases 1 and 2 reveals, this is true regardless of the value of σ_2; that is, regardless of whether the agent makes expectational errors when forecasting wages in every location.

There are two dimensions of the setting and of the way we build the moment inequalities that are important for the tightness of the confidence interval for θ_α in cases 1 and 2. First, one of the sufficient conditions listed in Corollary 3 for the inequality defined by equations (17) and (18) to point identify θ_α is that the locations l and l' being compared in such inequality must have equal amenity levels; i.e., $\Delta\kappa^{ll'} = 0$. As described in Section 4.1, this requirement holds in our simulation setting for locations 1 and 2, as it is the case that $\kappa^1 = \kappa^2 = 0$. To illustrate the importance of this characteristic of our simulated setting, we present in Table C.1 in Appendix C.2 confidence intervals for θ_α in settings that differ from that described in Section 4.1 exclusively in the value of the amenity terms κ^1, κ^2 and κ^3. As the results in Table C.1 show, the confidence interval for θ_α remains tight as long as $\Delta\kappa^{ll'} = 0$ for two locations l and l', and becomes wider as the differences in κ^l across all three locations $l = \{1, 2, 3\}$ increase.

Second, another of the sufficient conditions listed in Corollary 3 is that the observations s and r that get combined when forming the moment in equation (17) must verify that $\mathbb{E}[\Delta w_s^{ll'}|z_s] = \mathbb{E}[\Delta w_s^{l'l}|z_r]$. In our simulation setting, both wages and all its predictors are continuous variables and, consequently, there are no two observations s and r for which the predicted wages $\mathbb{E}[\Delta w_s^{ll'}|z_s]$ and $\mathbb{E}[\Delta w_s^{l'l}|z_r]$ coincide exactly. This is precisely why the confidence interval for θ_α reported in Table 1 includes values other than α even when all other sufficient conditions for point identification listed in Corollary 3 are satisfied. As the results in Table C.2 in Appendix C.3 reveal, the confidence interval for θ_α becomes wider as the average difference between $\mathbb{E}[\Delta w_s^{ll'}|z_s]$ and $\mathbb{E}[\Delta w_s^{l'l}|z_r]$ for the matched sectors s and r increases.

Cases 3 and 4 share the feature that $\sigma_1 > 0$ and, thus, the researcher only observes part of the agent's information set. While the true information set is (x_{1s}, x_{2s}) for every $s = 1, \dots, S$, the researcher only observes x_{2s}. From the perspective of our confidence interval for θ_α, the main consequence of this lack of complete information on the agent's information set is that the confidence interval becomes wider, incorporating not only the true value α, but also other values

Table 1: Simulation Results - Moment Inequality Confidence Intervals

Case	σ_1	σ_3	z_s	First Step	Second Step		
				α	Mom. Ineq.	κ_2	κ_3
1	0	0	x_{2s}	[1 , 1.02]	Bounding	[0 , 0]	[1 , 1]
					Odds-based	[0 , 0]	[1 , 1]
					Both	[0 , 0]	[1 , 1]
2	0	1	x_{2s}	[1 , 1.01]	Bounding	[0 , 0]	[1 , 1]
					Odds-based	[-0.33 , 0.32]	[0.68 , 1.33]
					Both	[0 , 0]	[1 , 1]
3(a)	1	0	x_{2s}	[0.82 , 1.29]	Bounding	[-0.31 , 0.31]	[0.70 , 1.30]
					Odds-based	[0 , 0]	[1 , 1.01]
					Both	[0 , 0]	[1 , 1.01]
3(b)	2	0	x_{2s}	[0.58 , 2.26]	Bounding	[-0.98 , 0.98]	[-0.02 , 1.98]
					Odds-based*	[-1.75 , 1.75]	[-0.44 , 2.75]
					Both	[-0.14 , 0.21]	[1 , 1.39]
4	1	1	x_{2s}	[0.82 , 1.31]	Bounding	[-0.31 , 0.31]	[0.69 , 1.31]
					Odds-based	[-0.38 , 0.39]	[0.68 , 1.45]
					Both	[-0.31 , 0.31]	[0.69 , 1.31]
5	0	1	w_s	[0.87 , 0.87]	Bounding	[-0.05 , -0.10]	[0.85 , 0.88]
					Odds-based	$\varnothing$	$\varnothing$
					Both	$\varnothing$	$\varnothing$

Note: The true parameter values are $\alpha = 1$, $\kappa^2 = 0$, and $\kappa^2 = 1$. The column α contains a 95% confidence interval for θ_α based on the moment inequality estimator introduced in Section 3.2 and described in detail in Appendix C.1. The columns κ^2 and κ^3 contain 95% confidence intervals for θ_2 and θ_3, respectively, based on the moment inequality estimators introduced in Section 3.1 and the inference procedure described in Section 3.3. The confidence intervals for θ_2 and θ_3 reported in the rows labeled *Bounding* use the moment inequalities introduced in Section 3.1.1; those reported in the row labeled *Odds-based* use the moment inequalities introduced in Section 3.1.2; and those reported in the row labeled *Both* combine the inequalities introduced in sections 3.1.1 and 3.1.2. For additional details on the moment inequalities used to compute the confidence sets reported in this table, see Appendix C.1. In all cases other than 3(b), confidence sets are computed using a 1-dimensional grids whose sides are $[0.5, 1.5]$ (for α), $[-0.5, 0.5]$ (for κ_2) and $[0.5, 1.5]$ (for κ_3). In case 3(b), we use grids whose sides are $[-0.75, 2.75]$ (for α), $[-1.75, 1.75]$ (for κ_2) and $[-0.75, 2.75]$ (for κ_3). All confidence sets are computed following the procedure in Andrews and Soares (2010). We mark with an asterisk when the confidence interval includes points outside the grid.

of the parameter θ_α.

In case 5, we consider a case in which the researcher wrongly assumes the agent has perfect information on wages. In this case, the confidence interval for θ_α includes only one value of this parameter (i.e., $\theta_\alpha = 0.87$), but this one does not coincide with the parameter's true value (i.e., $\alpha = 1$). As pointed out by Molinari (2020) and Andrews and Kwon (2024), this is one of the main risks of moment inequality models: when they are misspecified, they may yield confidence sets that are very tight but do not include the true parameter value. As shown in Table C.5 in Appendix C.6, the result that the confidence set for case 5 is non-empty is contingent on the small number of instruments we use to compute the confidence sets reported in Table 1. Once we increase from $q = 2$ to $q = 4$ the number of intervals in which we split the support of the wage predictor w_s when building our instruments (see Section 3.3 for extra details) our confidence set for θ_α in case 5 becomes empty.

4.2.2 Confidence Intervals for Amenity Components

Given confidence sets for θ_α, we then compute confidence sets for θ_2 and θ_3 one at a time, using the inequalities introduced in sections 3.1.1 and 3.1.2, and following the inference procedure described in Section 3.3. In Appendix C.1, we provide a more detailed description of the inequalities we use in this simulation exercise.

Consistently with the predictions of Corollary 1, the bounding moment inequalities (see Section 3.1.1) point identify the parameters θ_2 and θ_3 when $\sigma_1 = 0$ and, consequently, the agent's information sets are perfectly observed by the researcher. When $\sigma_1 > 0$ and, thus, the agent's information set is only partly observed by the researcher, the confidence intervals built using the bounding moment inequalities alone contain the true parameter values (consistently with Theorem 1) but also contain other parameter values. The set of parameter values included in the corresponding confidence set increases in σ_1.

The odds-based moment inequalities (see Section 3.1.2) have properties somewhat opposed to those of the bounding moment inequalities. When $\sigma_3 = 0$, these inequalities point identify the amenity parameters as long as the wage coefficient equals its true value α (see Corollary 2). This explains why the confidence intervals defined by the odds-based moment inequalities only include the true parameter value in case 1. When $\sigma_3 > 0$ (as in case 2) or when the confidence interval for θ_α includes values other than its true value (as in cases 3(a) and 3(b)), the odds-based moment inequalities may still point identify these parameters (as in case 3(a)) or may only partially identify them (as in case 3(b)).

The fact that the range of parameter values consistent with the bounding moment inequalities is insensitive to the extent to which agents make expectational errors when forecasting payoff-relevant variables (insensitive to σ_3) and the range of parameter values consistent with the odds-based moment inequalities is somewhat insensitive to whether the researcher observes every element in the agent's information set (partially insensitive to σ_1) implies that there are practical advantages of combining both types of inequalities in estimation. For example, in cases 2 and 3(a) in Table 1, the moment inequalities introduced in sections 3.1.1 and 3.1.2 jointly point identify the amenity parameters, although neither of these two inequality types would point identify them in both cases when considered in isolation; e.g., the confidence interval for θ_1 defined by the odds-based inequalities equals the range of values $[-0.33, 0.32]$ in case 2, and that defined by the bounding inequalities equals $[-0.31, 0.31]$ in case 3(a). In case 3(b), the confidence intervals obtained when combining the bounding and odds-based moment inequalities include values of the parameters of interest along with the true one, but to a much lesser extent than when each of the two types of moment inequalities are considered in isolation; e.g., the confidence interval for θ_2 generated by the bounding inequalities is $[-0.98, 0.98]$, that generated by the odds-based inequalities includes every point in a grid that goes between -1.75 and 1.75, but the confidence interval for θ_2 we obtain when we combine both types of inequalities shrinks to $[-0.14, 0.21]$. In Appendix C.7, we describe in more detail why the confidence intervals for θ_2 and θ_3 that we obtain when combining bounding and odds-based inequalities in cases 2, 3(a), and 3(b) are smaller than when each inequality type is considered in isolation.

In case 4, we illustrate the performance of our moment inequality estimation procedure in a setting in which the agent's information set is partly unobserved (i.e., $\sigma_1 > 0$) and the agent predicts location-specific wages with error (i.e., $\sigma_3 > 0$). This case is likely to be the most empirically relevant in our application. As the results in Table 1 show, the confidence intervals defined by the bounding inequalities, by the the odds-based inequalities, and by the combination of both, contain the true parameter values, but also other values.

In case 5, we show that the bounding moment inequalities fail to produce an empty confidence set when the researcher wrongly assumes that workers' have perfect information on location-specific wages (although, as discussed above, they do so when we use more detailed instruments). Conversely, the confidence intervals defined by the odds-based moment inequalities alone, or by both types of inequalities jointly, are empty.

4.3 Alternative Estimators

Maximum Likelihood Estimator. To facilitate the comparison of our moment inequality estimator with more traditional estimation approaches, we report in Table 2 maximum likelihood estimates (MLEs) computed under the assumption that every relevant variable in the agent's information set is observed by the researcher. That is, given a choice of an observable variable z_s assumed to be all information worker s has on w_s, we compute MLEs of $(\theta_\alpha, \theta_2, \theta_3)$ as

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

When $z_s = x_{2s}$ and, thus, $\mathbb{E}[w_s|z_s] = x_{2s}$, the MLE of $(\theta_\alpha, \theta_2, \theta_3)$ is consistent if and only if $\sigma_1 = 0$, as only then does the worker's true wage expectation coincide exactly with the researcher's assumed expectation; i.e., only then does it hold that $\mathbb{E}[w_s|\mathcal{W}_s] = \mathbb{E}[w_s|z_s]$. Conversely, when $z_s = x_{2s}$ and $\sigma_1 > 0$, the worker's true expectation and the researcher's assumed expectation do not coincide and, as a result, the MLE of all parameters are biased towards zero. When the researcher assumes workers have perfect information on wages (i.e., $z_s = w_s$ and, thus, $\mathbb{E}[w_s|z_s] = w_s$), the MLE $(\theta_\alpha, \theta_2, \theta_3)$ is biased if, contrary to the researcher's assumption, workers make errors when forecasting wages (i.e., when $\sigma_3 > 0$). A comparison of the results for cases 3(a) and 5 shows that, conditional on the difference between the worker's true expectations and the researcher's assumed expectation having the same marginal distribution (i.e., if $\sigma_1 = \sigma_3$), the bias in the MLE is smaller when the researcher assigns workers an information set that is too small than when it assigns them an information set that is too large. For more results on the bias in the MLE depending on the researcher's assumptions on the content of the agent's information set, see Dickstein and Morales (2018).

A comparison of the estimates in tables 1 and 2 yields two conclusions. When the researcher observes a subset of the worker's true information set (i.e., if $\sigma_1 > 0$), the MLE is biased and

Table 2: Simulation Results - MLE

Case	σ_1	σ_3	z_i	α	κ_2	κ_3
1	0	0	x_{2s}	1	0	1
2	0	1	x_{2s}	1	0	1
3	1	0	x_{2s}	0.91	0	0.92
3(b)	2	0	x_{2s}	0.75	0	0.75
4	1	1	x_{2s}	0.91	0	0.92
5	0	1	w_s	0.87	-0.03	0.87

Note: *MLE* denotes the maximum likelihood estimate.

our moment inequality estimator yields confidence intervals that contain the true parameter value. Furthermore, as illustrated by the results for both case 3(a), 3(b), and 5, the confidence intervals produced by our moment inequality estimator may not include the corresponding MLEs; e.g., in case 3(b), the 95% confidence interval for θ_3 when bounding and odds-based inequalities are used jointly is $[1, 1.39]$, while the MLE is 0.75.

One-step Moment Inequality Estimator When Choice Set is Small. Relative to the approach in Dickstein et al. (2023), the main contribution of the two-step moment inequality estimator described in Section 3 is that it yields valid confidence intervals for each parameter in a discrete-choice setting in which the agent's utility depends on a potentially very large number of choice-specific fixed effects. This added generality is important in our setting, as worker location choices typically involve large choice sets. However, the simulation setting described in Section 4.1 involves a choice set with only three alternatives and, thus, it is feasible to compute a confidence set for the parameter vector $(\theta_\alpha, \theta_2, \theta_3)$ in one step, following the moment inequality procedure introduced in Dickstein et al. (2023). This procedure uses moment inequalities analogous to those introduced in Section 3.1. The only difference is that, when applying the bounding and the odds-based inequalities described in that section, the wage coefficient is treated as an unknown parameter and the inequalities for all possible location pairs $l = 1, \dots, L$ and $l' \neq l$ are used jointly to estimate a confidence set for the whole parameter vector $(\theta_\alpha, \theta_2, \theta_3)$. We report in Table C.3 in Appendix C.4 the projection on each parameter of the three-dimensional confidence set computed following that procedure. A comparison of the results in tables 1 and C.3 shows that our procedure yields very similar confidence sets for the amenity parameters θ_2 and θ_3, but yields larger confidence intervals for θ_α when agents' information sets are partly unobserved by the researcher (i.e., when $\sigma_1 > 0$). There is thus a trade-off between, on the one side, feasibility and computation time and, on the other side, tightness of the resulting parameter-specific confidence intervals. Relative to the procedure in Dickstein et al. (2023), our moment inequality estimator makes it feasible to compute parameter-specific confidence intervals in settings with large choice sets and choice-specific fixed effects, and it is significantly faster even in settings in which applying the procedure in Dickstein et al. (2023) is feasible. Conversely, our procedure does not yield a confidence set for the parameter vector that is larger than that Dickstein et al. (2023).

Alternative Moment Inequality Estimator When Choice Set is Large. Using only inequalities analogous to those described in Section 3.1 (that is, without using the first-step inequalities in Section 3.2), one can devise an inference procedure that yields valid confidence intervals for each parameter even in settings with a large number of choice-specific fixed effects. This procedure fixes one of the location indices l and l' that define the inequalities in Section 3.1 to the outside choice (i.e., $l = 1$), and uses these inequalities to compute valid confidence intervals for the the parameter vector $(\theta_\alpha, \theta_l)$ for a specific location $l > 1$. Doing this repeatedly for all $L - 1$ inside choices, it yields $L - 1$ two-dimensional confidence sets. By projecting them on each of its elements, we obtain a valid confidence interval for each of the $L-1$ amenity terms $(\theta_2, \ldots, \theta_L)$, and $L-1$ valid confidence intervals for θ_α. In Table C.4 in Appendix C.5, we report the resulting confidence intervals for each parameter, with the reported one for θ_α computed as the union of the $L - 1$ valid confidence intervals described above.

5 Empirical Application

Our empirical application focuses on individual migration events across local labor markets in Brazil between 2002 and 2011. We define a labor market as the combination of a region and an industry, and we focus on migration events that entail a change of location but not of industry. In that context, we interpret a worker's type s, defined in Section 2, as the combination of their industry, demographic group, and current location.

5.1 Data

Our main data source for transitions across regional labor markets is the *Relação Anual de Informações Sociais* (RAIS). RAIS is an administrative dataset assembled yearly by the Brazilian Ministry of Labor, recording information about workers, establishments, and characteristics of the labor contract, for the Brazilian formal labor market. We utilize the establishment's geographic location (microregion, the closest equivalent to a commuting zone) and industry to define labor markets, and worker-level information, including gender, age, education (two categories), reported race (two categories), and average monthly earnings.

Since RAIS only covers formally employed workers, a worker may be absent from the dataset either because they are out of the labor force, unemployed, informally employed, or self-employed. The informality rate was close to 50 percent during our sample period, our sample hence omits a sizeable share of individuals in Brazil. We restrict our sample to individuals who have a strong attachment to the formal labor market, selecting only those that are recorded in RAIS for at least seven years during our ten-year analysis period. We further limit our sample to include working-age individuals who are likely to have finished their education, aged 25–64; we omit those working in public administration and those without valid information on the industry of employment.

To ensure a large enough number of individuals observed in each labor market, we select 1,000 labor markets consisting of the combinations of the 50 microregions (out of 558) and 20 industry sectors (out of 51) with the largest total employment reported in RAIS. The results in this section are obtained from our primary working sample obtained after selecting 1 million individuals per year among those 1,000 labor markets that belong to the largest demographic group in RAIS, consisting of white males with at least a high school education. In Appendix D, we provide more details on the RAIS data and the construction of our working sample. We also report summary statistics on the migration rates, earnings, and other characteristics of the selected individuals and labor markets.

5.2 Estimation of Preference Parameters

5.2.1 Specification of Moments

We implement the empirical equivalents to (17) to estimate the migration elasticity α, and to (8) and (13) to estimate the migration costs $\{\kappa^l_{nt}\}$, with the information on migration and wages in our dataset. We describe how we define each element of the moment functions.

Migration. Each individual in our working sample is observed for two consecutive years. We define the indicator variable y^l_{ist} to take value 1 if worker i of type s, which codes their demographic group, sector, and microregion in year $t-1$, is observed in microregion l in year t. By construction, our sample only includes individuals whose demographic group and sector remain unchanged across the two years. We define all possible observations of (17) by pairing each worker i with all workers j that are also of type s in year $t-1$.

Wages. While our data contains worker-specific information on the realized wages for those sector-region pairs in which each worker is employed in any given year, they naturally do not include information on the hypothetical wages that this worker could have obtained if they had chosen a region different than that observed in the data. We construct a proxy for those unobserved wage levels. We allow the log wages that any worker would obtain if employed in a particular sector and region to depend on individual worker characteristics (age, gender, education level) with year-specific coefficients, region-year-specific fixed effects, worker-sector-specific fixed effects, and an unexpected region-sector-year-individual-specific unobserved term.[14] In Appendix E.1, we provide more detail on the exact specification and report the results of our regressions. From these regressions, we obtain estimates of local labor market wage residuals $\{w^l_{st}\}_l$ for each sector s and period t.

[14]This wage equation does not include a region-, sector- and individual-specific term that is constant over time, unobserved by the researcher, and potentially known by workers when deciding on their preferred region and employment sector. See Kennan and Walker (2011); Dix-Carneiro (2014); Traiberman (2019) for specifications of wage equations that allow for such a term in the context of dynamic discrete choice models of labor mobility. Allowing for such a component of wages poses estimation challenges when, as is our case, one imposes only a minimum content requirement on agents' information sets.

Wage Predictors. The instruments that appear in the moment inequalities depend on the assumptions we make about the content of workers' information sets. These assumptions determine which wage predictors x^l_{st} of the log wage residuals w^l_{st} in each labor market belong to workers' information sets and, therefore, satisfy the exclusion restriction. For example, if we assume that workers know the previous year's wage residuals, then we set $x^l_{st} \equiv w^l_{st-1}$. In other cases, however, we assume that workers can only observe whether a given lagged wage belongs to one of b quantiles, or bins, of the lagged wage distribution, where $b \in \mathbb{N}$. To compute a wage predictor that is known by workers under this assumption, we define the vector $(B^b_0, \ldots, B^b_b)$ of b-quantiles of the distribution of w^l_{st-1} across all sectors and locations (i.e., $\Pr(w^l_{st-1} \leqslant B^b_k) = k/b$ for $k \in [0, b]$), and denote by $k^{ll'}_{st}$ the index of the largest b-quantile below w^l_{st-1}. We define the predictor as the mean value of w^l_{st-1}, conditional on belonging to the interval above the b-quantile $k = k^{ll'}_{st}$,

$$x^l_{st} \equiv \mathbb{E}[w^l_{st-1} | B^b_{k-1} < w^l_{st-1} \leqslant B^b_k], \quad \text{s.t. } k = k^{ll'}_{st} \in [1, b]. \tag{27}$$

We refer to the number of bins, b, of lagged wages that workers can distinguish as the level of precision of their information. Note that when b becomes arbitrarily large, the b-bin predictor becomes arbitrarily close to w^l_{st-1}.

