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

Charly Porcher* Georgetown University

Eduardo Morales* Princeton University

Thomas Fujiwara* Princeton University

February 2025

Abstract

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

JEL Classifications: C13, J61, R23

Keywords: migration, moment inequalities, information frictions

*We thank Treb Allen, Stéphane Bonhomme, Lorenzo Casaburi, Rafael Dix-Carneiro, Dave Donaldson, Cécile Gaubert, Gabriel Kreindler, Melanie Morten, Diego Puga, and James Tybout, for insightful conversations, and seminar participants at Aarhus, UC Berkeley, Carlos III, CEMFI, Chicago-Princeton Conference on Spatial Economics, Cornell, Dartmouth, Johns Hopkins SAIS, MIT, Penn State, Princeton, Stanford, University of Toronto, University of Virginia, University of Zurich, UPF, Yale, and the NBER Summer Institute for helpful comments. We especially thank Milena Almagro for her thoughtful discussion, and Mark Bamba, Juanma Castro-Vincenzi, Pedro Cabra-Acela, and Anaïs Galdin for excellent research assistance. All errors