Instruments. The moments in equations (17), (8), and (13), depend on a generic instrument vectors z_s. We construct instruments for our empirical application $z^{ll'}_{st}$ following the intuition in Corollary 3 that forming separate moments with positive and negative wage differences identifies tighter bounds on the migration elasticity. First, for each pair of potential destinations l and l', we follow (27) to form a predictor $\Delta x^{ll'}_{st}$ of the wage difference $\Delta w^{ll'}_{st}$ consistent with the information available to workers of type s about locations l and l'. We build distinct instruments for predicted wage differences that belong to different quantiles of their distribution. We also weight observations by the magnitude of the predicted wage difference, building two instruments for each quantile of the predicted wage differences: one that weights them by their absolute value, the other by the inverse of their absolute value. Hence, given the vector $(Q^q_0, \ldots, Q^q_q)$ of q-quantiles of the distribution of $\Delta x^{ll'}_s$ across all sectors s and location pairs l and l' (i.e., $\Pr(\Delta x^{ll'}_s \leqslant Q^q_k) = k/q$ for $k \in [0, q]$), and for $d \in \{-1, 1\}$, we set our instrument vector for the wage difference $\Delta w^{ll'}_{st}$ as

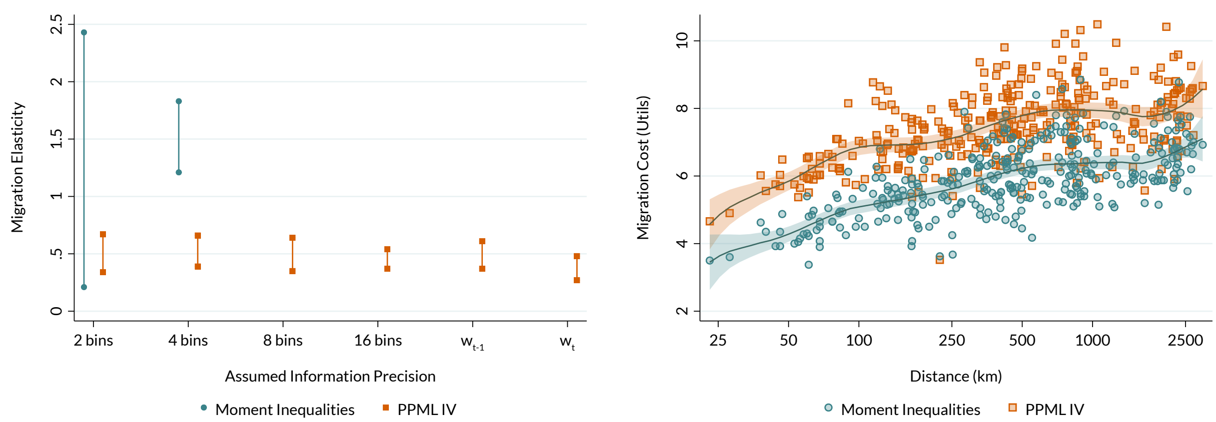
$$z^{ll'(d)}_{st} = (z^{ll'(d)}_{st,1}, \ldots, z^{ll'(d)}_{st,q})' \quad \text{with} \quad z^{ll'(d)}_{st,k} \equiv \mathbb{1}\{Q^q_{k-1} < \Delta x^{ll'}_{st} \leqslant Q^q_k\} |\Delta x^{ll'}_{st}|^d \quad \text{for } k \in [1, q]. \tag{28}$$

In our application, we set $q = 4$, yielding four quantiles (two for negative predicted wage differences and two for positive differences) and eight instruments after raising the absolute differences in (28) by either $d = 1$ or $d = -1$. We build the instruments for (17), which involves two wages differences for sectors s and r for each pair of destinations l and l', as the product $z^{ll'(d)}_{st} z^{ll'(d)}_{rt}$.

Points of Linear Approximation. The bounding inequalities in (8) and (17) are linear in the parameters $\{\kappa^l_{nt}\}$ and α, respectively. The linearity results from the first-order approximation of

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

(a) Migration Elasticity (95% conf. int)   (b) Migration Costs (midpoints of 95% conf. ints)



This figure displays the main estimates of the preference parameters from our working sample. Panel (a) reports the 95-percent confidence intervals of the migration elasticity α under different information assumptions. Starting from the weaker assumption that individuals only know the 2-bin accuracy of lagged wages, we increase the assumed precision to 4, 8, and 16 bins and finally to the exact values of lagged wages. The blue circles delimit the intervals from our moment inequality estimator. The orange squares mark the intervals from the two-step PPML-IV procedure described in Appendix E.2.2 for comparison. The intervals for any level of precision above or equal to 8 bins are empty, which leads us to reject that workers possess those levels of information. In Panel (b), each point shows the midpoint of the 95-percent confidence interval for a given bilateral migration cost $\{\kappa^l_{nt}\}$ in the year 2011, expressed in utils. The blue circles depict the estimates from our moment inequalities estimator, while the orange squares result from the two-step PPML-IV procedure. The fit lines represent the kernel-weighted local polynomial smoothing of each series, with the shaded areas representing the 95-percent confidence interval around the smoothed series.

the concave, non-linear, moment functions obtained from the initial revealed-preference inequality. Equations and (10) and (19) provide the expressions of the points of approximation $h^{ll'}_{is}(z_s, \Delta\theta_{ll'})$ and $g^{ll'}_{ijsr}(z_s, z_r, \alpha)$ that yield the tightest identified set for the migration costs and migration elasticity. For each pair of locations l, l', and sectors s, r, these expressions depend on α—one of the parameters we aim to identify—and on the expectation of the mean wage difference between two sectors held by individuals who moved to labor markets (s, l) and (r, l'), respectively. Since neither is known to us, for each of our eight instruments, we define M moments, each at a different point of approximation g_m, $m = 1, \dots, M$, that approximates the optimal functions h and g by replacing α with a guessed value α_m, and the expectation in (10) and (19) by the difference in wage predictors:

$$h_m = \alpha_m \Delta x^{ll'}_{st}, \qquad g_m = 0.5\alpha_m \left(\Delta x^{ll'}_{st} + \Delta x^{l'l}_{rt}\right). \tag{29}$$

We set $M = 30$, and α_m taking equidistant values in $[0, 3]$. Hence, our estimation of the migration elasticity α involves $8M = 240$ distinct moments.

5.2.2 Results

We estimate the migration elasticity under several assumptions on the precision of workers' information about wages. Our main results from the estimation of the migration elasticity and migration costs are summarized in Figure 1.

Migration Elasticity. Panel (a) in Figure 1 reports the 95-percent confidence intervals of the migration elasticity α under different information assumptions. Starting from the weaker assumption that individuals only know the 2-bin accuracy of lagged wages, we increase the assumed precision to 4, 8, and 16 bins and finally to the exact values of lagged wages.

For the four highest levels of information precision, we obtain an empty interval for the migration elasticity. This leads us to reject that workers know any level of precision of lagged wages above 4 bins, or that they know the exact contemporaneous wages. When we only assume that workers know the 4-bin precision of lagged wages, we obtain a 95-percent confidence interval for α equal to [1.21, 1.83]. We conclude that we cannot reject that workers can discern the 4-bin precision of lagged wages. When we relax the assumption further and only assume that workers know the 2-bin precision of lagged wages (above or below its median), we obtain a non-empty but wide confidence interval equal to [0.23, 2.46]. This interval being wider is consistent with the assumption of 2-bin wages having less identification power.

For comparison, we also include the 95-percent confidence intervals obtained from the two-step PPML-IV procedure derived in Artuç and McLaren (2015). As we describe in Appendix E.2.2, this alternative point-identified estimation method allows for some level of incomplete information but is only valid in this context if all workers have exactly the same information within the same sector, so that their expectations of the wages in other labor markets are identical. This is a much stronger assumption than the one required by our moment inequality estimator. For example, testing for the 4-bin level of precision of lagged wages with our estimator does not assume that all workers know exactly the 4-bin precision. Some individuals may have more information than others within any sector, location, period, or demographic group, as long as they all know at least the 4-bin precision. For this reason, it is not surprising that the estimates from these two methods are quantitatively different, with the PPML-IV estimates contained between 0.3 and 0.6, while the lowest value in our preferred moment inequality estimator is 1.21.

Our larger estimate of the migration elasticity relative to the PPML-IV procedure indicates that workers value income relative to other non-income location payoffs to a higher degree. As our counterfactual exercises will confirm in Section 5.4, a larger value of theta implies that the cost of incomplete information about wages is higher since obtaining a lower-than-predicted wage in a location leads to a larger decline in utility.

Migration Costs. Panel (b) in Figure 1 displays our estimates of migration costs across all pairs of locations for the year 2011 that result from the moment inequalities estimator in (24). As discussed in Section 3.3, for each pair of locations (n, l) and period t, we compute a 95-percent confidence interval for κ_{nt}^{l} by computing a 96-percent confidence interval for κ_{nt}^{l} for all values of α inside the 99-percent confidence interval.

Although we obtain a confidence interval around each bilateral migration cost κ_{nt}^{l}, for clarity, Panel (b) only depicts the midpoints of the 95-percent confidence intervals, expressed in utils.[15] For comparison, we also display the estimates of migration costs from the two-step PPML-IV procedure for the same locations and year (Artuç and McLaren, 2015).[16] The median value of migration cost estimates is 7.45 from the PPML-IV method and 5.88 from the moment inequalities. Hence, our moment inequalities estimator yields migration costs that are 21 percent smaller. If we convert migration costs into their log-wage equivalents by dividing them by the estimates of α produced by each method, our estimates are 74 percent smaller.

To interpret the magnitude of these migration costs, we can compare them to the utility gain from a typical increase in earnings from a migration decision. In our sample, the standard deviation of residual log wages across labor markets is 1.70. Hence, if we set the migration elasticity to 1.5, the central value of our preferred confidence interval in Panel (a), this one standard deviation increase in log wages results in a $1.70 \times 1.5 = 2.55$ additional utils. Therefore, to be indifferent between paying the median migration cost of 5.88, a worker would need to obtain a log wage that is higher than their current log wage by $5.88 / 2.55 = 2.31$ standard deviations of the cross-labor market log wage distribution.

5.3 Tests of Information Heterogeneity

Our results in Section 5.2.2 showed that our moment inequalities estimator only identifies a non-empty interval for the migration elasticity when we assume that the precision of the information held by all workers about wages in all labor markets is not higher than 4 bins. It is likely, however, that some workers have more information than others and that most workers might have more information in some labor markets than others. For example, workers may know more about wages in nearby destinations or about destinations with stronger ties in the form of past migration flows, larger populations, and better access to the Internet.

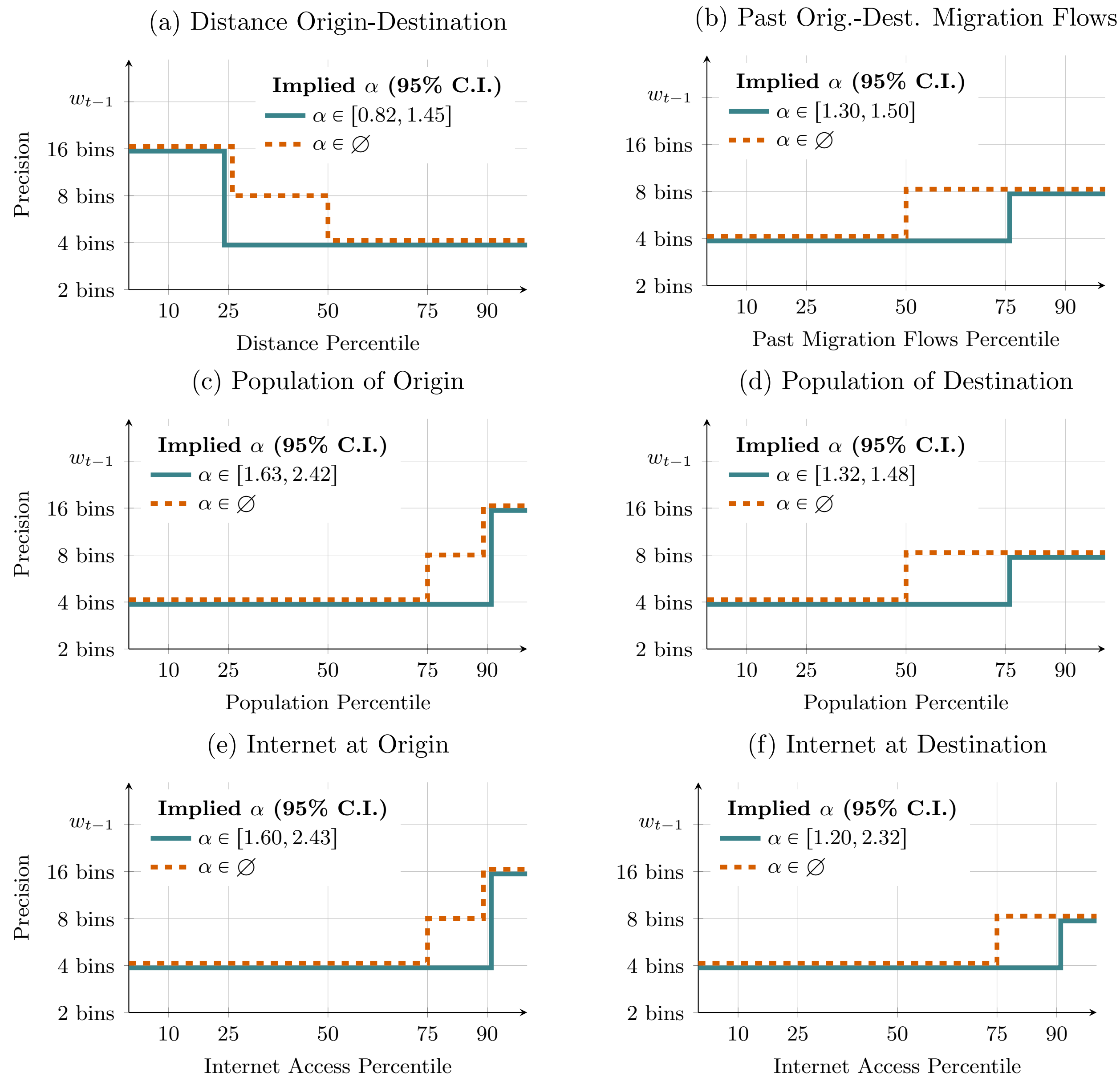
In this section, we explore whether the migration patterns in our sample are consistent with some agents having higher precision about some, but not all, locations. We implement those tests by performing the estimation of the migration elasticity α under assumptions that let workers have more information about one key dimension at a time. We consider six dimensions: geographic distance, past migration flows, the population of origin and destination, and the share of households with internet access in the origin and destination.

We measure each of these dimensions in the following ways: the geographic distance between any pair of microregions is defined as the geodesic distance between their population centroids; past migration flows from any origin to any destination are set to the total number of workers

[15] See Figure E.3 in Appendix E.3 for a representation of the confidence intervals around each of these estimates. The median width of the confidence interval is 0.50 utils.

[16] The two-step PPML-IV procedure does not allow for computing confidence intervals around the origin-destination-year fixed effects κ_{nt}^{l}.

Figure 2: Testing for Heterogeneous Information Sets



This figure displays patterns of information precision that can and cannot be rejected in the data, with each panel showing how these patterns vary along a key dimension, including distance, past migration flows, population of origin and destination, and the share of households with internet access in the origin and destination. Patterns of information precision that can be rejected are shown in dotted lines, while patterns that cannot be rejected are shown in solid lines. We test each hypothesis by building an instrument function that defines wage proxies according to the assumed precision of information. These wage proxies reflect the characteristics of the origin and destination labor markets. For example, in the case of distance, the non-rejected pattern is tested by defining the wage predictor in a far-away labor market (beyond the 25^{th} quantile of distance) with lower precision (4 bins of lagged wages) than wage predictor in a nearby labor market (16 bins). We report the estimated 95% confidence interval of the migration elasticity that is consistent with the information assumption in each test.

recorded in RAIS as having migrated from that origin to that destination during the three years before our sample analysis (1999-2001); the population of each microregion is computed as the total employment in RAIS in the years 1999 to 2001; the share of households with internet access in each microregion is defined as the mean share of households with broadband internet access between 2007 and 2011, the period for which that information is recorded by the *Agência Nacional de Telecomunicações* (ANATEL). We provide more detail about the definition of these variables and descriptive statistics in Section D.

We investigate the heterogeneity of information by following a similar iterative approach for each of the six dimensions. We divide the support of each heterogeneity variable into six intervals delimited by the 10th, 25th, 50th, 75th, and 90th percentiles. Starting from a baseline level of precision of 4 bins, we first postulate that workers have a higher level of precision for values of the heterogeneity variable in the highest or lowest interval, depending on the direction along which we suspect information might be more precise. Our first test in this procedure always assigns a level of precision of 8 bins to labor markets that fall in that extreme interval, keeping 4 bins for all others. This translates into wage predictors (28) with different levels of precision across different combinations of origin and destination labor markets, affecting the values of instruments and, therefore, changing the identified set. If the 95-percent confidence interval for α is empty, we reject that assumption and end the procedure. If it is not empty, we increase the level of precision to 16 bins on that extreme interval and perform a new estimation. Calling B_j the maximum precision level obtained on the jth interval tested that is not rejected, the next iterations maintain B_j on that interval and test for all levels of precision on the interval $j+1$ up to precision B_j. This procedure yields a weakly monotone information schedule along each dimension.

Figure 2 displays patterns of information precision that result from this procedure. For all dimensions, we are able to estimate a non-empty confidence interval for the migration elasticity α under information assumptions that are richer than 4 bins on the extreme intervals. Panel (a) shows that we can't reject that workers know the 16-bin level of precision of lagged wages in labor markets that are within the 25th percentile of distance (383 km), and 4 bins for all other locations. However, we reject any higher information level on that interval, and also reject that workers also know the 8-bin precision over the 25th-50th percentile interval. In Panel (b), we can't reject that workers know the 8-bin precision, of lagged wages about labor markets that are connected to their current location by past migration flows above the 75th percentile. Any higher level of precision on that interval is rejected, and so is the assumption that the 8-bin precision extends to locations with past migration flows in the 50th-75th percentile interval. Panel (c) shows that the migration patterns in our data are consistent with workers in the microregions with populations above the 90th percentile knowing the 16-bin precision of lagged wages. Since our sample selects the 50 largest microregions, the 90th percentile essentially selects the 5 largest microregions in Brazil. Panel (d) shows we can't reject that workers know the 8-bin precision of lagged wages in the destination labor markets with populations above the 75th percentile. Finally, Panels (e) and (f) show that we don't reject that workers know the 16-bin precision if they live in microregions with the highest internet access or the 8-bin precision if they consider destinations with high internet access.

Our results confirm that when workers consider moving to other labor markets, they have more information about some labor markets than others, and this level of information partly depends on where they are. Our analysis, however, only correlates the underlying structure of information with observable variables without identifying a causal relationship between those variables and the level of information. For example, internet access and population size are positively correlated, and while information appears to improve along both these dimensions, we are not able to identify which of these two, if any, has the strongest causal effect.

5.4 Counterfactuals

In this section, we undertake counterfactual exercises in a simulated economy calibrated to our sample to illustrate the effects of providing better information to individuals about wages when they consider migrating. These exercises allow us to compare the changes in migration decisions resulting from only improving the precision of information about wages, keeping constant all other aspects of the economic environment constant.

We simulate an economy with 50 regions and 20 sectors over 10 years. We calibrate the migration costs between regions as a function of distance by regressing the midpoint estimates of the bilateral migration costs obtained from the moment inequality estimation (see Panel (b) in Figure 1) on a fixed cost and a variable cost linear with log distance. We set the migration elasticity to the midpoint value of our baseline interval, to 1.51 (see Panel (a) of Figure 1). From the residual wages specific to each location, sector, and year, estimated in Appendix E.1, we fit an AR(1) process with a common persistence and variance of innovation, but an intercept composed of a sector and a location fixed effect. The mean intercept across labor markets is 0.36, the persistence parameter is 0.93, and the standard deviation of the innovation process is 0.43.

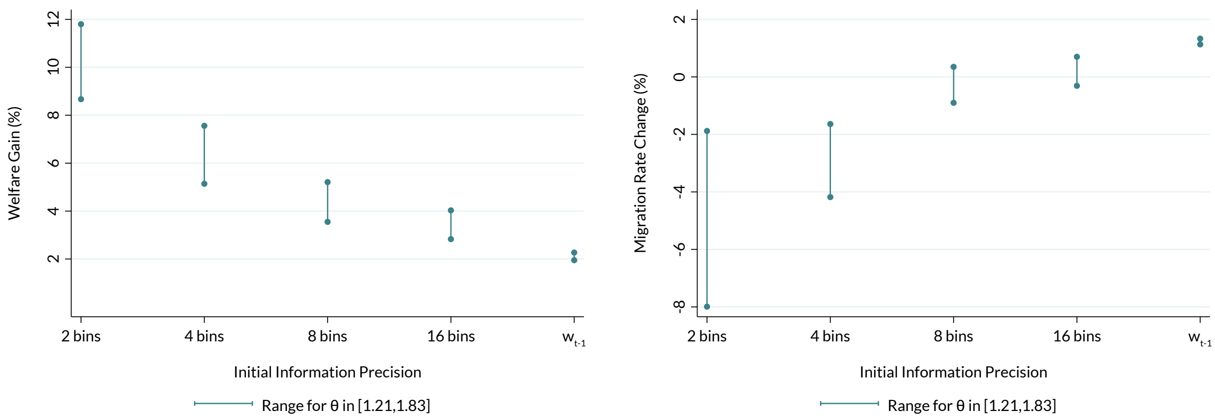
Our counterfactual exercises compare the simulated migration decisions and welfare of agents that either (i) have complete information about the realization of the wages in each location and year, or (ii) only know in which of Q quantiles of the economy-wide distribution each labor market's lagged wage belongs to, with a different simulation for $Q \in \{2, 4, 8, 16\}$ and the limit when Q is arbitrarily large and the exact lagged wages are known.

Our counterfactuals allow us to predict, in a specific partial-equilibrium context, the potential gains from expanding workers' informational content about the spatial distribution of wages. We measure welfare by averaging the per-period utility of all agents, including the contribution of their idiosyncratic tastes for locations and, importantly, using the realized wages as their income. Hence, individuals moving with complete information are exactly maximizing their utility, while individuals with incomplete information are prone to mistakes, and choose locations that offer a lower utility.

Each simulation depends on the realized draws from the AR(1) processes in each labor market. Therefore, we report the results from 100 simulations. We report the percentage loss in average welfare between the two simulated economies when agents respond to the same realized wages but

Figure 3: Counterfactual Changes from Providing Full Information About Wages

(a) Welfare Changes ($\alpha \in [1.21, 1.83]$) (b) Migration Rate Changes ($\alpha \in [1.21, 1.83]$)



This figure displays the counterfactual changes in simulated economies transitioning to complete information from initial levels of incomplete information with precision ranging from 2 bins of lagged wages to exact lagged wages. The intervals display the range of counterfactual changes obtained for all values of the migration elasticity, α, in the 95-percent confidence interval. We calibrate our simulated economy to our working sample on all aspects besides information; there are 50 locations and 20 sectors; we draw 1 million individuals each year and simulate their migration decisions over ten periods; wages follow an AR(1) process in each labor market that is estimated on observed wages in the data; migration costs are set as a fixed moving cost and a variable linear cost in log distance, both calibrated from our estimated migration costs; we consider values of the migration elasticity inside our estimated 95-percent confidence interval. Since the outcomes of each counterfactual depend on the realized sequence of wages, we compute the mean change over 100 simulated wage sequences for each value of α.

with complete or incomplete information.

The results are displayed in Figure 3. Panel (a) shows that the largest average welfare gains are obtained from a counterfactual change from an economy in which workers only know the 2-bin precision of lagged wages to one in which they have complete information. The gains are between 8.3 and 11.9 percent, for a range of migration elasticities α inside our 95-percent interval. The gains decrease but remain sizeable for information changes that start from higher levels of precision. They are between 5.0 and 7.8 percent when workers start from the 4-bin precision, between 3.6 and 5.4 when they start from the 8-bin precision, between 2.7 and 4.1 when they start from the 16-bin precision, and between 2.0 and 2.2 percent when they start from the exact lagged wages.

Panel (b) illustrates that the changes in total migration flows from moving to complete information vary significantly depending on the initial level of information. For the lowest levels of information precision, migration flows decrease once workers move with incomplete information. At higher levels of precision, the effect can be positive, especially when workers start from the knowledge of lagged wages. These results emphasize that information frictions may not always reduce the probability that workers move between labor markets, even though it increases the probability that they are located in a labor market that doesn't achieve the highest possible utility for them.

6 Conclusion

In this paper, we study the extent to which workers' limited migration responses to local labor demand shocks are due to their lack of information about their potential net gains from regional migration. Using data from 2002 to 2011 on the work location of every legally employed Brazilian worker, we propose an empirical approach to estimate information frictions in a context where migration decisions result from agents maximizing their expected utility from moving to any location. Using our revealed preference method, we first show that estimates of the elasticity of migration with respect to wages, a parameter with critical welfare implications in structural analyses of migration patterns, are highly dependent on the assumptions about the information available to migrants. Our preferred estimates, which rely only on agents knowing the wage quartile to which wages in a labor market belonged to the previous year, are almost three times larger than the estimates obtained under the common assumption that agents know the exact wage difference at the microregion level in the previous year.

The results from the implementation of our testing procedure reveal two key patterns. First, in stark contrast with the common assumption in the literature of perfect information about wages, we find strong evidence of incomplete information for all workers in our sample. Second, workers' information sets are heterogeneous, with characteristics of where they reside and the location they consider affecting the extent to which they are informed about annual wages. Our results show that gravity in terms of geographic distance, as well as the degree of internet penetration and population density in the region of residence of a worker, determine the extent to which workers are informed about the net gains of migrating to any other Brazilian region.

Our partial-equilibrium counterfactual exercises emphasize the importance of accounting for incomplete information when predicting the effect of policies affecting variables relevant to the migration decision. They also indicate the significant potential gains from improving access to information. The heterogeneity in workers' information sets we document will likely have important additional welfare implications. Workers living in the best-informed regions and part of the best-informed demographic categories may have a significant advantage when making their migration decisions. By exploiting their more precise information when choosing where to move (and whether to move), more informed workers are likely to enjoy higher wages and amenity levels on average than less informed workers.

References

Andrews, Donald W. K. and Soonwoo Kwon, "Misspecified Moment Inequality Models: Inference and Diagnostics," *Review of Economic Studies*, 2024, *91* (3), 45–76. [19]

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, 14, 19, 20, 21, 22]

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. [3, 4, 27, 28]

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]

Bayer, Patrik, Robert McMillan, Alvin Murphy, and Christopher Timmins, "A Dynamic Model of Demand for Houses and Neighborhoods," *Econometrica*, 2016, *84* (3), 893–942. [4]

Bei, Xinyue, "Inference on Union Bounds with Applications to DiD, RDD, Bunching, and Structural Counterfactuals," *mimeo*, January 2024. [16]

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]

Bishop, Kelly C. and Alvin D. Murphy, "Estimating the Willingness to Pay to Avoid Violent Crime: A Dynamic Approach," *American Economic Review: Papers & Proceedings*, 2011, *101* (3), 625–629. [4]

Borjas, George J., "Self-Selection and the Earnings of Immigrants," *American Economic Review*, 1987, *77* (4), 531–553. [4]

Bryan, Gharad, Shyamal Chowdhury, and Ahmed Mushfiq Mobarak, "Under-investment in a Profitable Technology: The Case of Seasonal Migration in Bangladesh," *Econometrica*, 2014, *82* (5), 1671–1748. [1, 5]

Caliendo, Lorenzo, Luca David Opromolla, Fernando Parro, and Alessandro Sforza, "Goods and Factor Market Integration: A Quantitative Assessment of the EU Enlargement," *Journal of Political Economy*, 2021, *129* (12), 3491–3545. [4]

_ , Maximiliano Dvorkin, and Fernando Parro, "Trade and Labor Market Dynamics: General Equilibrium Analysis of the China Trade Shock," *Econometrica*, 2019, *87* (3), 741–835. [4]

Canay, Ivan A., Gastón Illanes, and Amilcar Velez, "A User's Guide for Inference in Models Defined by Moment Inequalities," *Journal of Econometrics*, 2023, *accepted.* [2]

Clemens, Michael A., "Economics and Emigration: Trillion-Dollar Bills on the Sidewalk?," *Journal of Economic Perspectives*, September 2011, *25* (3), 83–106. [1]

Dahl, Gordon B., "Mobility and the Return to Education: Testing a Roy Model with Multiple Markets," *Econometrica*, 2002, *70* (6), 2367–2420. [4]

De la Roca, Jorge and Diego Puga, "Learning by Working in Big Cities," *Review of Economic Studies*, January 2017, *84* (1), 106–142. [1]

Desmet, Klaus, Dáavid Kristián Nagy, and Esteban Rossi-Hansberg, "The Geography of Development," *Journal Political Economy*, 2018, *126* (3), 903–983. [4]

Diamond, Rebecca, Tim McQuade, and Franklin Qian, "Who Benefits from Rent Control? The Equilibrium Consequences of San Francisco's Rent Control Expansion," *Working Paper*, 2019. [7]

Dickstein, Michael J. and Eduardo Morales, "What do Exporters Know?," *The Quarterly Journal of Economics*, 2018, *133* (4), 1753–1801. [2, 3, 5, 21, 14]

_ , Jihye Jeon, and Eduardo Morales, "Patient Costs and Physicians' Information," *mimeo*, 2023. [9, 10, 22, 7, 16, 20]

Dix-Carneiro, Rafael, "Trade Liberalization and Labor Market Dynamics," *Econometrica*, 2014, *82* (3), 825–885. [3, 8, 24, 25]

_ and Brian K. Kovak, "Trade Liberalization and Regional Dynamics," *American Economic Review*, 2017, *107* (10), 2908–2946. [25]

_ and _ , "Margins of Labor Market Adjustment to Trade," *Journal of International Economics*, 2019, *117*, 125–142. [25]

Dustmann, Christian and Ian P. Preston, "Free Movement, Open Borders, and the Global Gains from Labor Mobility," *Annual Review of Economics*, 2019, *11* (Volume 11, 2019), 783–808. [1]

Felix, Mayara, "Trade, labor market concentration, and wages," *Working paper*, 2022. [25]

Ho, Katherine and Ariel Pakes, "Hospital Choices, Hospital Prices, and Financial Incentives to Physicians," *American Economic Review*, 2014, *104* (12), 3841–3884. [5]

Humlum, Anders, "Robot Adoption and Labor Market Dynamics," *Working Paper*, 2020. [7, 8]

Jensen, Robert and Nolan H Miller, "Keepin' 'em Down on the Farm: Migration and Strategic Investment in Children's Schooling," Working Paper 23122, National Bureau of Economic Research February 2017. [5]

Jia, Ning, Raven Molloy, Christopher Smith, and Abigail Wozniak, "The Economics of Internal Migration: Advances and Policy Questions," *Journal of Economic Literature*, 2023, *61* (1), 144–180. [1]

Kaplan, Greg and Sam Schulhofer-Wohl, "Understanding the Long-run Decline in Interstate Migration," *International Economic Review*, 2017, *58* (1). [5, 6]

Kennan, John, "Open borders," *Review of Economic Dynamics*, 2013, *16* (2), L1–L13. [1]

_ and James R. Walker, "The Effect of Expected Income on Individual Migration Decisions," *Econometrica*, 2011, *79* (1), 211–251. [1, 4, 8, 24]

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. [19]

Morten, Melanie and Jaqueline Oliveira, "The Effects of Roads on Trade and Migration: Evidence from a Planned Capital City," *American Economic Journal: Applied*, 2024, *Forthcoming.* [4]

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, 6]

Redding, Stephen J. and Esteban Rossi-Hansberg, "Quantitative Spatial Economics," *Annual Review of Economics*, 2017, *9* (1), 21–58. [7, 8]

Roy, A. D., "Some Thoughts on the Distribution of Earnings," *Oxford Economic Papers, New Series*, 1951, *3* (2), 135–146. [4]

Scott, Paul, "Dynamic Discrete Choice Estimation of Agricultural Land Use," *Working Paper*, 2013. [5, 7]

Sjaastad, Larry A., "The Costs and Returns of Human Migration," *Journal of Political Economy*, 1962, *70* (5), 80–93. [1]

Szerman, Christiane, "The labor market effects of disability hiring quotas," *Working paper*, 2024. [25]

Tombe, Trevor and Xiaodong Zhu, "Trade, Migration and Productivity: A Quantitative Analysis of China," *American Economic Review*, 2019, *forthcoming.* [4]

Traiberman, Sharon, "Occupations and Import Competition," *American Economic Review*, 2019, *109* (12), 4260–4301. [5, 7, 8, 24]

van der Klauuw, Wilbert, "On the Use of Expectations Data in Estimating Structural Dynamic Choice Models," *Journal of Labor Economics*, 2012, *30* (3), 554–571. [5]

Wilson, Riley, "Moving to Jobs: The Role of Information in Migration Decisions," *Journal of Labor Economics*, 2021, *39* (4), 1083–1128. [5]

Online Appendix for "Measuring Information Frictions in Migration Decisions: A Revealed-Preference Approach"

Thomas Fujiwara
Princeton University

Eduardo Morales
Princeton University

Charly Porcher
Georgetown University

March, 2023

A Model with Locations and Sectors

Time is discrete and indexed by t. Workers indexed by i belong to different demographic types indexed by d. We model workers' choice of the market to supply labor at each period t. Each labor market is defined by a sector $s = 1, \ldots, S$ and a location $l = 1, \ldots, L$, and we index each market by the combination of indices sl. A worker's choice of market at t may depend on the worker's history, i.e., the sequence of sectors and locations of employment up until period $t-1$. We index histories by h.

A.1 Static Utility

We index a worker's prior sector and location by $r = 1, \ldots, S$ and $n = 1, \ldots, L$, respectively, and index the prior market by the combination of indices rn. Thus, the static utility of choosing market sl at period t for a worker i of demographic group d that has a labor history h consistent with a prior market rn is

$$\mathcal{U}^{sl}_{hit} = u^{sl}_{hit} + \varepsilon^{sl}_{i}, \tag{A.1a}$$

$$u^{sl}_{hit} = \kappa^{l}_{ndt} + \kappa^{s}_{rdt} + w^{sl}_{hit}, \tag{A.1b}$$

$$w^{sl}_{hit} = w^{sl}_{dt} + \alpha^{s}_{dt} + \alpha^{s}_{i} + \beta^{s}_{d} x^{s}_{hit} + \nu^{sl}_{i}, \tag{A.1c}$$

with $x^{s}_{hit} = x^{s}_{h'it}$ if h and h' imply the same history of sectors up to t. More specifically, in our estimation, we impose

$$\beta^{s}_{d} x^{s}_{hit} = \beta^{s}_{de} exp^{s}_{ht} + \beta^{s}_{dee} (exp^{s}_{ht})^2 + \beta^{s}_{da} age_i + \beta^{s}_{daa} age^2_i,$$

where exp^{s}_{ht} denotes experience in sector s at period t implied by history h.

A.2 Optimal Labor Market Choice

We allow worker i's labor market choice at any period t to depend on her prior history of employment, i.e., the complete sequence of sectors and locations of employment of worker i since the period she joined the labor market up until period $t-1$. Denoting the optimal choice of a worker i at period t with previous history h as sl_{hit}, we assume that

$$sl_{hit} = \underset{sl \in \mathcal{B}_{hit}}{\text{argmax}} \, \mathbb{E}[\mathcal{U}^{sl}_{hit} | \mathcal{J}_{hit}]. \tag{A.2}$$

If s is the sector of employment at period $t-1$ of a worker i, then

$$\{sl\}_l \in \mathcal{B}_{hit}.$$

A.3 Distributional Assumptions

First, we assume that every worker i knows at period t the value of the idiosyncratic preference shocks $\varepsilon_i \equiv \{\varepsilon_i^{sl}\}_{sl}$; i.e.,

$$\mathcal{J}_{hit} = (\mathcal{W}_{hit}, \varepsilon_i). \tag{A.3}$$

Furthermore, we assume the shocks have a distribution with the following cumulative distribution function:

$$F(\varepsilon_i | \{\mathcal{W}_{hit}\}_i) = \exp\left(- \sum_{s=1}^{S} \left(\sum_{l=1}^{L} \exp(-\theta_d \varepsilon_i^{sl}) \right)^{\lambda_d} \right), \tag{A.4}$$

and are independent across workers, i.e.,

$$\varepsilon_i \perp \varepsilon_{i't} \qquad \text{for any } i \neq i'. \tag{A.5}$$

Second, we assume worker i ignores the value of the idiosyncratic wage shocks $\nu_i \equiv \{\nu_i^{sl}\}_{sl}$ when choosing their preferred market at period t and, additionally, we assume that these are mean-independent of the other terms in the wage equation. Formally,

$$\mathbb{E}[\nu_i^{sl} | \mathcal{J}_{hit}, w_{dt}^{sl}, \alpha_{dt}^s, \alpha_i^s, x_{hit}^s] = 0. \tag{A.6}$$

Third, defining $w_{dt}^s \equiv \{w_{dt}^{sl}\}_l$ and $\kappa_d \equiv \{\kappa_{ndt}^l\}_{nlt}$, we assume that, for any two workers i and i' that share demographic group and location at period $t-1$, it holds that

$$\mathbb{E}[w_{dt}^s, \kappa_d | \mathcal{J}_{hit}, \mathcal{J}_{h'i't}] = \mathbb{E}[w_{dt}^s, \kappa_d | \mathcal{J}_{hit}] \tag{A.7}$$

if i's sector at $t-1$ is also s. Additionally, we assume that

$$\mathbb{E}[w_{dt}^s, \kappa_d | \mathcal{J}_{hit}] = \mathbb{E}[w_{dt}^s, \kappa_d | \mathcal{W}_{hit}]. \tag{A.8}$$

A.4 Normalization

For every origin n, demographic group d, and period t, the parameters $\{\kappa_{ndt}^l\}_l$ are identified only up to an additive constant. We thus impose the following normalization

$$\kappa_{ndt}^n = 0 \qquad \text{for all } n, d, \text{ and } t. \tag{A.9}$$

B Proofs and Additional Derivations

In Appendix B.1, we prove Theorem 1. In Appendix B.2, we present additional derivations showing how we obtain several expressions appearing in Section 3.1.1. In Appendix B.4, we prove Theorem 2. In Appendix B.5, we present additional derivations showing how we obtain several expressions appearing in Section 3.1.2. In Appendix B.7, we prove Theorem 3. In Appendix B.8, we present additional derivations showing how we obtain several expressions appearing in Section 3.2. In Appendix B.9, we prove Corollary 3.

B.1 Second-Step Bounding Moment Inequalities: Proof of Theorem 1

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

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

Equation (3) implies we can rewrite this equality as

$$(y_{is}^{l} + y_{is}^{l'})(\mathbb{1}\{\mathbb{E}[\Delta\kappa^{ll'} + \alpha\Delta w_{is}^{ll'} + \Delta\varepsilon_{is}^{ll'}|\mathcal{J}_{is}] \geqslant 0\} - y_{is}^{l}) = 0,$$

where $\Delta\kappa^{ll'} = \kappa^{l} - \kappa^{l'}$, $\Delta w_{is}^{ll'} = w_{is}^{l} - w_{is}^{l'}$, and $\Delta\varepsilon_{is}^{ll'} = \varepsilon_{is}^{l} - \varepsilon_{is}^{l'}$. Equation (4) further implies that

$$(y_{is}^{l} + y_{is}^{l'})(\mathbb{1}\{\Delta\kappa^{ll'} + \alpha\mathbb{E}[\Delta w_{is}^{ll'}|\mathcal{J}_{is}] + \Delta\varepsilon_{is}^{ll'} \geqslant 0\} - y_{is}^{l}) = 0,$$

and, given equation (5), we can simplify

$$(y_{is}^{l} + y_{is}^{l'})(\mathbb{1}\{\Delta\kappa^{ll'} + \alpha\mathbb{E}[\Delta w_{is}^{ll'}|\mathcal{W}_{is}] + \Delta\varepsilon_{is}^{ll'} \geqslant 0\} - y_{is}^{l}) = 0. \tag{B.1}$$

This equality holds for any worker i of any type s, 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 $\mathcal{W}_{is}$ and a dummy variable that equals one if worker i of type s chooses either location l or location l'; i.e.,

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

Equation (6) implies we can rewrite this moment equality as

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

or, equivalently,

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

Multiplying by $1 + \exp(-\Delta\kappa^{ll'} - \alpha\mathbb{E}[\Delta w_{s}^{ll'}|\mathcal{W}_{is}])$, we obtain

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

or, equivalently,

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

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

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

As $-\exp(-x)$ is concave in $x \in \mathbb{R}$, a linear approximation to this function at any value $a \in \mathbb{R}$ will bound it from above. The formula for this linear approximation is

$$-\exp(-a) + \exp(-a)(x-a) = -\exp(-a)(1+a) + \exp(-a)x = \exp(-a)(-(1+a)+x).$$

Therefore, given any deterministic function $h^{ll'}_{is}\colon \mathcal{Z}_s \times \Theta_{ll'} \to \mathbb{R}$ and the equality in equation (B.2), we can derive the following inequality

$$\mathbb{E}[y^{l'}_{is} + y^{l}_{is}\exp(-h^{ll'}_{is}(z_s,\Delta\kappa^{ll'}))(-(1+h^{ll'}_{is}(z_s,\Delta\kappa^{ll'})) + \Delta\kappa^{ll'} + \alpha\mathbb{E}[\Delta w^{ll'}_{is}|\mathcal{W}_{is}])|\mathcal{W}_{is}] \geqslant 0. \tag{B.3}$$

Defining as $\nu^{ll'}_{is}$ the expectational error worker i of type s makes when forecasting the wage difference Δw^{ll}_{s}, equation (2) implies

$$\nu^{ll'}_{is} \equiv \Delta w^{ll'}_{is} - \mathbb{E}[\Delta w^{ll'}_{s}|\mathcal{W}_{is}], \qquad \Longrightarrow \qquad \mathbb{E}[\nu^{ll'}_{is}|\mathcal{W}_{is}] = 0. \tag{B.4}$$

Let's consider the moment

$$\mathbb{E}[y^{l'}_{is} + y^{l}_{is}\exp(-h^{ll'}_{is}(z_s,\Delta\kappa^{ll'}))(-(1+h^{ll'}_{is}(z_s,\Delta\kappa^{ll'})) + \Delta\kappa^{ll'} + \alpha\Delta w^{ll'}_{is})|\mathcal{W}_{is}]. \tag{B.5}$$

or, equivalently,

$$\mathbb{E}[y^{l'}_{is} + y^{l}_{is}\exp(-h^{ll'}_{is}(z_s,\Delta\kappa^{ll'}))(-(1+h^{ll'}_{is}(z_s,\Delta\kappa^{ll'})) + \Delta\kappa^{ll'} + \alpha(\mathbb{E}[\Delta w^{ll'}_{is}|\mathcal{W}_{is}] + \nu^{ll'}_{is}))|\mathcal{W}_{is}].$$

Given that $\mathcal{W}_{is} \subset \mathcal{J}_{is}$, we can use the Law of Iterated Expectations to rewrite this moment as

$$\mathbb{E}[\mathbb{E}[y^{l'}_{is} + y^{l}_{is}\exp(-h^{ll'}_{is}(z_s,\Delta\kappa^{ll'}))(-(1+h^{ll'}_{is}(z_s,\Delta\kappa^{ll'})) + \Delta\kappa^{ll'} + \alpha(\mathbb{E}[\Delta w^{ll'}_{is}|\mathcal{W}_{is}] + \nu^{ll'}_{is}))|\mathcal{J}_{is}]|\mathcal{W}_{is}].$$

Equation (1) implies $\mathbb{E}[y^{l}_{is}|\mathcal{J}_{is}] = y^{l}_{is}$, and equation (4) implies $(\Delta\kappa^{ll'},\alpha) \subset \mathcal{J}_{is}$. Consequently, if $z_s \subseteq \mathcal{W}_{is}$, it is then the case that $z_s \subset \mathcal{J}_{is}$ and we can further rewrite

$$\mathbb{E}[y^{l'}_{is} + y^{l}_{is}\exp(-h^{ll'}_{is}(z_s,\Delta\kappa^{ll'}))(-(1+h^{ll'}_{is}(z_s,\Delta\kappa^{ll'})) + \Delta\kappa^{ll'} + \alpha\mathbb{E}[\mathbb{E}[\Delta w^{ll'}_{is}|\mathcal{W}_{is}] + \nu^{ll'}_{is}|\mathcal{J}_{is}])|\mathcal{W}_{is}].$$

As $\Delta w^{ll'}_{is} = \mathbb{E}[\Delta w^{ll'}_{is}|\mathcal{W}_{is}] + \nu^{ll'}_{is}$, equation (5) further implies that

$$\mathbb{E}[y^{l'}_{is} + y^{l}_{is}\exp(-h^{ll'}_{is}(z_s,\Delta\kappa^{ll'}))(-(1+h^{ll'}_{is}(z_s,\Delta\kappa^{ll'})) + \Delta\kappa^{ll'} + \alpha\mathbb{E}[\mathbb{E}[\Delta w^{ll'}_{s}|\mathcal{W}_{is}] + \nu^{ll'}_{is}|\mathcal{W}_{is}])|\mathcal{W}_{is}],$$

and equation (B.4) implies we can rewrite the moment in equation (B.5) as

$$\mathbb{E}[y^{l'}_{is} + y^{l}_{is}\exp(-h^{ll'}_{is}(z_s,\Delta\kappa^{ll'}))(-(1+h^{ll'}_{is}(z_s,\Delta\kappa^{ll'})) + \Delta\kappa^{ll'} + \alpha\mathbb{E}[\Delta w^{ll'}_{is}|\mathcal{W}_{is}])|\mathcal{W}_{is}].$$

However, this moment is exactly the same entering the moment inequality in equation (B.3), which implies that the following inequality involving the moment in equation (B.5) is equivalent to that in equation (B.3):

$$\mathbb{E}[y^{l'}_{is} + y^{l}_{is}\exp(-h^{ll'}_{is}(z_s,\Delta\kappa^{ll'}))(-(1+h^{ll'}_{is}(z_s,\Delta\kappa^{ll'})) + \Delta\kappa^{ll'} + \alpha\Delta w^{ll'}_{is})|\mathcal{W}_{is}] \geqslant 0.$$

Finally, if $z_s \subset \mathcal{W}_{is}$, we can use the Law of Iterated Expectations and conclude that

$$\mathbb{E}[y^{l'}_{is} + y^{l}_{is}\exp(-h^{ll'}_{is}(z_s,\Delta\kappa^{ll'}))(-(1+h^{ll'}_{is}(z_s,\Delta\kappa^{ll'})) + \Delta\kappa^{ll'} + \alpha\Delta w^{ll'}_{is})|z_s] \geqslant 0. \tag{B.6}$$

The moment in this inequality is precisely the moment in equation (8) when evaluated at $\Delta\theta_{ll'} = \Delta\kappa^{ll'}$. Thus, equation (B.6) implies Theorem 1. ■

B.2 Second-Step Bounding Moment Inequalities: Additional Derivations

Derivation of equation (10). Given $z_s \in \mathcal{Z}_s$, we compute the function in equation (10) by finding the value of $h^{ll'}_{is}(z_s,\Delta\theta_{ll})$ that minimizes the moment in equation (8) at each value of $\Delta\theta_{ll'}$. Specifically, given z_s and

$\Delta\theta_{ll'}$ the first-order condition of the moment in equation (8) with respect to the scalar $h_{is}^{ll'}(z_s, \Delta\theta_{ll})$ is

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

or, equivalently,

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

Setting this moment condition to zero and bearing in mind that, according to equation (4), it holds that $\alpha \subset \mathcal{J}_{is}$, we can solve for $h_{is}^{ll'}(z_s, \Delta\theta_{ll'})$ to obtain the solution in equation (10); i.e.,

$$h_{is}^{ll'}(z_s, \Delta\theta_{ll'}) = \Delta\theta_{ll'} + \alpha\mathbb{E}[\Delta w_s^{ll'}|z_s, y_{is}^l = 1]. \tag{B.7}$$

Derivation of equation (11). Equations (8) to (10) imply the following inequality

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

or, equivalently,

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

We can simplify this inequality as

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

or, equivalently,

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

We can further rewrite this inequality as

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

or equivalently,

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

Eliminating terms that cancel each other, we obtain the inequality

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

Rearranging terms, we obtain the expression in equation (11); i.e.,

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

Derivation of equation (12). Swapping the indices l and l' in equation (11) we obtain the following inequality

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

Rearranging terms, we immediately obtain the inequality in equation (12); i.e.,

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

B.3 Second-Step Bounding Inequalities: Proof of Corollary 1

Equations (8) to (10) imply the following moment inequality:

$$\mathbb{E}[y_{is}^{l}\exp(-(\Delta\theta_{ll'} + \alpha\mathbb{E}[\Delta w_s^{ll'}|z_s, y_s^{l} = 1])) - y_{is}^{l'}|z_s] \geqslant 0. \tag{B.8}$$

As Corollary 1 assumes $z_s \subset \mathcal{J}_{is}$, the definition $\mathcal{W}_{is} \equiv \mathcal{J}_{is}/\varepsilon_{is}$ implies $z_s \subseteq \mathcal{W}_{is}$. Using the LIE, we write

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

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

$$\mathbb{E}[\mathbb{E}[y_{is}^{l}|\mathcal{W}_{is}]\exp(-(\Delta\theta_{ll'} + \alpha\mathbb{E}[\Delta w_s^{ll'}|z_s, y_s^{l} = 1])) - \mathbb{E}[y_{is}^{l'}|\mathcal{W}_{is}]|z_s] \geqslant 0. \tag{B.9}$$

As y_s^{l} is a function of $(\mathcal{W}_{is}, \varepsilon_i)$, equation (5) implies that

$$\mathbb{E}[\Delta w_s^{ll'}|z_s, \mathcal{W}_{is}, y_s^{l}] = \mathbb{E}[\Delta w_s^{ll'}|z_s, \mathcal{W}_{is}].$$

Given that $\mathbb{E}[\Delta w_s^{ll'}|z_s] = \mathbb{E}[\Delta w_s^{ll'}|\mathcal{J}_{is}]$ according to Corollary 1, we can rewrite equation (B.9) as

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

Given equation (7), we can further rewrite

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

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

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

where we have used the assumption (imposed in Corollary 1) that $\mathbb{E}[\Delta w_s^{ll'}|z_s] = \mathbb{E}[\Delta w_s^{ll'}|\mathcal{W}_{is}]$. We can rewrite this inequality as

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

The expectation in this inequality is always strictly positive. Thus, we can rewrite the inequality as

$$\exp(\Delta\kappa_{ll'} - \Delta\theta_{ll'}) - 1 \geqslant 0.$$

or, equivalently,

$$\Delta\kappa_{ll'} \geqslant \Delta\theta_{ll'}. \tag{B.10}$$

This inequality holds for any two locations l and l'. Thus, swapping the location indices l and l', we similarly obtain the following inequality:

$$\Delta\kappa_{l'l} \geqslant \Delta\theta_{l'l},$$

which can be rewritten as

$$\Delta\kappa_{ll'} \leqslant \Delta\theta_{ll'}, \tag{B.11}$$

Combining the inequalities in equations (B.10) and (B.11), we obtain the following equality:

$$\Delta\kappa_{ll'} = \Delta\theta_{ll'}. \tag{B.12}$$

Thus, Corollary 1 holds. ■

B.4 Second-Step Odds-Based Moment Inequalities: Proof of Theorem 2

To prove Theorem 2, we follow the approach in Dickstein et al. (2023), adapting their notation to our particular application. The point of departure of our proof is equation (B.2). Let's consider the moment

$$\mathbb{E}[y_{is}^{l'} + y_{is}^{l}(-\exp(-(\Delta\kappa^{ll'} + \alpha\Delta w_{is}^{ll'})))|\mathcal{W}_{is}] \tag{B.13}$$

or, equivalently,

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

where $\nu_{is}^{ll'}$ is defined in equation (B.4). Given that $\mathcal{W}_{is} \subset \mathcal{J}_{is}$, we can use the Law of Iterated Expectations to rewrite this moment as

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

Equation (1) implies $\mathbb{E}[y_{is}^{l}|\mathcal{J}_{is}] = y_{is}^{l}$. Consequently, we can further rewrite

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

Equation (4) and the definition of $\mathcal{W}_{is}$ as including every element of $\mathcal{J}_{is}$ other than ε_{is} implies that $(\Delta\kappa^{ll'}, \alpha) \subset \mathcal{W}_{is}$. As $\Delta w_{is}^{ll'} = \mathbb{E}[\Delta w_{is}^{ll'}|\mathcal{W}_{is}] + \nu_{is}^{ll'}$, equation (5) further implies that $\mathbb{E}[\mathbb{E}[\Delta w_{is}^{ll'}|\mathcal{W}_{is}] + \nu_{is}^{ll'}|\mathcal{J}_{is}] = \mathbb{E}[\mathbb{E}[\Delta w_{is}^{ll'}|\mathcal{W}_{is}] + \nu_{is}^{ll'}|\mathcal{W}_{is}]$. Thus, we can rewrite the moment above as

$$\mathbb{E}[y_{is}^{l'} + y_{is}^{l}\mathbb{E}[-\exp(-(\Delta\kappa^{ll'} + \alpha(\mathbb{E}[\Delta w_{is}^{ll'}|\mathcal{W}_{is}] + \nu_{is}^{ll'})))|\mathcal{W}_{is}]|\mathcal{W}_{is}]. \tag{B.14}$$

As $-\exp(x)$ is concave in $x \in \mathbb{R}$, equation (B.4) and Jensen's inequality imply the following inequality

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

The right-hand side of this inequality coincides with equation (B.2) and, thus, we can conclude that

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

As the moments in equations (B.13) and (B.14) are equivalent, we can further rewrite this inequality as

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

Multiplying by -1 on both sides of this equation, we obtain the following inequality

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

Finally, if $z_s \subset \mathcal{W}_{is}$, we can use the Law of Iterated Expectations to conclude

$$\mathbb{E}[y_{is}^{l} \exp(-(\Delta\kappa^{ll'} + \alpha\Delta w_{is}^{ll'})) - y_{is}^{l'} | z_s] \geqslant 0. \tag{B.15}$$

The moment in this inequality is precisely the moment in equation (13) when evaluated at $\Delta\theta_{ll'} = \Delta\kappa^{ll'}$. Thus, equation (B.15) implies Theorem 2. ■

B.5 Second-Step Odds-Based Moment Inequalities: Additional Derivations

Derivation of equation (15)*.* Equation (13) and (14) imply the following inequality

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

We can rewrite this inequality as

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

or, equivalently,

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

Rearranging terms, we obtain the expression in equation (15); i.e.,

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

Derivation of equation (16)*.* Swapping the indices l and l' in equation (15), we obtain the following inequality

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

Rearranging terms, we immediately obtain the inequality in equation (16); i.e.,

$$\frac{\mathbb{E}[y_{is}^{l} | z_s]}{\mathbb{E}[y_{is}^{l'} | z_s]} (\mathbb{E}[\exp(-\alpha\Delta w_{is}^{l'l}) | z_s, y_{is}^{l'} = 1])^{-1} \leqslant \exp(\Delta\theta_{ll'}). \tag{B.16}$$

B.6 Second-Step Odds-Based Moment Inequalities: Proof of Corollary 2

Equations (13) and (14) imply the following moment inequality:

$$\mathbb{E}[y_{is}^{l} \exp(-(\Delta\theta_{ll'} + \alpha\Delta w_{s}^{ll'})) - y_{is}^{l'} | z_s] \geqslant 0. \tag{B.17}$$

The assumption that $z_s \subset \mathcal{J}_{is}$ in Corollary 2, and the definition of $\mathcal{W}_{is} \equiv \mathcal{J}_{is}/\varepsilon_{is}$, imply that $z_s \subseteq \mathcal{W}_{is}$. Using the LIE, we can then write

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

Given the assumption that $\Delta w_{s}^{ll'} = \mathbb{E}[\Delta w_{s}^{ll'} | \mathcal{W}_{is}]$ in Corollary 2, we can rewrite

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

or, equivalently,

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

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

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

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

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

or, equivalently,

$$(\exp(\Delta\kappa_{ll'} - \Delta\theta_{ll'}) - 1)\mathbb{E}[(\sum_{l''=1}^{L}\exp(\Delta\kappa_{l''l'} + \alpha\mathbb{E}[\Delta w_s^{l''l'}|\mathcal{W}_{is}]))^{-1}|z_s] \geqslant 0. \tag{B.18}$$

The conditional expectation in this inequality will always be strictly positive. Thus, we can rewrite the inequality in equation (B.18) as

$$\exp(\Delta\kappa_{ll'} - \Delta\theta_{ll'}) - 1 \geqslant 0.$$

or, equivalently,

$$\Delta\kappa_{ll'} \geqslant \Delta\theta_{ll'}. \tag{B.19}$$

This inequality holds for any two locations l and l'. Thus, swapping the location indices l and l', we similarly obtain the following inequality:

$$\Delta\kappa_{l'l} \geqslant \Delta\theta_{l'l},$$

which can be rewritten as

$$\Delta\kappa_{ll'} \leqslant \Delta\theta_{ll'}, \tag{B.20}$$

Combining the inequalities in equations (B.19) and (B.20), we obtain the following equality:

$$\Delta\kappa_{l'l} = \Delta\theta_{l'l}. \tag{B.21}$$

Thus, Corollary 2 holds. ■

B.7 First-Step Moment Inequalities: Proof of Theorem 3

For any locations l and l' and any worker i of type s, equation (B.1) indicates the following equality holds

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

For any locations l and l', any worker i of type s, and any worker j of type r, we can thus derive the following equality:

$$y_{jr}^{l'}(y_{is}^{l} + y_{is}^{l'})(\mathbb{1}\{\Delta\kappa^{ll'} + \alpha\mathbb{E}[\Delta w_{is}^{ll'}|\mathcal{W}_{is}] + \Delta\varepsilon_{is}^{ll'} \geqslant 0\} - y_{is}^{l}) = 0. \tag{B.22}$$

As the next step, we take the expectation of both sides of this equality conditional on $\mathcal{W}_{is}$, on $\mathcal{W}_{jr}$, and on a dummy variable that equals one if worker i of type s chooses either location l or location l'; i.e.,

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

Given equations (1) and (6), we can rewrite this moment equality as

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

or, equivalently,

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

Multiplying by $1+\exp(-\Delta\kappa^{ll'}-\alpha\mathbb{E}[\Delta w_{s}^{ll'}|\mathcal{W}_{is}])$, we obtain

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

or, equivalently,

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

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

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

or, equivalently,

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

As the function $-\exp(-x)$ is concave in $x\in\mathbb{R}$, given any deterministic function $g_{ijsr}^{ll'}\colon\mathcal{Z}_s\times\mathcal{Z}_r\times\Theta_\alpha\to\mathbb{R}$, we can derive the following inequality

$$\begin{aligned}&\mathbb{E}[y_{is}^{l'}y_{jr}^{l'}+y_{is}^{l}y_{jr}^{l'}\exp(-g_{ijsr}^{ll'}(z_s,z_r,\alpha))\times\\&(-(1+g_{ijsr}^{ll'}(z_s,z_r,\alpha))+\Delta\kappa^{ll'}+\alpha\mathbb{E}[\Delta w_{is}^{ll'}|\mathcal{W}_{is}])|\mathcal{W}_{is},\mathcal{W}_{jr}]\geqslant 0.\end{aligned}\tag{B.23}$$

Let's consider the moment

$$\mathbb{E}[y_{is}^{l'}y_{jr}^{l'}+y_{is}^{l}y_{jr}^{l'}\exp(-g_{ijsr}^{ll'}(z_s,z_r,\alpha))(-(1+g_{ijsr}^{ll'}(z_s,z_r,\alpha))+\Delta\kappa^{ll'}+\alpha\Delta w_{is}^{ll'})|\mathcal{W}_{is},\mathcal{W}_{jr}],\tag{B.24}$$

or, equivalently,

$$\mathbb{E}[y_{is}^{l'}y_{jr}^{l'}+y_{is}^{l}y_{jr}^{l'}\exp(-g_{ijsr}^{ll'}(z_s,z_r,\alpha))(-(1+g_{ijsr}^{ll'}(z_s,z_r,\alpha))+\Delta\kappa^{ll'}+\alpha(\mathbb{E}[\Delta w_{is}^{ll'}|\mathcal{W}_{is}]+\nu_{is}^{ll'}))|\mathcal{W}_{is},\mathcal{W}_{jr}],$$

where, as a reminder, $\nu_{is}^{ll'}\equiv\Delta w_{is}^{ll'}-\mathbb{E}[\Delta w_{s}^{ll'}|\mathcal{W}_{is}]$. Equation (5) implies we can also write $\nu_{is}^{ll'}\equiv\Delta w_{is}^{ll'}-\mathbb{E}[\Delta w_{s}^{ll'}|\mathcal{W}_{is},\mathcal{W}_{jr}]$ and, consequently, we can conclude that

$$\mathbb{E}[\nu_{is}^{ll'}|\mathcal{W}_{is},\mathcal{W}_{jr}]=0.\tag{B.25}$$

Given that $\mathcal{W}_{is}\subset\mathcal{J}_{is}$ and $\mathcal{W}_{jr}\subset\mathcal{J}_{jr}$, we can use the Law of Iterated Expectations to rewrite the moment inequality above as

$$\begin{aligned}&\mathbb{E}[\mathbb{E}[y_{is}^{l'}y_{jr}^{l'}+y_{is}^{l}y_{jr}^{l'}\exp(-g_{ijsr}^{ll'}(z_s,z_r,\alpha))\times\\&(-(1+g_{ijsr}^{ll'}(z_s,z_r,\alpha))+\Delta\kappa^{ll'}+\alpha(\mathbb{E}[\Delta w_{is}^{ll'}|\mathcal{W}_{is}]+\nu_{is}^{ll'}))|\mathcal{J}_{is},\mathcal{J}_{jr}]|\mathcal{W}_{is},\mathcal{W}_{jr}].\end{aligned}$$

Equation (1) implies $\mathbb{E}[y_{is}^{l}y_{jr}^{l'}|\mathcal{J}_{is},\mathcal{J}_{jr}]=y_{is}^{l}y_{jr}^{l'}$, and equation (4) implies $(\Delta\kappa^{ll'},\alpha)\subset(\mathcal{J}_{is}\cap\mathcal{J}_{jr})$. Consequently, if $z_s\subseteq\mathcal{W}_{is}$ and $z_r\subseteq\mathcal{W}_{jr}$, it is then the case that $z_s\subset\mathcal{J}_{is}$ and $z_r\subset\mathcal{J}_{jr}$, and we can thus further rewrite

$$\mathbb{E}[y_{is}^{l'}y_{jr}^{l'}+y_{is}^{l}y_{jr}^{l'}\exp(-g_{ijsr}^{ll'}(z_s,z_r,\alpha))\times$$

$$(-(1+g_{ijsr}^{ll'}(z_s,z_r,\alpha))+\Delta\kappa^{ll'}+\alpha\mathbb{E}[\mathbb{E}[\Delta w_{is}^{ll'}|\mathcal{W}_{is}]+\nu_{is}^{ll'}|\mathcal{J}_{is},\mathcal{J}_{jr}])|\mathcal{W}_{is},\mathcal{W}_{jr}].$$

As $\Delta w_{is}^{ll'}=\mathbb{E}[\Delta w_{is}^{ll'}|\mathcal{W}_{is}]+\nu_{is}^{ll'}$, equation (5) further implies that

$$\mathbb{E}[y_{is}^{l'}y_{jr}^{l'}+y_{is}^{l}y_{jr}^{l'}\exp(-g_{ijsr}^{ll'}(z_s,z_r,\alpha))\times$$
$$(-(1+g_{ijsr}^{ll'}(z_s,z_r,\alpha))+\Delta\kappa^{ll'}+\alpha\mathbb{E}[\mathbb{E}[\Delta w_{is}^{ll'}|\mathcal{W}_{is}]+\nu_{is}^{ll'}|\mathcal{W}_{is},\mathcal{W}_{jr}])|\mathcal{W}_{is},\mathcal{W}_{jr}],$$

and equation (B.25) implies we can rewrite this moment as

$$\mathbb{E}[y_{is}^{l'}y_{jr}^{l'}+y_{is}^{l}y_{jr}^{l'}\exp(-g_{ijsr}^{ll'}(z_s,z_r,\alpha))(-(1+g_{ijsr}^{ll'}(z_s,z_r,\alpha))+\Delta\kappa^{ll'}+\alpha\mathbb{E}[\Delta w_{is}^{ll'}|\mathcal{W}_{is}])|\mathcal{W}_{is},\mathcal{W}_{jr}].$$

However, this moment is exactly the same entering the moment inequality in equation (B.23), which implies that the following inequality involving the moment in equation (B.24) is equivalent to that in equation (B.23):

$$\mathbb{E}[y_{is}^{l'}y_{jr}^{l'}+y_{is}^{l}y_{jr}^{l'}\exp(-g_{ijsr}^{ll'}(z_s,z_r,\alpha))(-(1+g_{ijsr}^{ll'}(z_s,z_r,\alpha))+\Delta\kappa^{ll'}+\alpha\Delta w_{is}^{ll'})|\mathcal{W}_{is},\mathcal{W}_{jr}]\geqslant 0. \quad \text{(B.26)}$$

This moment inequality is one of the two moment inequalities we will combine to obtain the inequality that we use to bound the parameter α. To obtain the second moment inequality, we start from the following expression

$$y_{is}^{l}(y_{jr}^{l}+y_{jr}^{l'})(\mathbb{1}\{\Delta\kappa^{l'l}+\alpha\mathbb{E}[\Delta w_{jr}^{l'l}|\mathcal{W}_{jr}]+\Delta\varepsilon_{jr}^{l'l}\geqslant 0\}-y_{jr}^{l'})=0, \quad \text{(B.27)}$$

which is analogous to that in equation (B.22). Following the same steps described above to go from equation (B.22) to equation (B.26), we can derive from equation (B.27) the following inequality

$$\mathbb{E}[y_{is}^{l}y_{jr}^{l}+y_{is}^{l}y_{jr}^{l'}\exp(-g_{ijsr}^{ll'}(z_s,z_r,\alpha))(-(1+g_{ijsr}^{ll'}(z_s,z_r,\alpha))+\Delta\kappa^{l'l}+\alpha\Delta w_{jr}^{l'l})|\mathcal{W}_{is},\mathcal{W}_{jr}]\geqslant 0. \quad \text{(B.28)}$$

As the moments in equations (B.26) and (B.28) have the same conditioning set, we can add them, obtaining the following moment inequality:

$$\mathbb{E}[y_{is}^{l}y_{jr}^{l}+y_{is}^{l'}y_{jr}^{l'}+y_{is}^{l}y_{jr}^{l'}\exp(-g_{ijsr}^{ll'}(z_s,z_r,\alpha))(-(1+g_{ijsr}^{ll'}(z_s,z_r,\alpha))+\alpha(\Delta w_{is}^{ll'}+\Delta w_{is}^{l'l}))|\mathcal{W}_{is},\mathcal{W}_{jr}]\geqslant 0.$$

Finally, if $z_s\subset\mathcal{W}_{is}$ and $z_r\subset\mathcal{W}_{jr}$, we can use the Law of Iterated Expectations and conclude that

$$\mathbb{E}[y_{is}^{l}y_{jr}^{l}+y_{is}^{l'}y_{jr}^{l'}+y_{is}^{l}y_{jr}^{l'}\exp(-g_{ijsr}^{ll'}(z_s,z_r,\alpha))\times$$
$$(-(1+g_{ijsr}^{ll'}(z_s,z_r,\alpha))+\alpha(\Delta w_{is}^{ll'}+\Delta w_{is}^{l'l}))|z_s,z_r]\geqslant 0. \quad \text{(B.29)}$$

The moment in this inequality is precisely the moment in equation (17) when evaluated at $\theta_\alpha=\alpha$. Thus, equation (B.29) implies Theorem 3. ∎

B.8 First-Step Moment Inequalities: Additional Derivations

Derivation of equation (19). Given values $z_s\in\mathcal{Z}_s$ and $z_r\in\mathcal{Z}_r$, we compute the function in equation (19) by finding the value of $g_{ijsr}^{ll'}(z_s,z_r,\theta_\alpha)$ that minimizes the moment in equation (17) at each value of θ_α. Specifically, given z_s, z_r, and θ_α, the first-order condition of the moment in equation (17) with respect to the scalar $g_{ijsr}^{ll'}(z_s,z_r,\theta_\alpha)$ is

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

or, equivalently,

$$\mathbb{E}[2g_{ijsr}^{ll'}(z_s,z_r,\theta_\alpha)-\theta_\alpha(\Delta w_s^{ll'}+\Delta w_r^{l'l})|z_s,z_r,y_{is}^{l}y_{is'}^{l'}=1].$$

Setting this moment condition to zero, we can solve for $g_{ijsr}^{ll'}(z_s, z_r, \theta_\alpha)$ to obtain the solution in equation (19); i.e.,

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

Derivation of equation (20). Equations (17) to (19) imply the following inequality

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

or, equivalently,

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

Using the Law of Iterated Expectations, we can rewrite this inequality as

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

Simplifying this expression, we obtain

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

Rearranging terms, we obtain the expression in equation (20); i.e.,

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

B.9 First-Step Moment Inequalities: Proof of Corollary 3

If $z_s \subset \mathcal{W}_{is}$ and $z_r \subset \mathcal{W}_{jr}$, we can use the Law of Iterated expectations to rewrite equation (20) as

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

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

$$\frac{\mathbb{E}[\mathbb{E}[y_{is}^l|\mathcal{W}_{is}]\mathbb{E}[y_{jr}^{l'}|\mathcal{W}_{jr}]|z_s, z_r]}{\mathbb{E}[0.5(\mathbb{E}[y_{is}^l|\mathcal{W}_{is}]\mathbb{E}[y_{jr}^l|\mathcal{W}_{jr}] + \mathbb{E}[y_{is}^{l'}|\mathcal{W}_{is}]\mathbb{E}[y_{jr}^{l'}|\mathcal{W}_{jr}])|z_s, z_r]} \leqslant \\ \exp(\theta_\alpha \mathbb{E}[0.5(\Delta w_s^{ll'} + \Delta w_r^{l'l})|z_s, z_r, y_{is}^l y_{jr}^{l'} = 1]). \tag{B.30}$$

Given equations (1) to (6), it holds that, for any $l_1 = 1, \ldots, L$ and $l_2 = 1, \ldots, L$, 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 can then rewrite the inequality in equation (B.30) as

$$\frac{\mathbb{E}[\exp(\Delta\kappa^{ll'} + \alpha\mathbb{E}[\Delta w_s^{ll'}|\mathcal{W}_{is}]) \exp(\Delta\kappa^{l'l} + \alpha\mathbb{E}[\Delta w_r^{l'l}|\mathcal{W}_{jr}])|z_s, z_r]}{\mathbb{E}[0.5(\exp(\Delta\kappa^{ll'} + \alpha\mathbb{E}[\Delta w_s^{ll'}|\mathcal{W}_{is}]) + \exp(\Delta\kappa^{l'l} + \alpha\mathbb{E}[\Delta w_r^{l'l}|\mathcal{W}_{jr}]))|z_s, z_r]}$$

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

Simplifying this expression, we obtain

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

If $\mathbb{E}[\Delta w_s^{ll'}|z_s] = \mathbb{E}[\Delta w_s^{ll'}|\mathcal{J}_{is}] = \Delta\bar{w}$, and $\mathbb{E}[\Delta w_r^{l'l}|z_r] = \mathbb{E}[\Delta w_r^{l'l}|\mathcal{J}_{ir}] = \Delta\bar{w}$, for a common constant $\Delta\bar{w} \in \mathbb{R}$ this inequality becomes

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

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

$$\frac{\exp(\alpha\Delta\bar{w})\exp(\alpha\Delta\bar{w})}{0.5(\exp(\alpha\Delta\bar{w}) + \exp(\alpha\Delta\bar{w}))} \leqslant \exp(\theta_\alpha\Delta\bar{w}),$$

which may be rewritten as

$$\frac{\exp(\alpha\Delta\bar{w})\exp(\alpha\Delta\bar{w})}{\exp(\alpha\Delta\bar{w})} \leqslant \exp(\theta_\alpha\Delta\bar{w}),$$

and

$$\exp(\alpha\Delta\bar{w}) \leqslant \exp(\theta_\alpha\Delta\bar{w}).$$

Thus, two inequalities of this type, one with $\Delta\bar{w} > 0$ and the other one with $\Delta\bar{w} < 0$, will point identify α.

B.10 Using Inequalities for Estimation of Wage Coefficient

We describe here how we use the moment inequalities introduced in Section 3.2 to compute a confidence set for the wage parameter θ_α.

The moment in equation (17) depends on generic instrument vectors z_s and z_r. We construct z_s and z_r following equation (28). In our empirical application, we equate $\Delta x_s^{ll'}$ and $\Delta x_r^{ll'}$ to the one-year lag value of $\Delta w_s^{ll'}$ and $\Delta w_r^{ll'}$, respectively, and build vectors of instruments $z_s^{ll'}$ and $z_r^{ll'}$ for $q \in \{2, 4, 8, 16\}$.

The instruments in $z_s^{ll'}$ and $z_r^{ll'}$ are indicator functions and, thus, weakly positive. Consequently, given Theorem 3 and the LIE implies that, for any locations l and l', worker i of type s, worker j of type r, and and deterministic function $g_{ijsr}^{ll'}: \mathcal{Z}_s \times \mathcal{Z}_r \times \Theta_\alpha \to \mathbb{R}$, the $q^2 \times 1$ vector of moment inequalities

$$\mathbb{E}[\mathbb{M}_{ijsr}^{ll'}(z_s, z_r, \theta_\alpha, g_{ijsr}^{ll'}(\cdot))(z_s^{ll'} \otimes z_r^{ll'})] \geqslant 0, \tag{B.31}$$

holds for $\Delta\theta_{ll'} = \Delta\kappa^{ll'}$ if $z_s^{ll'} \subset \mathcal{J}_{is}$ and $z_r^{ll'} \subset \mathcal{J}_{jr}$. The choice of the number of intervals q is consequential for the validity of the inequalities in equation (B.31). If $q = 2$, $z_s^{ll'}$ includes the following two elements:

$$\begin{aligned} z_{s,1}^{ll'} &= \mathbb{1}\{-\infty < \Delta x_s^{ll'} \leqslant \operatorname{med}(\Delta x_s^{ll'})\}, \\ z_{s,2}^{ll'} &= \mathbb{1}\{\operatorname{med}(\Delta x_s^{ll'}) < \Delta x_s^{ll'} \leqslant \infty\}; \end{aligned} \tag{B.32}$$

and $z_r^{ll'}$ includes the following two elements:

$$\begin{aligned} z_{r,1}^{ll'} &= \mathbb{1}\{-\infty < \Delta x_r^{ll'} \leqslant \operatorname{med}(\Delta x_r^{ll'})\}, \\ z_{r,2}^{ll'} &= \mathbb{1}\{\operatorname{med}(\Delta x_r^{ll'}) < \Delta x_r^{ll'} \leqslant \infty\}. \end{aligned} \tag{B.33}$$

Thus, if $q = 2$, the set of values of $\Delta\theta_{ll'}$ consistent with the inequalities in equation (B.31) includes $\Delta\kappa^{ll'}$ if, for locations l and l', worker i of type s knows whether the realized value of $\Delta x_s^{ll'}$ is above or below the median of the distribution of sector-wage differences across locations and sectors, and worker j of type r knows whether the realized value of $\Delta x_r^{ll'}$ is above or below the same median.

The data setting described in Section 2 includes one observation per worker. The sample analogue of the moment inequality in equation (22) thus averages over only one observation. However, as this inequality is valid for every worker i of every type s, every worker j of every type r, and every pair of locations l and l', it holds that, for any function $g_{ijsr}^{ll'} \colon \mathcal{Z}_s \times \mathcal{Z}_r \times \Theta_\alpha \to \mathbb{R}$, the vector of moment inequalities

$$\sum_{s=1}^{S} \sum_{i=1}^{I_s} \sum_{r>s} \sum_{j=1}^{I_r} \mathbb{E}[\mathbb{M}_{ijsr}^{ll'}(z_s, z_r, \theta_\alpha, g_{ijsr}^{ll'}(\cdot))(z_s^{ll'} \otimes z_r^{ll'})] \geqslant 0 \tag{B.34}$$

is satisfied at $\Delta\theta_{ll'} = \Delta\kappa^{ll'}$ if $z_{s'}^{ll'} \subset \mathcal{J}_{is'}$ for every worker $i = 1, \ldots, I_{s'}$ of every type $s' = 1, \ldots, S$.

If the number of worker types S is small, it may be convenient to further aggregate the inequality in equation (B.34) across all location pairs (l, l'). However, it may be the case that the instrument vector $z_{s'}^{ll'}$ belongs to the information set of every worker only for a subset of location pairs (l, l'); e.g., only for urban locations. If this the case, we may aggregate the inequality in equation (B.34) only over location pairs (l, l') that belong to some subset specified by the researcher.

Given any significance level, we compute a confidence interval for θ_α by applying the inference procedure in Andrews and Soares (2010) to the sample analogue of the moment inequalities in equation (B.34); see Appendix B.11 for details.

B.11 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 Online Appendix A.7 in Dickstein and Morales (2018).

We base the construction of our confidence set for the true parameter on the modified method of moments (MMM) statistic. Denote as γ a generic parameter for which we want to compute a $1 - \alpha$ confidence set. In our context, the parameter γ may equal either θ_α or θ_l for some location l. Assume we use K sample moment inequalities to compute a confidence set for γ, and denote each of these inequalities as

$$\bar{m}_k(\gamma) \geqslant 0, \qquad k = 1, \ldots, K. \tag{B.35}$$

where, for each $k = 1, \ldots, K$,

$$\bar{m}_k(\gamma) \equiv \frac{1}{N} \sum_{c=1}^{C} \sum_{n=1}^{N_c} m_k(x_{nc}, \gamma), \tag{B.36}$$

and where observations are grouped into clusters $c = 1, \ldots, C$ and indexed by $n = 1, \ldots, N_c$ within each cluster c. The variable x_n is a generic vector of observed covariates. For example, in the context of the moment in equation (24), each of the clusters c in equation (B.36) corresponds to a sector s, and each observation n within a cluster c corresponds to a worker i within a sector s. In the context of the moment in equation (B.34), each cluster may correspond to a pair of sectors (s, r) and each observation n within a cluster may correspond to a tuple of individual indices and location indices (i, j, l, l'). Regardless of the definition of what an observation is, the variable N_c denotes the number of sample observations within a cluster c.

The MMM statistic is defined as

$$T(\gamma) = \sum_{k=1}^{K} (\min\{\sqrt{N} \frac{\bar{m}_k(\gamma)}{\hat{\sigma}_k(\gamma)}, 0\})^2, \tag{B.37}$$

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

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

Given a set of K inequalities and a grid Γ_g covering the parameter space of γ, we implement the following steps to compute a confidence set for this parameters:

Step 1: choose a point $\gamma_p \in \Gamma_g$. Steps 2 to 8 test the null hypothesis that γ^* equals γ_p:

$$H_0 : \gamma^* = \gamma_p \qquad \text{vs.} \qquad H_0 : \gamma^* \neq \gamma_p.$$

Step 2: evaluate the MMM test statistic at γ_p**:**

$$T(\gamma_p) = \sum_{k=1}^{K}(\min\{\sqrt{N}\frac{\bar{m}_k(\gamma_p)}{\hat{\sigma}_k(\gamma_p)}, 0\})^2, \tag{B.38}$$

Step 3: compute the correlation matrix of the moments evaluated at γ_p**:**

$$\hat{\Omega}(\gamma_p) = Diag^{-\frac{1}{2}}(\hat{\Sigma}(\gamma_p))\hat{\Sigma}(\gamma_p)Diag^{-\frac{1}{2}}(\hat{\Sigma}(\gamma_p)), \tag{B.39}$$

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

$$\hat{\Sigma}(\gamma_p) = \frac{1}{N}\sum_{c=1}^{C}(\sum_{n=1}^{N_c}(m(x_{nc},\gamma_p) - \bar{m}(\gamma_p)))(\sum_{n=1}^{N_c}(m(x_i,\gamma_p) - \bar{m}(\gamma_p)))', \tag{B.40}$$

where

$$m(x_{nc},\gamma_p) = (m_1(x_{nc},\gamma_p),\ldots,m_K(x_{nc},\gamma_p))', \tag{B.41}$$

$$\bar{m}(\gamma_p) = (\bar{m}_1(\gamma_p),\ldots,\bar{m}_K(\gamma_p))'. \tag{B.42}$$

Step 4: simulate the asymptotic distribution of $T(\gamma_p)$**.** Take D draws from the multivariate normal distribution $\mathbb{N}(0_K, I_K)$ where 0_K is a vector of 0s of dimension K and I_K is the identity matrix of dimension L. Denote each of these draws as ζ_d. Define the criterion function $T_d^{AA}(\gamma_p)$ as

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

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

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 γ_p**.** Include γ_p in the $(1-\delta)\%$ confidence set, $\hat{\Gamma}^{1-\delta}$, if $T(\gamma_p) \leqslant \hat{c}^{AA}(\gamma_p, 1-\delta)$.

Step 7: repeat steps 2 to 6 for every γ_p **in the grid** Γ_g**.**

Step 8: compare $\hat{\Gamma}^{1-\alpha}$ **to** Γ_g**.** If none of the points in $\hat{\Gamma}^{1-\alpha}$ are at the boundary of Γ_g, define $\hat{\Gamma}^{1-\alpha}$ as the 95% confidence set for γ^*. Otherwise, expand the limits of Θ_g and repeat steps 1 to 8.

C Additional Simulation Results

In Appendix C.1, we provide a detailed description of the moment inequalities we use to compute the confidence intervals reported in Table 1. In Appendix C.2, we explore the behavior of our estimation procedure in a setting analogous to that in Section 4.1 except for the fact that the amenity terms κ^l for $l = \{1, 2, 3\}$ differ across the three locations. In Appendix C.3, we build confidence intervals under alternative ways of building the moment inequalities described in Section 3.2. In Appendix C.4, we compare our moment inequality estimation procedure to that in Dickstein et al. (2023). In Appendix C.5, we study the performance of an alternative estimator to that described in Section 3 that may also be applied in settings with the parameter vector to estimate includes a large number of choice-specific fixed effects. In Appendix C.6, we present estimates analogous to those in reported in Table 1, but computed using a larger set of instruments. In Appendix C.7, we present figures that illustrate some of the confidence sets reported in Table 1. In Appendix C.8, we study the performance of our estimator in a setting that differs from that in Section 4.1 in that the precision of workers' wage information differs across locations.

C.1 Inequalities Used in Estimation of Confidence Intervals in Table 1

First Step: Cases 1 to 4. To compute the confidence intervals for θ_α corresponding to cases 1 to 4 in Table 1, we use the following sample moment inequality

$$\sum_{s=1}^{S}(y_s^l y_{r(s)}^l + y_s^{l'} y_{r(s)}^{l'} - y_s^l y_{r(s)}^{l'} \exp(-g_s^{ll'}(\Delta x_{2s}^{ll'}, \Delta x_{2r(s)}^{l'l}, \theta_\alpha))$$
$$(2 + 2g_s^{ll'}(\Delta x_{2s}^{ll'}, \Delta x_{2r(s)}^{l'l}, \theta_\alpha) - \theta_\alpha(\Delta w_s^{ll'} + \Delta w_{r(s)}^{l'l})))(z_s^{ll'} \otimes z_{r(s)}^{l'l}) \geqslant 0, \tag{C.1}$$

where $r(s)$ indexes a sector that we match with each sector s when computing the inequality in equation (C.1). The sector $r(s)$ may thus potentially vary across $s = 1, \ldots, S$. More specifically, we select the sector $r(s)$ to match with a given sector s randomly among those that satisfy the restriction

$$\mathbb{1}\{|\hat{\mathbb{E}}[\Delta w_s^{ll'}|\Delta x_{2s}^{ll'}, y_s^l = 1] - \hat{\mathbb{E}}[\Delta w_{r(s)}^{l'l}|\Delta x_{2r(s)}^{l'l}, y_{r(s)}^{l'} = 1]| \leqslant \tau\} = 1, \tag{C.2}$$

with $\tau = 0.02$. For any sector s, the term $\hat{\mathbb{E}}[\Delta w_s^{ll'}|\Delta x_{2s}^{ll'}, y_s^l = 1]$ denotes the predicted value of $\Delta w_s^{ll'}$ computed using a linear regression of $\Delta w_s^{ll'}$ on $\Delta x_{2s}^{ll'}$ estimated on the subsample of observations $s' = 1, \ldots, S$ with $y_{s'}^l = 1$. To understand the restriction in equation (C.2), one should first notice that the moment inequality introduced in equations (17) to (19) holds for any pair of sectors s and r. However, Corollary 3 indicates that the moment inequality defined by equations (17) and (19) has desirable properties when, combined with other conditions, sectors s and r satisfy the restriction

$$\mathbb{E}[\Delta w_s^{ll'}|\mathcal{J}_s] = \mathbb{E}[\Delta w_r^{l'l}|\mathcal{J}_r]. \tag{C.3}$$

We cannot impose this restriction as the information sets $\mathcal{J}_{is}$ and $\mathcal{J}_r$ are unobserved. Given these data limitations, we approximate the restriction in Corollary 3 when the variable x_2 is continuous by imposing the restriction in equation (C.2). In Appendix C.2, we indeed show that the confidence set for θ_α we obtain when implementing the moment inequality defined in equations (C.1) and (C.2) indeed increases as we increase the value of τ entering the expression in equation (C.2).

To complete the description of the inequality in equation (C.1), we must determine the function $g_s^{ll'}(\cdot)$ and the instrument vectors $z_s^{ll'}$ and $z_{r(s)}^{l'l}$. Building on the expression in equation (19), we impose

$$g_s^{ll'}(\Delta x_{2s}^{ll'}, \Delta x_{2r(s)}^{l'l}, \theta_\alpha) = \theta_\alpha 0.5(\hat{\mathbb{E}}[\Delta w_s^{ll'}|\Delta x_{2s}^{ll'}, y_s^l = 1] + \hat{\mathbb{E}}[\Delta w_{r(s)}^{l'l}|\Delta x_{2r(s)}^{l'l}, y_{r(s)}^{l'} = 1]). \tag{C.4}$$

We build the instrument vector $z_s^{ll'}$ as integrating two elements that we construct as

$$z_{s,1}^{ll'} = \mathbb{1}\{-\infty < \Delta x_{2s}^{ll'} \leqslant 0\}, \qquad \text{and} \qquad z_{s,2}^{ll'} = \mathbb{1}\{0 < \Delta x_{2s}^{ll'} \leqslant \infty\}. \tag{C.5}$$

Similarly, we build the instrument vector $z^{l'l}_{r(s)}$ as integrating two elements that we construct as

$$z^{l'l}_{r(s),1} = \mathbb{1}\{-\infty < \Delta x^{l'l}_{2r(s)} \leqslant 0\}, \qquad \text{and} \qquad z^{l'l}_{r(s),2} = \mathbb{1}\{0 < \Delta x^{l'l}_{2r(s)} \leqslant \infty\}. \tag{C.6}$$

First Step: Case 5. To compute the confidence interval for θ_α corresponding to case 5 in Table 1, we use

$$\sum_{s=1}^{S} (y^l_s y^l_{r(s)} + y^{l'}_s y^{l'}_{r(s)} - y^l_s y^{l'}_{r(s)} \exp(-g^{ll'}_s(\Delta w^{ll'}_s, \Delta w^{l'l}_{r(s)}, \theta_\alpha))$$
$$(2 + 2g^{ll'}_s(\Delta w^{ll'}_s, \Delta w^{l'l}_{r(s)}, \theta_\alpha) - \theta_\alpha(\Delta w^{ll'}_s + \Delta w^{l'l}_{r(s)})))(z^{ll'}_s \otimes z^{l'l}_{r(s)}) \geqslant 0, \tag{C.7}$$

where, for each $s = 1, \ldots, S$, the sector $r(s)$ is selected randomly among those that satisfy the restriction

$$\mathbb{1}\{|\Delta w^{ll'}_s - \Delta w^{l'l}_{r(s)}| \leqslant 0.02\} = 1; \tag{C.8}$$

the function $g^{ll'}_s(\cdot)$ is determined as

$$g^{ll'}_s(\Delta x^{ll'}_{2s}, \Delta x^{l'l}_{2r(s)}, \theta_\alpha) = \theta_\alpha 0.5(\Delta w^{ll'}_s + \Delta w^{l'l}_{r(s)}); \tag{C.9}$$

the vector $z^{ll'}_s$ includes the following two elements

$$z^{ll'}_{s,1} = \mathbb{1}\{-\infty < \Delta w^{ll'}_s \leqslant 0\}, \qquad \text{and} \qquad z^{ll'}_{s,2} = \mathbb{1}\{0 < \Delta w^{ll'}_s \leqslant \infty\}; \tag{C.10}$$

and the instrument vector $z^{l'l}_{r(s)}$ includes the following two elements

$$z^{l'l}_{r(s),1} = \mathbb{1}\{-\infty < \Delta w^{l'l}_{r(s)} \leqslant 0\}, \qquad \text{and} \qquad z^{l'l}_{r(s),2} = \mathbb{1}\{0 < \Delta w^{l'l}_{r(s)} \leqslant \infty\}. \tag{C.11}$$

Second Step: Bounding Inequalities for Cases 1 to 4. Given a value $\check{\theta}_\alpha$ of the parameter θ_α and a location $l = 2, \ldots, L$, we use the following moment inequalities to compute the confidence interval for θ_l in cases 1 to 4 in Table 1:

$$\sum_{s=1}^{S} (y^1_s - y^l_s \exp(-h^{l1}_s(\Delta x^{l1}_{2s}, \theta_l))(1 + h^{l1}_s(\Delta x^{l1}_{2s}, \theta_l) - (\theta_l + \check{\theta}_\alpha \Delta w^{l1}_s)))z^{l1}_s \geqslant 0, \tag{C.12a}$$

$$\sum_{s=1}^{S} (y^l_s - y^1_s \exp(-h^{1l}_s(\Delta x^{1l}_{2s}, -\theta_l))(1 + h^{1l}_s(\Delta x^{1l}_{2s}, -\theta_l) + (\theta_l + \check{\theta}_\alpha \Delta w^{l1}_s)))z^{1l}_s \geqslant 0; \tag{C.12b}$$

with the instrument vector z^{l1}_s being column vector that includes the following two elements

$$z^{l1}_{s,1} = \mathbb{1}\{-\infty < \Delta x^{l1}_{2s} \leqslant 0\}, \qquad \text{and} \qquad z^{l1}_{s,2} = \mathbb{1}\{0 < \Delta x^{l1}_{2s} \leqslant \infty\}; \tag{C.13}$$

the instrument vector z^{1l}_s being column vector that includes the following two elements

$$z^{1l}_{s,1} = \mathbb{1}\{-\infty < \Delta x^{1l}_{2s} \leqslant 0\}, \qquad \text{and} \qquad z^{1l}_{s,2} = \mathbb{1}\{0 < \Delta x^{1l}_{2s} \leqslant \infty\}; \tag{C.14}$$

and

$$h^{l1}_s(\Delta x^{l1}_{2s}, \theta_l) = \theta_l + \check{\theta}_\alpha \hat{\mathbb{E}}[\Delta w^{l1}_s | \Delta x^{l1}_{2s}, y^l_{is} = 1], \tag{C.15a}$$

$$h^{1l}_s(\Delta x^{1l}_{2s}, -\theta_l) = -\theta_l + \check{\theta}_\alpha \hat{\mathbb{E}}[\Delta w^{1l}_s | \Delta x^{1l}_{2s}, y^{l'}_{is} = 1]. \tag{C.15b}$$

For any sector s and locations l and l', the term $\hat{\mathbb{E}}[\Delta w^{ll'}_s | \Delta x^{ll'}_{2s}, y^l_s = 1]$ denotes the predicted value of $\Delta w^{ll'}_s$ computed using a linear regression of $\Delta w^{ll'}_{s'}$ on $\Delta x^{ll'}_{2s'}$ estimated on the subsample of observations $s' = 1, \ldots, S$ with $y^l_{s'} = 1$.

Second Step: Bounding Inequalities for Case 5. Given a value $\check{\theta}_\alpha$ of the parameter θ_α and a location

$l = 2, \ldots, L$, we use the following sample moment inequalities to compute the confidence interval for θ_l in case 5 in Table 1:

$$\sum_{s=1}^{S} (y_s^1 - y_s^l \exp(-h_s^{l1}(\Delta w_s^{l1}, \theta_l))(1 + h_s^{l1}(\Delta w_s^{l1}, \theta_l) - (\theta_l + \check{\theta}_\alpha \Delta w_s^{l1})))z_s^{l1} \geqslant 0, \tag{C.16a}$$

$$\sum_{s=1}^{S} (y_s^l - y_s^1 \exp(-h_s^{1l}(\Delta w_s^{1l}, -\theta_l))(1 + h_s^{1l}(\Delta w_s^{1l}, -\theta_l) + (\theta_l + \check{\theta}_\alpha \Delta w_s^{l1})))z_s^{1l} \geqslant 0; \tag{C.16b}$$

with the instrument vector z_s^{l1} being column vector that includes the following two elements

$$z_{s,1}^{l1} = \mathbb{1}\{-\infty < \Delta w_s^{l1} \leqslant 0\}, \qquad \text{and} \qquad z_{s,2}^{l1} = \mathbb{1}\{0 < \Delta w_s^{l1} \leqslant \infty\}; \tag{C.17}$$

the instrument vector z_s^{1l} being column vector that includes the following two elements

$$z_{s,1}^{1l} = \mathbb{1}\{-\infty < \Delta w_s^{1l} \leqslant 0\}, \qquad \text{and} \qquad z_{s,2}^{1l} = \mathbb{1}\{0 < \Delta w_s^{1l} \leqslant \infty\}; \tag{C.18}$$

and

$$h_s^{l1}(\Delta w_s^{l1}, \theta_l) = \theta_l + \check{\theta}_\alpha \Delta w_s^{l1}, \tag{C.19}$$

with $h_s^{1l}(\Delta w_s^{1l}, -\theta_l) = -h_s^{l1}(\Delta w_s^{l1}, \theta_l)$.

Second Step: Odds-based Inequalities for Cases 1 to 4. Given a value $\check{\theta}_\alpha$ of the parameter θ_α and a location $l = 2, \ldots, L$, we use the following inequalities to compute the confidence interval for θ_l in cases 1 to 4 in Table 1:

$$\sum_{s=1}^{S} (y_s^l \exp(-(\theta_l + \check{\theta}_\alpha \Delta w_s^{l1})) - y_s^1) z_s^{l1} \geqslant 0, \tag{C.20a}$$

$$\sum_{s=1}^{S} (y_s^1 \exp(\theta_l + \check{\theta}_\alpha \Delta w_s^{l1}) - y_s^l) z_s^{1l} \geqslant 0; \tag{C.20b}$$

with the instrument vector z_s^{l1} being column vector that includes the following two elements

$$z_{s,1}^{l1} = \mathbb{1}\{-\infty < \Delta x_{2s}^{l1} \leqslant 0\}, \qquad \text{and} \qquad z_{s,2}^{l1} = \mathbb{1}\{0 < \Delta x_{2s}^{l1} \leqslant \infty\}; \tag{C.21}$$

the instrument vector z_s^{1l} being column vector that includes the following two elements

$$z_{s,1}^{1l} = \mathbb{1}\{-\infty < \Delta x_{2s}^{1l} \leqslant 0\}, \qquad \text{and} \qquad z_{s,2}^{1l} = \mathbb{1}\{0 < \Delta x_{2s}^{1l} \leqslant \infty\}. \tag{C.22}$$

Second Step: Odds-based Inequalities for Case 5. Given a value $\check{\theta}_\alpha$ of θ_α and a location $l = 2, \ldots, L$, we compute the confidence interval for θ_l in case 5 in Table 1 using inequalities analogous to those in equation (C.20), with the only difference that z_s^{l1} is column vector that includes the following two elements

$$z_{s,1}^{l1} = \mathbb{1}\{-\infty < \Delta w_s^{l1} \leqslant 0\}, \qquad \text{and} \qquad z_{s,2}^{l1} = \mathbb{1}\{0 < \Delta w_s^{l1} \leqslant \infty\}; \tag{C.23}$$

and the instrument vector z_s^{1l} includes the following two elements

$$z_{s,1}^{1l} = \mathbb{1}\{-\infty < \Delta w_s^{1l} \leqslant 0\}, \qquad \text{and} \qquad z_{s,2}^{1l} = \mathbb{1}\{0 < \Delta w_s^{1l} \leqslant \infty\}. \tag{C.24}$$

C.2 Amenity Differences Across All Locations

Table C.1: Simulation Results - Moment Inequality Confidence Intervals With Amenity Differences

Case	σ_1	σ_3	z_s	$(\kappa_1, \kappa_2, \kappa_3)$	1st Step
					α
2	0	1	x_{2s}	$(0, 0, 1)$	[1 , 1.01]
2	0	1	x_{2s}	$(0, 0.5, 1)$	[0.92 , 1.09]
2	0	1	x_{2s}	$(0, 0, 2)$	[1 , 1.02]
2	0	1	x_{2s}	$(0, 1, 2)$	[0.73 , 1.29]
2	0	1	x_{2s}	$(0, 0, 3)$	[1 , 1.02]
2	0	1	x_{2s}	$(0, 1.5, 3)$	[0.55 , 1.49]

Note: The column α contains a 95% confidence interval for θ_α based on the moment inequality estimator introduced in Section 3.2 and described in detail in Appendix C.1. The column $(\kappa_1, \kappa_2, \kappa_3)$ displays the true value of the location-specific amenities in the simulated data. Confidence sets are computed using a 1-dimensional grids whose limits are $[0.5, 1.5]$. All confidence sets are computed following the procedure in Andrews and Soares (2010).

C.3 First-step Moment Inequalities with Loose Sectoral Matches

Table C.2: Simulation Results - Moment Inequality Confidence Intervals With Loose Matches

Case	σ_1	σ_3	z_s	τ	1st Step
					α
2	0	1	x_{2s}	8	[0.73 , 1.32]
2	0	1	x_{2s}	4	[0.79 , 1.25]
2	0	1	x_{2s}	2	[0.94 , 1.08]
2	0	1	x_{2s}	1	[0.98 , 1.03]
2	0	1	x_{2s}	0.8	[0.99 , 1.03]
2	0	1	x_{2s}	0.08	[1 , 1.02]
2	0	1	x_{2s}	0.008	[1 , 1.02]
2	0	1	x_{2s}	0.002	[1 , 1.01]

Note: The column α contains a 95% confidence interval for θ_α based on the moment inequality estimator introduced in Section 3.2 and described in detail in Appendix C.1. The column τ displays the maximum feasible distance between $\mathbb{E}[\Delta w_s^{ll'}|z_s]$ and $\mathbb{E}[\Delta w_r^{l'l}|z_r]$ in each moment inequality. Confidence sets are computed using a 1-dimensional grids whose limits are $[0.5, 1.5]$. All confidence sets are computed following the procedure in Andrews and Soares (2010). The case with $\tau = 0.002$ corresponds to the baseline case reported in Table 1.

C.4 One-step Moment Inequality Estimator When Choice Set is Small

Table C.3: Simulation Results - Confidence Intervals à la Dickstein et al. (2023).

Case	σ_1	σ_3	z_i	Mom. Ineq.	α	κ_2	κ_3
1	0	0	x_{2i}	Bounding	[1 , 1]	[0 , 0]	[1 , 1]
				Odds-based	[1 , 1]	[0 , 0]	[1 , 1]
				Both	[1 , 1]	[0 , 0]	[1 , 1]
2	0	1	x_{2i}	Bounding	[1 , 1]	[0 , 0]	[1 , 1]
				Odds-based*	[0.92 , 1.50]	[-0.33 , 0.33]	[0.67 , 1.33]
				Both	[1 , 1]	[0 , 0]	[1 , 1]
3(a)	1	0	x_{2i}	Bounding	[0.80 , 1.10]	[-0.30 , 0.30]	[0.70 , 1.30]
				Odds-based*	[1 , 1]	[0 , 0]	[1 , 1]
				Both	[1 , 1]	[0 , 0]	[1 , 1]
3(b)	2	0	x_{2i}	Bounding	[0.50 , 1.50]	[-1 , 1]	[0 , 1.95]
				Odds-based*	[1 ,1]∪[1.15 , 2.50]	[-1.50 , 1.50]	[-0.50 , 2.50]
				Both	[1 ,1]∪[1.15 , 1.50]	[-0.15 , 0.15]	[1 , 1.40]
4	1	1	x_{2i}	Bounding	[0.80 , 1.10]	[-0.30 , 0.30]	[0.70 , 1.30]
				Odds-based*	[0.92 , 1.50]	[-0.48 , 0.50]	[0.65 , 1.50]
				Both	[0.92 , 1.10]	[-0.33 , 0.30]	[0.70 , 1.30]
5	0	1	p_i	Bounding	[0.87 , 0.87]	[-0.05 , -0.03]	[0.85 , 0.88]
				Odds-based	∅	∅	∅
				Both	∅	∅	∅

Note: *Odds-based*, *Bounding*, and *Both* contain projections on each parameter of 95% confidence sets computed as in Andrews and Soares (2010). *Odds-based* indicates the confidence set is computed using inequalities of the type in Section 3.1.2; *Bounding* indicates the confidence set is computed using inequalities of the type in Section 3.1.1; *Both* indicates it is computed using both types of inequalities. In all cases other than 3(b), confidence sets are computed using a 3-dimensional grid whose sides are $[0.5, 1.5]$ (for α), $[-0.5, 0.5]$ (for κ_2) and $[0.5, 1.5]$ (for κ_3). In case 3(b), we use a grid whose sides are $[-0.5, 2.5]$ (for α), $[-1.5, 1.5]$ (for κ_2) and $[-0.5, 2.5]$ (for κ_3). We mark with an asterisk when the confidence set includes points outside the grid.

C.5 Alternative Moment Inequality Estimator When Choice Set is Large

Table C.4: Simulation Results - Alternative Confidence Intervals With Large Choice Set.

Case	σ_1	σ_3	z_i	Mom. Ineq.	α	κ_2	κ_3
1	0	0	x_{2i}	Bounding	[1 , 1]	[0 , 0]	[1 , 1]
				Odds-based	[1 , 1]	[0 , 0]	[1 , 1]
				Both	[1 , 1]	[0 , 0]	[1 , 1]
2	0	1	x_{2i}	Bounding	[1 , 1]	[0 , 0]	[1 , 1]
				Odds-based*	[0.91 , 1.50]	[-0.33 , 0.32]	[0.68 , 1.33]
				Both	[1 , 1]	[0 , 0]	[1 , 1]
3(a)	1	0	x_{2i}	Bounding	[0.80 , 1.10]	[-0.31 , 0.31]	[0.70 , 1.30]
				Odds-based*	[1 , 1]	[0 , 0]	[1 , 1.01]
				Both	[1 , 1]	[0 , 0]	[1 , 1.01]
3(b)	2	0	x_{2i}	Bounding	[0.55 , 1.46]	[-0.98 , 0.98]	[-0.02 , 1.98]
				Odds-based*	[1 , 2.75]	[-1.75 , 1.75]	[-0.75 , 2.75]
				Both	[1 ,1.46]	[-0.14 , 0.21]	[1 , 1.39]
4	1	1	x_{2i}	Bounding	[0.79 , 1.10]	[-0.31 , 0.31]	[0.69 , 1.30]
				Odds-based*	[0.92 , 1.50]	[-0.49 , 0.50]	[0.64 , 1.50]
				Both	[0.92 , 1.10]	[-0.31 , 0.31]	[0.69 , 1.31]
5	0	1	p_i	Bounding	[0.87 , 0.88]	[-0.05 , -0.01]	[0.86 , 0.88]
				Odds-based	$\varnothing$	$\varnothing$	$\varnothing$
				Both	$\varnothing$	$\varnothing$	$\varnothing$

Note: *Odds-based*, *Bounding*, and *Both* contain projections on each parameter of 95% confidence sets computed as in Andrews and Soares (2010). *Odds-based* indicates the confidence set is computed using inequalities of the type in Section 3.1.2; *Bounding* indicates the confidence set is computed using inequalities of the type in Section 3.1.1; *Both* indicates it is computed using both types of inequalities. In all cases other than 3(b), confidence sets are computed using a 3-dimensional grid whose sides are $[0.5, 1.5]$ (for α), $[-0.5, 0.5]$ (for κ_2) and $[0.5, 1.5]$ (for κ_3). In case 3(b), we use a grid whose sides are $[-0.75, 2.75]$ (for α), $[-1.75, 1.75]$ (for κ_2) and $[-0.75, 2.75]$ (for κ_3). We mark with an asterisk when the confidence set includes points outside the grid.

C.6 Two-step Moment Inequalities with Additional Instruments

Table C.5: Simulation Results - Confidence Intervals With Additional Instruments

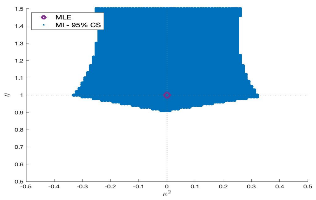
Case	σ_1	σ_3	z_s	First Step	Second Step		
				α	Mom. Ineq.	κ_2	κ_3
1	0	0	x_{2s}	[1 , 1.02]	Bounding	[0 , 0]	[1 , 1]
					Odds-based	[0 , 0]	[1 , 1]
					Both	[0 , 0]	[1 , 1]
2	0	1	x_{2s}	[1 , 1.01]	Bounding	[0 , 0]	[1 , 1]
					Odds-based	[-0.33 , 0.32]	[0.68 , 1.33]
					Both	[0 , 0]	[1 , 1]
3	1	0	x_{2s}	[0.82 , 1.29]	Bounding	[-0.31 , 0.31]	[0.70 , 1.30]
					Odds-based	[0 , 0]	[1 , 1.01]
					Both	[0 , 0]	[1 , 1.01]
3(b)	2	0	x_{2s}	[0.58 , 2.26]	Bounding	[-0.98 , 0.98]	[-0.02 , 1.98]
					Odds-based	[-1.75 , 1.75]	[-0.44 , 2.75]
					Both	[-0.14 , 0.21]	[1 , 1.39]
4	1	1	x_{2s}	[0.82 , 1.31]	Bounding	[-0.31 , 0.31]	[0.69 , 1.31]
					Odds-based	[-0.38 , 0.39]	[0.68 , 1.45]
					Both	[-0.31 , 0.31]	[0.69 , 1.31]
5	0	1	w_s	[0.87 , 0.87]	Bounding	[-0.05 , -0.10]	[0.85 , 0.88]
					Odds-based	$\varnothing$	$\varnothing$
					Both	$\varnothing$	$\varnothing$

Note: The true parameter values are $\alpha = 1$, $\kappa^2 = 0$, and $\kappa^2 = 1$. The column α contains a 95% confidence interval for θ_α based on the moment inequality estimator introduced in Section 3.2 and described in detail in Appendix C.1. The columns κ^2 and κ^3 contain 95% confidence intervals for θ_2 and θ_3, respectively, based on the moment inequality estimators introduced in Section 3.1 and the inference procedure described in Section 3.3. The confidence intervals for θ_2 and θ_3 reported in the rows labeled *Bounding* use the moment inequalities introduced in Section 3.1.1; those reported in the row labeled *Odds-based* use the moment inequalities introduced in Section 3.1.2; and those reported in the row labeled *Both* combine the inequalities introduced in sections 3.1.1 and 3.1.2. For additional details on the moment inequalities used to compute the confidence sets reported in this table, see Appendix C.1. In all cases other than 3(b), confidence sets are computed using a 1-dimensional grids whose sides are $[0.5, 1.5]$ (for α), $[-0.5, 0.5]$ (for κ_2) and $[0.5, 1.5]$ (for κ_3). In case 3(b), we use grids whose sides are $[-0.5, 2.5]$ (for α), $[-1.5, 1.5]$ (for κ_2) and $[-0.5, 2.5]$ (for κ_3). All confidence sets are computed following the procedure in Andrews and Soares (2010).

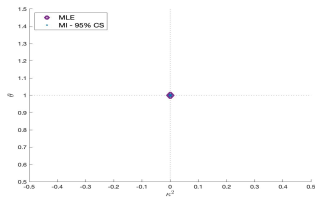
C.7 Advantages of Combining Bounding and Odds-based Moment Inequalities

Figure C.1: Case 2 – $\sigma_1 = 0$ and $\sigma_3 = 1$

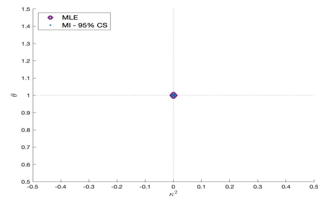
(a) Projected Confidence Set for (κ_2, θ)



Odds-based Ineq.

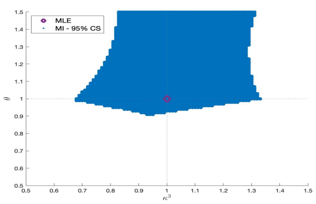


Bounding Ineq.

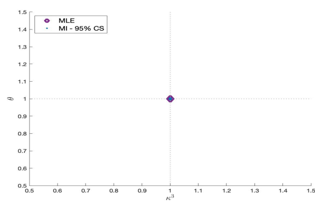


Both Types

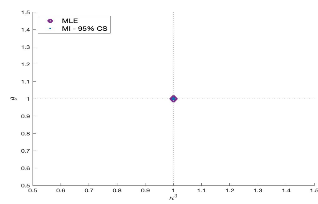
(b) Projected Confidence Set for (κ_3, θ)



Odds-based Ineq.



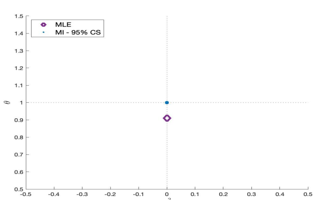
Bounding Ineq.



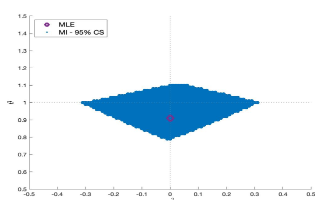
Both Types

Figure C.2: Case 3(a) – $\sigma_1 = 1$ and $\sigma_3 = 0$

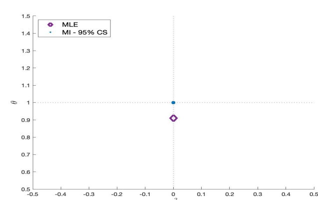
(a) Projected Confidence Set for (κ_2, θ)



Odds-based Ineq.

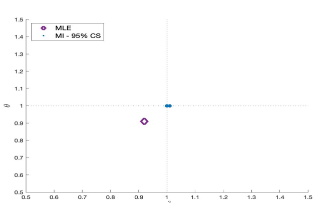


Bounding Ineq.

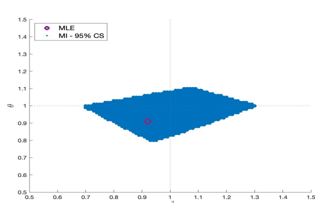


Both Types

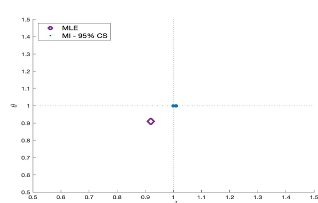
(b) Projected Confidence Set for (κ_3, θ)



Odds-based Ineq.



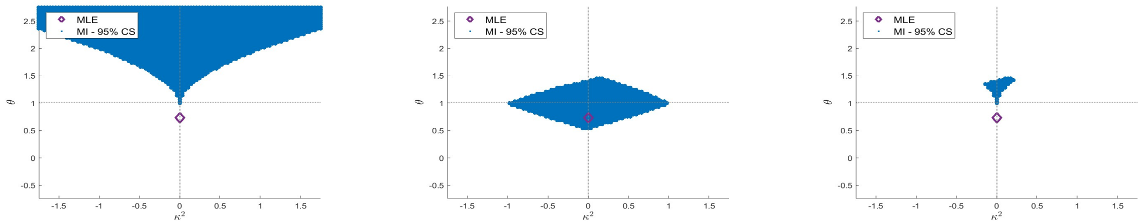
Bounding Ineq.



Both Types

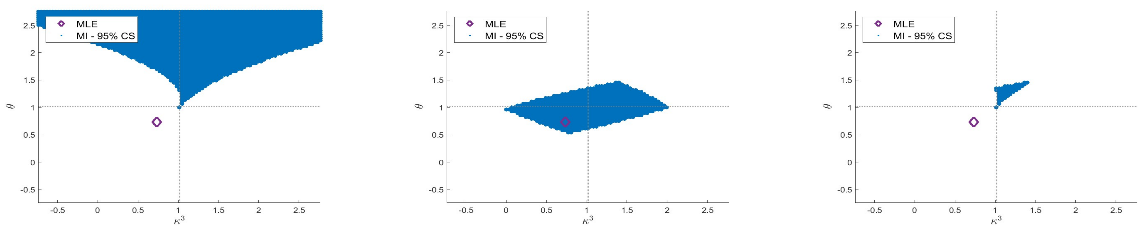
Figure C.3: Case 3(b) – $\sigma_1 = 2$ and $\sigma_3 = 0$

(a) Projected Confidence Set for (κ_2, θ)



Odds-based Ineq. | Bounding Ineq. | Both Types

(b) Projected Confidence Set for (κ_3, θ)



Odds-based Ineq. | Bounding Ineq. | Both Types

C.8 Location-specific Information Sets

D Data and Summary Statistics

D.1 Data Sources and Sample Construction

The RAIS data. Our primary data source is the *Relação Anual de Informações Sociais* (RAIS), an administrative dataset maintained by Brazil's Ministry of Labor and Employment (local acronym MTE). It includes the universe of formal Brazilian employment spells in the private and public sectors. Individual workers are identified by their government-issued identification numbers (PIS/PASEP and CPF), allowing us to track them as they change employers (and employment location). For all employment 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 occupation (according to the *Classificação Brasileira de Ocupações*, CBO), 2-digit sector (industry) of production (according to the *Classificação Nacional de Atividades Econômicas*, CNAE), as well as information on the worker's gender, age, race and level of education. All information on RAIS is reported by the employers.

RAIS only contains information on the formal employment of workers in Brazil. Thus, we have no information on the location of workers without formal jobs in a given year. These workers may be employed in the informal sector, self-employed, unemployed, or out of the labor force. Based on the 2010 Census, which directly asks respondents about their job status, 51% of the Brazilian labor force was in the formal sector. The implied total number of formal sector workers in the Census also closely matches the number of individual workers at RAIS.

Geography and income definitions. To determine workers' location and migration decisions, we use the microregion of the *plant* at which the worker is employed. This has the advantage of correctly locating workers in cases where a firm has several plants in different locations. Microregions are groups of municipalities that span the entirety of the Brazilian territory. They are officially defined by the Brazilian Institute of Geography and Statistics (local acronym IBGE). During our sample period, Brazil had 558 microregions. Microregions are also grouped into larger 136 mesoregions and are contained within 26 states and the federal district. While RAIS does not contain information on the residence of workers, Dix-Carneiro and Kovak (2017) use 2000 Census data to report that only 3.4% of individuals lived and worked in different microregions.

Previous research has used microregions as local labor markets (similar to commuting zones in the United States). For examples and further discussion, see Dix-Carneiro (2014), Dix-Carneiro and Kovak (2017), Dix-Carneiro and Kovak (2019), Felix (2022), and Szerman (2024).

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

Sample restrictions and sampling. We limit our data to workers between 25 and 64 years of age. This ensures we observe them after the vast majority of the population completed their education and limits the age before a large share is retired. We limit our sample to 2002-2011. However, we use the information from 1993-2001 to measure each worker's experience in each sector, occupation, 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.

For computational reasons, our empirical application uses a sample of 10 million worker-year pairs. We randomly draw one thousand observations from each of the 10 thousand origin-sector-year cells defined by the 50 most microregions with the largest formal employment, the 20 2-digit sectors with the largest formal employment, and each of the 10 years in 2002-2011.

Additional data sources. Microregion population and demographic characteristics are from the 2000 and 2010 Censuses collected by the *Instituto Brasileiro de Geografia e Estatística* (IBGE). We calculate distances between microregions by using the geodesic distance between their 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.

D.2 Summary Statistics

D.2.1 Employment in RAIS versus Census

Before describing our selected sample, we validate our sample selection criterion based on employment recorded in RAIS by we comparing the employment measures in RAIS with the employment recorded in the 2010 decennial Census wave. As discussed in Section D.1, RAIS only covers the subset of individuals with formal employment. Despite this limitation, Table D.1 and Table D.2 confirm that the employment measures we use to define microregions' populations align well with those implied by the Census dataset. We conclude that the largest labor markets recorded in RAIS are also the largest for the whole Brazilian workforce.

Table D.1: RAIS: Top / Bottom 10 Microregions by Population Size

Rank	**Name**	**Unit number**	**Total Employees**	**Share Employees**
Top1	Sao Paulo	35061	3,091,449	0.139
Top2	Rio de Janeiro	33018	1,868,663	0.0825
Top3	Belo Horizonte	31030	1,045,383	0.0459
Top4	Porto Alegre	43026	735,832	0.0327
Top5	Curitiba	41037	721,342	0.0323
Top6	Salvador	29021	617,272	0.0270
Top7	Brasilia	53001	539,744	0.0239
Top8	Campinas	35032	519,499	0.0232
Top9	Recife	26017	506,985	0.0223
Top10	Fortaleza	23016	493,997	0.0219
Bottom10	Baixa Verde	24013	348.2	1.71e-05
Bottom9	Serrana do Sertao Alagoano	27001	315.8	1.55e-05
Bottom8	Baixo Parnaiba Maranhense	21013	309.3	1.52e-05
Bottom7	Caracarai	14003	287.9	1.41e-05
Bottom6	Mazagao	16004	276.3	1.35e-05
Bottom5	Alvorada D'Oeste	11005	274.2	1.34e-05
Bottom4	Pio IX	22014	271.4	1.33e-05
Bottom3	Rosario Oeste	51016	269	1.32e-05
Bottom2	Camaqua	43028	229.2	1.12e-05
Bottom1	Tabuleiro	42017	218.9	1.07e-05

This Table lists the 10 microregions that have the largest total employment in our RAIS dataset, before drawing our working sample with 1 million observations per year. Column 4 reports the mean number of employees over the 10 years of our sample.

Table D.2: Census: Top / Bottom 10 Microregions by Population Size

Rank	**Name**	**Unit number**	**Total Employees**
Top1	Sao Paulo	35061	807,643
Top2	Rio de Janeiro	33018	692,180
Top3	Belo Horizonte	31030	266,329
Top4	Porto Alegre	43026	217,302
Top5	Recife	26017	178,077
Top6	Salvador	29021	172,621
Top7	Curitiba	41037	165,188
Top8	Fortaleza	23016	159,207
Top9	Campinas	35032	140,114
Top10	Brasilia	53001	111,041
Bottom10	Caracarai	14003	2,531
Bottom9	Almeirim	15003	2,503
Bottom8	Paranatinga	51008	2,386
Bottom7	Meruoca	23004	2,109
Bottom6	Boca do Acre	13011	1,957
Bottom5	Sena Madureira	12003	1,954
Bottom4	Oiapoque	16001	1,743
Bottom3	Japura	13002	1,533
Bottom2	Amapa	16002	1,409
Bottom1	Fernando de Noronha	26019	235

This Table lists the 10 microregions that have the largest total employment in the 2010 Census wave. Column 4 reports the number of employees sampled that year in the Census.

D.2.2 Employment By Sector and Microregion

We describe the composition of employment across our 1,000 selected labor markets, comprising the 50 largest microregions and 20 largest sectors. The sample contains 1 million observations per year between 2002 and 2011, with a total of 10 million observations. All individuals belong to the largest demographic group of white males with at least a high school education. Table D.3 lists the total employment over our sample period in each of our 20 selected sectors. The largest sector is "Service Activities to Businesses," comprising about 20 percent of our sample's employment. The sectors are all represented by at least 150,000 employees over the sample period and cover a wide range of manufacturing and services industries.

Table D.4 lists the total employment over our sample period in each of the 50 selected microregions. In line with the analysis in Table D.1, São Paulo and Rio de Janeiro are the largest microregions of employment for the selected demographic group.

D.2.3 Transitions Between Microregions and Sectors

We computed the number of individuals who moved across microregions for each demographic type- year-sector. We created a transition dataset to record the frequency of transitions from or towards a given microregion by year, sector, and demographic type and the income changes associated with the transitions.

For each year, demographic type, sector (2 digits), microregion of origin, and microregion of destination, it records the number of individuals in the origin, the number of individuals who moved from the origin to the destination, the mean income in the origin, the mean income in the destination and the mean income of the individuals who moved.

Table D.3: Distribution of Employees by Sector (Total over Years)

Sector	**Sector (2 digit)**	**Nb. Employees**	**Share Employees**
Service Activities to Businesses	74	2,051,755	20.52
Wholesale Trade in Goods and General Merchandise	51	828,122	8.28
Rail and Road Passengers transportation	60	794,868	7.95
Construction Services and Installation Works	45	742,102	7.42
Motor Vehicle Parts and Accessories Manufacturing	34	628,111	6.28
Education and Teaching Activities	80	611,787	6.12
Household Appliance, Machinery Manufacturing	29	554,905	5.55
Food Processing and Manufacturing	15	473,506	4.74
Health Related Activities and Social Services	85	391,755	3.92
Metal Product Manufacturing	28	376,045	3.76
Software and Computer Dvpment, Consulting	72	325,470	3.25
Plastics Product Manufacturing	25	307,699	3.08
Activities Related to Freight Transport	63	299,258	2.99
Professional, Political Organizations and Trade Unions	91	290,901	2.91
Real Estate	70	276,707	2.77
Nonferrous Metal Foundries	27	276,349	2.76
Medias Publishing, Printing and Reproducing	22	214,406	2.14
Production and Distribution of Electrical, Energy	40	193,728	1.94
Performing, Arts, Sports and Leisure Activities	92	187,785	1.88
Electrical Machinery, and Supplies Manufacturing	31	174,741	1.75
Total		10,000,000	100.00

This Table summarizes the employment composition of our main sample, obtained after drawing a random subsample of 1 million employees each year between 2002 and 2011 from the 20 largest sectors and 50 microregions.

D.2.4 Summary Statistics: Migration

This subsection provides summary statistics describing key migration patterns. It focuses on white male workers with at least a high school education (the demographic group that is the focus of our application) in the 2002-2011 period. All figures are constructed using all workers in this demographic group (i.e., not only the random subsample focused on the 50 largest sectors and microregions).

Figure D.1 provides migration rates (the share of workers that change microregion of employment between year t and $t-1$) by year. It shows an upward trend over the sample period, from close to 6% in 2002 to 8% in 2011. As a comparison, the overall migration rate in our random subsample in the same period is 6.3%. Figure D.1 also provides migration rates conditional on the distance between origin and destination microregions, indicating that about a third of moves are to microregions within 100 km from the origin, and less than a sixth of moves involve migration over a distance larger than a 1,000 km. Figure D.2 provides a similar figure for changes in the employment sector, which are more common. It also provides the share of workers that change both microregion and sector of employment in a given year, indicating that most changes in the employment sector are not accompanied by migration.

Figure D.3 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 of moves occur at larger distances, including substantial mass in distances over 500 km.

Figure D.4 provides a scatter plot depicting the distribution of in-migration and out-migration rates for the 50 largest microregions which are the focus of our random subsample. Each marker represents one of these microregions, indicating its overall out- and in-migration rates (share of workers moving into and out of it between two consecutive years) in the sample period. In- and out-migration rates strongly correlated across microregions, varying close to one to one. Note the figure also depicts the distribution of both variables. The bulk of microregions have migration rates in the 3%-10% with four other with migration rates of roughly 14% and two outliers that experience large migration flows. D.5 provides similarly constructed figures exploring the relationship between migration rates and microregion size (measured as the number of workers observed in the data), indicating their correlation is quite low.

Lastly, Figures D.6 and D.7 provide similar scatter plots depicting the distribution of internet access over time and its relationship with microregion characteristics. Internet access in 2007 is increasing in municipality size, and this relationship is accentuated by 2011. Internet access was lower in more remote microregions (proxied here by distance to São Paulo, Brazil's largest city). This relationship also becomes steeper by 2017. Both figures indicate a sizable internet penetration expansion between 2007 and 2011.

Figure D.1: Migration Rates by Year

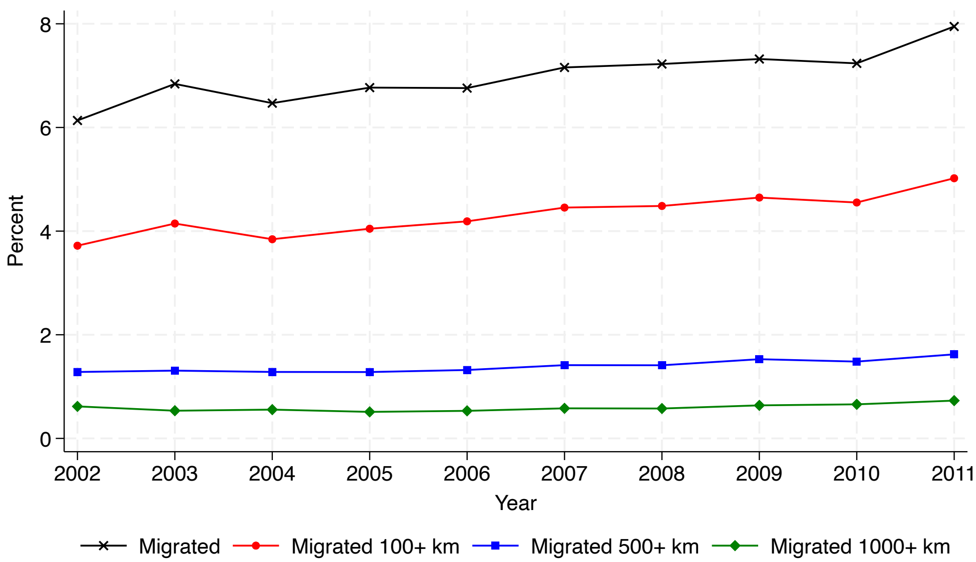
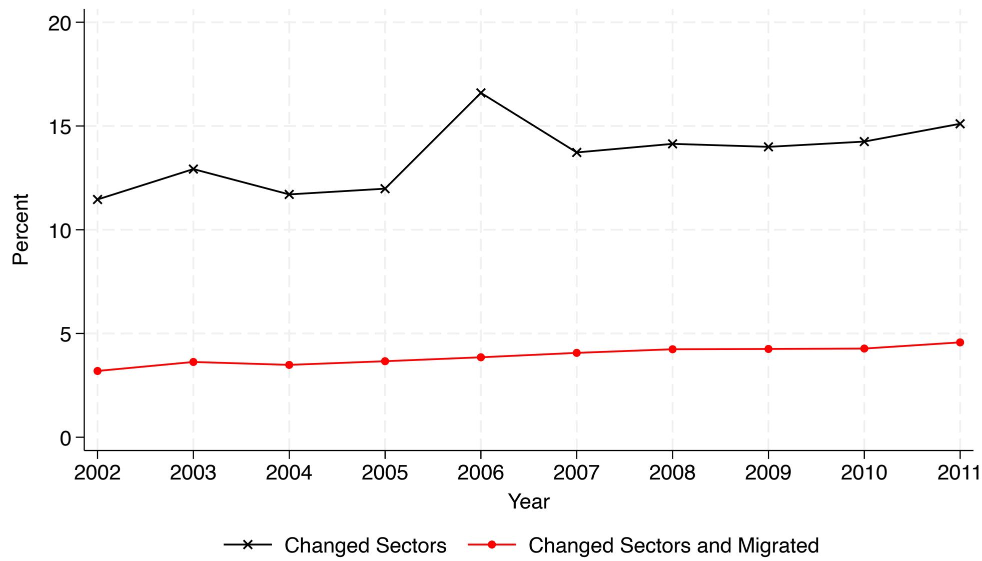


Figure shows migration rates (share of individuals who changed microregion from the previous year) for each year. Migration rates are further refined by distance (only including workers that moved to microregions 100km or more, 500km or more, 1000km or more). Data includes all white male workers with at least a high school education in the relevant period.

Figure D.2: Share of Individuals Changing Sector and Microregion by Year



The top line shows the share of workers that changed sectors from the previous year. The bottom line shows the share that changed both sector and migrated (changed microregion from the previous year). Data includes all white male workers with at least a high school education in the relevant period.

Figure D.3: Histogram: Migration Distance

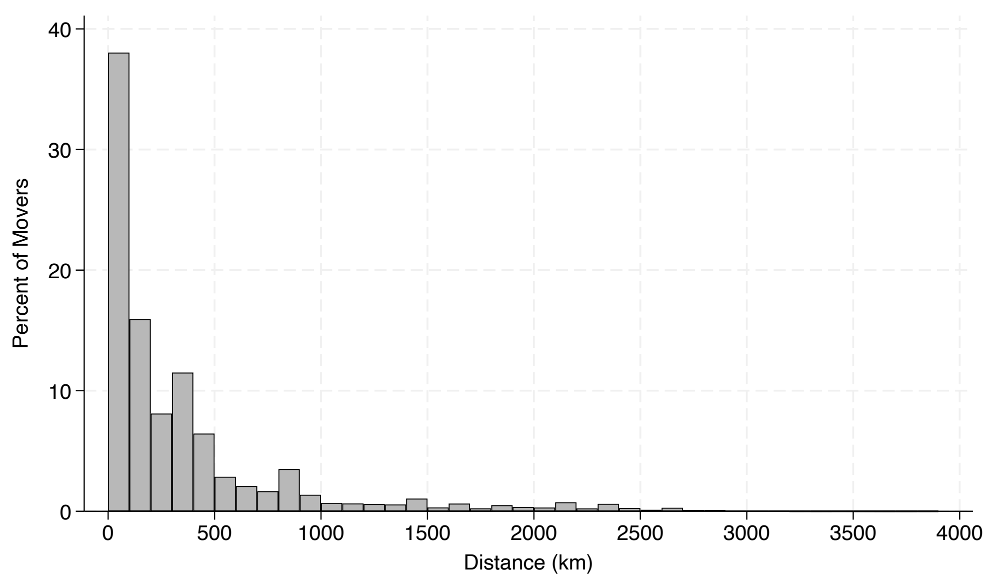
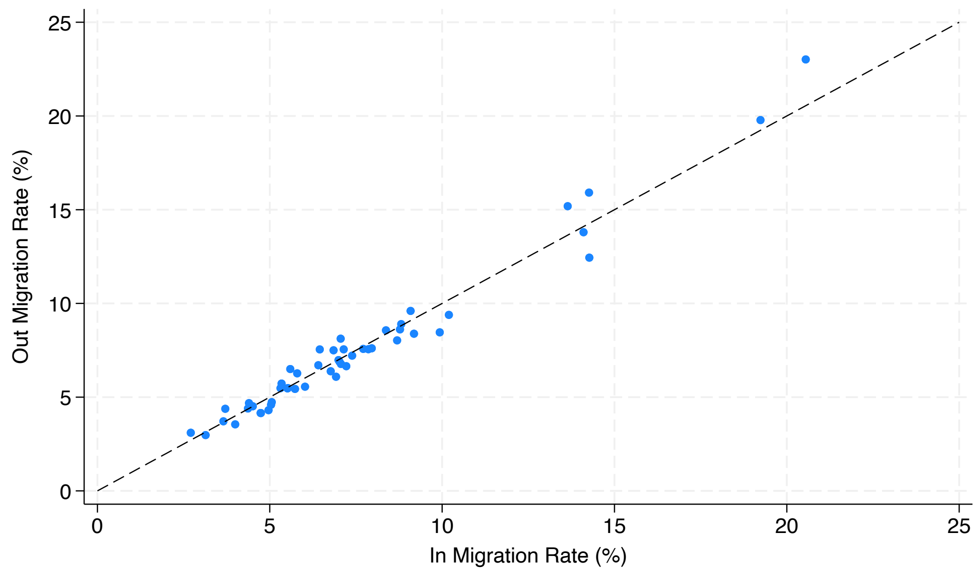


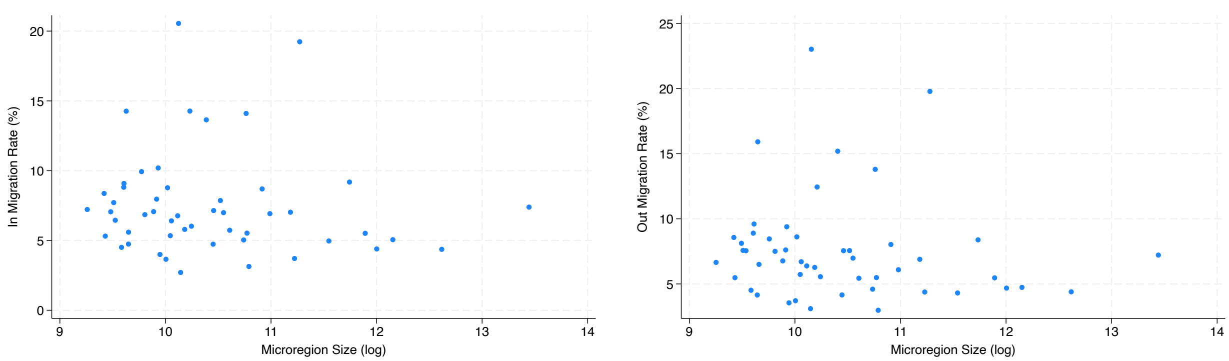
Figure provides a histogram of the distribution of distance between origin and destination of migrants in the 2002-2011 period. Data includes all white male workers with at least a high school education in the relevant period, conditional on having migrated.

Figure D.4: Out- versus In-Migration: Sample of 50 Largest Microregions



Each marker in the figure represents one of the 50 microregions in the subsample. The y-axis measures the out-migration rate (share of workers that move out of the microregion in a year), while the x-axis measure the in-migration (share of workers that move into the microregion). Data includes all white male workers with at least a high school degree in the 2002-2011 period. We consider migration with origins and destinations to all microregions (including outside the 50 largest ones). Dashed line represents the 45 degree line.

Figure D.5: Migration and Municipality Size

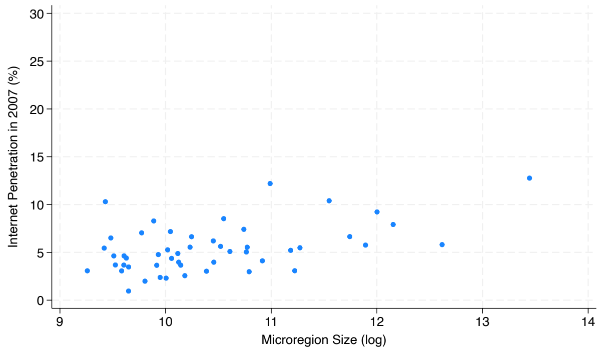


(a) In-Migration Rates

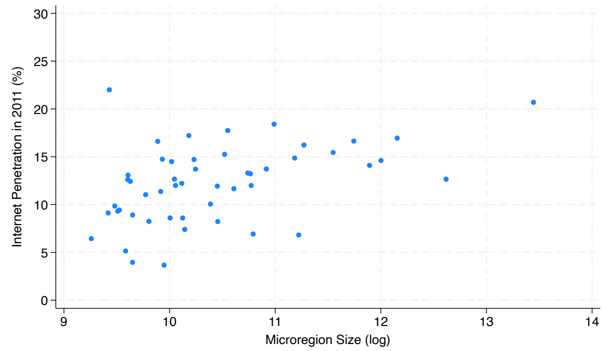
(b) Out-Migration Rates

Each marker in the figure represents one of the 50 microregions in the subsample. The y-axis measures the in-migration and out-migration rates (share of workers that move into or out of a microregion between two consecutive years), respectively. The x-axis measures the log of average yearly number of workers. All variables are based on data including all white male workers with at least a high school degree in the 2002-2011 period.

Figure D.6: Internet Access and Microregion Size: Sample of Largest 50 Microregions



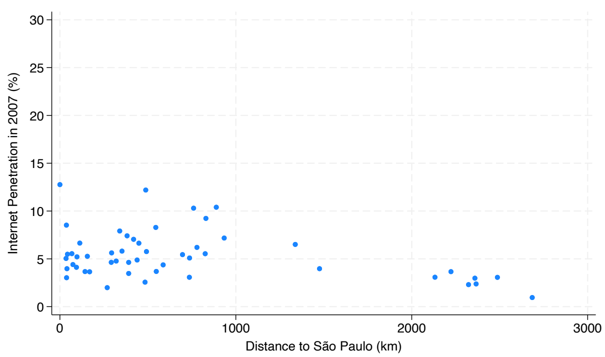
(a) 2007 Internet Access



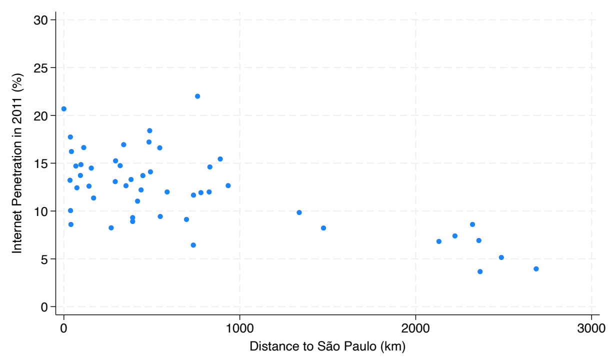
(b) 2011 Internet Access

Each marker in the figure represents one of the 50 microregions in the subsample. The y-axis measures the internet penetration rate in 2007 and 2011 for panels (a) and (b), respectively. The x-axis measures the log of average yearly number of workers, based on data including all white male workers with at least a high school degree in the 2002-2011 period.

Figure D.7: Internet Access and Distance to São Paulo: Sample of Largest 50 Microregions



(a) 2007 Internet Access



(b) 2011 Internet Access

Each marker in the figure represents one of the 50 microregions in the subsample. The y-axis measures the internet penetration rate in 2007 and 2011 for panels (a) and (b), respectively. The x-axis measures the distance between the microregion and São Paulo.

Table D.4: Distribution of Employees by Microregion (Total over Years)

Rank	**Microregion**	**Code**	**Nb. Employees**	**Share Employees**
1	São Paulo	35061	2,548,738	25.48
2	Rio de Janeiro	33018	946,725	9.46
3	Curitiba	41037	626,351	6.26
4	Porto Alegre	43026	571,251	5.71
5	Belo Horizonte	31030	500,020	5.00
6	Campinas	35032	423,185	4.23
7	Osasco	35057	308,658	3.08
8	Brasília	53001	223,655	2.23
9	São José dos Campos	35050	223,002	2.23
10	Recife	26017	221,367	2.21
11	Joinville	42008	205,087	2.05
12	Sorocaba	35046	203,146	2.03
13	Guarulhos	35059	184,586	1.84
14	Caxias do Sul	43016	146,702	1.46
15	Santos	35063	146,037	1.46
16	Florianópolis	42016	137,472	1.37
17	Goiânia	52010	135,607	1.35
18	Vitória	32009	131,470	1.31
19	Fortaleza	23016	120,507	1.20
20	Ribeirão Preto	35014	119,289	1.19
21	Salvador	29021	111,002	1.11
22	Mogi das Cruzes	35062	107,911	1.07
23	Jundiaí	35047	105,198	1.05
24	Itapecerica da Serra	35060	97,556	0.97
25	Londrina	41011	94,602	0.94
26	Piracicaba	35028	87,526	0.87
27	Blumenau	42012	80,488	0.80
28	Vale do Paraíba Fluminense	33011	79,172	0.79
29	Uberlândia	31018	73,926	0.73
30	São José do Rio Preto	35004	70,729	0.70
31	Limeira	35027	65,605	0.65
32	Bauru	35020	63,077	0.63
33	Maringá	41009	62,499	0.62
34	Bragança Paulista	35048	54,873	0.54
35	Itajaí	42013	53,024	0.53
36	Natal	24018	52,274	0.52
37	Moji Mirim	35031	52,034	0.52
38	Juiz de Fora	31065	51,698	0.51
39	Campo Grande	50004	51,646	0.51
40	Manaus	13007	50,312	0.50
41	Aglomeração Urbana de São Luís	21002	48,119	0.48
42	Araraquara	35024	45,630	0.04
43	Ponta Grossa	41021	44,008	0.44
44	São Carlos	35025	43,977	0.43
45	Macaé	33004	41,508	0.41
46	Passo Fundo	43010	40,939	0.40
47	João Pessoa	25022	40,913	0.40
48	Cascavel	41023	39,312	0.39
49	Presidente Prudente	35036	35,877	0.35
50	Criciúma	42019	31,710	0.31
	Total		10,000,000	1.00

This Table summarizes the employment composition of our main sample, obtained after drawing a random subsample of 1 million employees each year between 2002 and 2011 from the 20 largest sectors and 50 microregions.

E Appendix to Empirical Analysis

E.1 Wage Regressions

In this section, we compute proxies for the log wages in all labor markets for all individuals in our sample. We consider a wage process that depends on a rich individual-sector-year component, that workers consider as constant when deciding which location to move to. The relevant component of wages for their decision is then the location-sector-year component of log wages. The objective is to obtain a wage proxy that is as precise as possible after accounting for the individual and sector heterogeneity, in order to credibly interpret our proxies as local labor market demand shifters.

Workers indexed by i belong to different demographic types indexed by d. We construct eight demographic types from the combinations of three two-category variables: education (less than high school, at least high school), gender (male, female), and race (white, non-white). A market is defined by the intersection of a sector s, a location l, and a demographic group d, which we omit in the expressions below to lighten the notation. A worker's choice of a labor market at t may depend on their history h, i.e., the sequence of sectors of employment until period $t-1$.

Rewriting (3a) and (3b) to make explicit the role of sectors, locations, and histories, the static utility of choosing market sl at period t for a worker i of a demographic group d that has a labor history h consistent with a prior labor market sn is given by

$$U_{hit}^{sl} = u_{hit}^{sl} + \varepsilon_{it}^{sl}, \tag{E.1}$$

$$u_{hit}^{sl} = \kappa_{nt}^{l} + w_{hit}^{sl}, \tag{E.2}$$

where we only consider individuals who remain in the same sector s. We assume that an individual's log wage w_{hit}^{sl} in a sector can be expressed as the sum of a labor market-specific term w_t^{sl}, common to all agents, an individual skill for that sector that depends on a persistent match term and the number of years of age and of experience in the sector, and an unexpected wage shock,

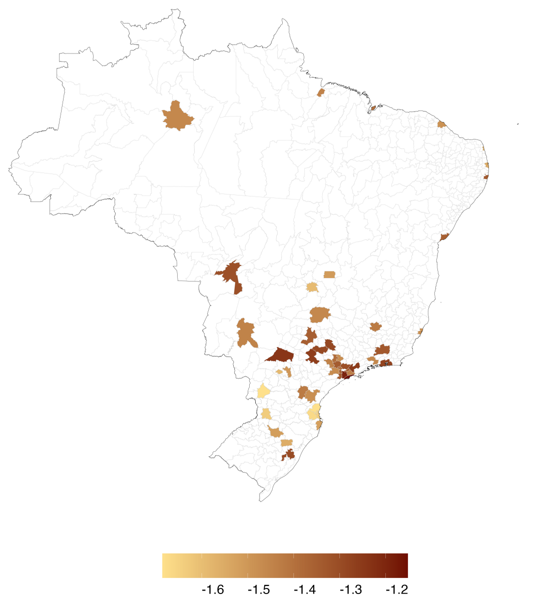
$$w_{hit}^{sl} = w_t^{sl} + \underbrace{\alpha_i^s + \beta_e^s exp_{hit}^s + \beta_{ee}^s (exp_{hit}^s)^2 + \beta_a^s age_{it} + \beta_{aa}^s age_{it}^2}_{\text{sector-specific skill}} + \nu_{it}^{sl}, \tag{E.3}$$

where we assume that the individual-sector skill is a function of age and age squared, and of the number of years of experience individual i has accumulated in sector s from 1994, the first year for which we collect individual employment information until the year they are observed. Hence, by construction, our measure of experience is capped at eight years at the beginning of our analysis period in 2002 and at 17 years in 2011.

We estimate (E.3) on the largest sample of individuals possible drawn from the selected 1,000 labor markets over the 10 years of analysis, i.e. before sampling only 1 million individuals per year. The sample for our main demographic group (white males with at least high school education) contains 15,313,848 observations. The mean value of experience across the 20 sectors and across the 10 years of our sample is 4,7 years, and the median is 3 years.

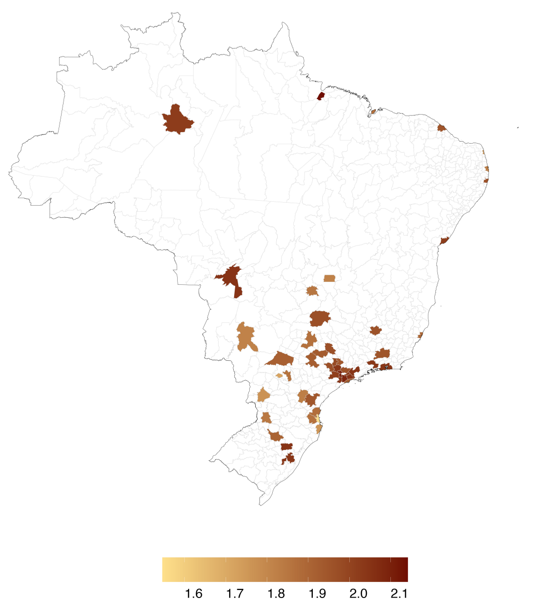
Figures E.1 and E.2 provide maps depicting the geographic distribution of w_t^{sl} in the first and final year of our sample period (2002 and 2011). Since in each microregion-year we observe different w_t^{sl} for each sector s, the figure provides averages weighed by the number of workers in each microregion-sector-year cell.

Figure E.1: Geographic Distribution of Predicted Wages in Sample - 2002



The map presents the log predicted wage (w_t^{sl} from equation E.3) for each microregion in the subsample of 50 largest microregions in the year 2002. Predicted wages vary by municipality, sector, and year and are weighted by the number of workers in the sample in a municipality-sector-year cell. Microregions not included in the sample are shaded white.

Figure E.2: Geographic Distribution of Predicted Wages in Sample - 2011



The map presents the log predicted wage (w_t^{sl} from equation E.3) for each microregion in the subsample of 50 largest microregions in the year 2011. Predicted wages vary by municipality, sector, and year and are weighted by the number of workers in the sample in a municipality-sector-year cell. Microregions not included in the sample are shaded white.

E.2 Connection to Point-Identified Estimation Strategies

E.2.1 Log-Linear Estimation

When agents' expectations are common to all agents i within a location-sector (n, s), the share μ^l_{nst} of workers in (n, s) who chose l is expressed as:

$$\mu^l_{nst} \quad = \quad \frac{\exp\left(\alpha \mathbb{E}[w^l_{st}|\mathcal{J}_{nst}] - \kappa^l_{nt}\right)}{\sum_k \exp\left(\alpha \mathbb{E}[w^k_{st}|\mathcal{J}_{nst}] - \kappa^k_{nt}\right)}. \tag{E.4}$$

The log ratio of the share of migrants to l relative to the share of stayers writes

$$\begin{aligned} \log \frac{\mu^l_{nst}}{\mu^n_{nst}} \quad &= \quad \alpha\left(\mathbb{E}[w^l_{st}|\mathcal{J}_{nst}] - \mathbb{E}[w^n_{st}|\mathcal{J}_{nst}]\right) - \kappa^l_{nt} \\ &= \quad \alpha \Delta w^{ln}_{st} - \kappa^l_{nt} - \alpha \Delta \xi^{ln}_{st}. \end{aligned} \tag{E.5}$$

We estimate (E.5) using 2SLS, using z^{ln}_{st} as instrument for Δw^{ln}_{st}:

$$\log \frac{\mu^l_{nst}}{\mu^n_{nst}} = \alpha \Delta w^{ln}_{st} - \kappa^l_{nt} + u^l_{nst}, \tag{E.6}$$

where u^l_{nst} is the measurement error.

E.2.2 PPML Estimation

When agents' expectations are common to all agents i within a sector s, the expected utility $U_{nst} \equiv \mathbb{E}_\epsilon[\max_l \mathcal{U}^l_{ins}]$ of a representative agent in labor market n, where $\mathcal{U}^l_{ins}$ is defined in (3a), writes:

$$U_{nst} = \log\left(\sum_l \exp\left(\alpha \mathbb{E}[w^l_{st}|\mathcal{J}_{st}] - \kappa^l_{nt}\right)\right). \tag{E.7}$$

The number of migrants y^l_{nst} between locations n and l can be expressed as:

$$\begin{aligned} y^l_{nst} \quad &= \quad \frac{\exp\left(\alpha \mathbb{E}[w^l_{st}|\mathcal{J}_{st}] - \kappa^l_{nt}\right)}{\sum_k \exp\left(\alpha \mathbb{E}[w^k_{st}|\mathcal{J}_{st}] - \kappa^k_{nt}\right)} L_{nst} \\ &= \quad \exp\left(\alpha \mathbb{E}[w^l_{st}|\mathcal{J}_{st}] - \kappa^l_{nt} - U_{nst} + \ln L_{nst}\right) \\ &= \quad \exp\left(\alpha w^l_{st} - \alpha \xi^l_{st} - \kappa^l_{nt} - U_{nst} + \ln L_{nst}\right) \\ &= \quad \exp\left(\lambda^l_{st} + \Gamma_{nst} + \Psi^l_{nt}\right), \end{aligned} \tag{E.8}$$

where $\xi^l_{st} \equiv w^l_{st} - \mathbb{E}[w^l_{st}|\mathcal{J}_{st}]$ and

$$\lambda^l_{st} \equiv \alpha w^l_{st} - \alpha \xi^l_{st}, \qquad \Gamma_{nst} \equiv -U_{nst} + \ln L_{nst}, \qquad \Psi^l_{nt} \equiv -\kappa^l_{nt}.$$

In three steps, we recover the migration elasticity α and migration costs κ^l_{nt}. First, we estimate (E.8) by PPML:

$$y^l_{nst} = \exp\left(\lambda^l_{st} + \Gamma_{nst} + \Psi^l_{nt}\right) + u^l_{nst},$$

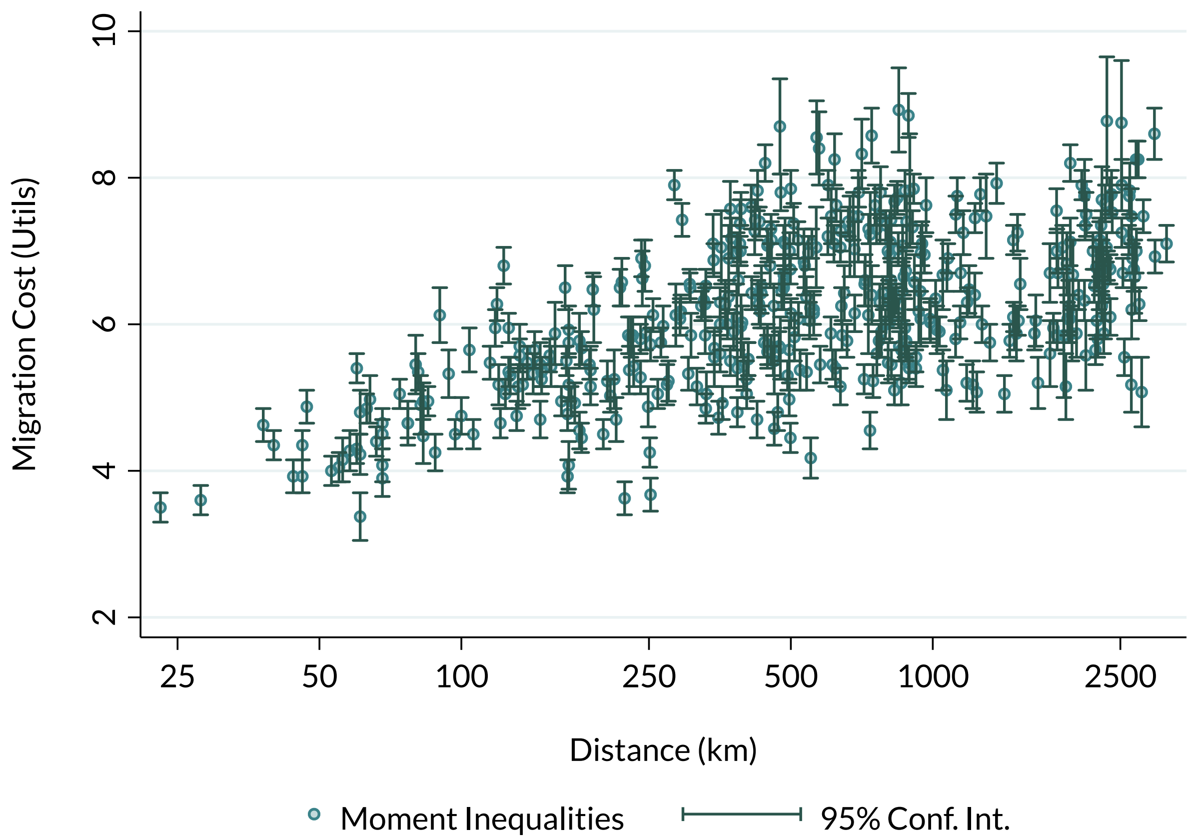
where u^l_{nst} is the measurement error. This provides estimates $\hat{\lambda}^l_{st}$, $\hat{\Gamma}_{nst}$, $\hat{\Psi}^l_{nt}$. Second, we estimate α from the regression of the location-sector-year fixed effect on wages,

$$\hat{\lambda}^l_{st} = \alpha w^l_{st} + u^l_{st},$$

using z^l_{st} as instrument for w^l_{st}, and interpreting the measurement error as $u^l_{st} = -\alpha \xi^l_{st}$. Third, we recover $\hat{\kappa}^l_{nt} = -\hat{\Psi}^l_{nt}$.

E.3 Additional Results

Figure E.3: Migration Costs from Moment Inequalities with Confidence Intervals



This figure displays the main estimates of the preference parameters from our working sample. Each point shows the midpoint of the 95-percent confidence interval for a given bilateral migration cost $\{\kappa^l_{nt}\}$ in the year 2011, expressed in utils, and the associated 95-percent confidence interval.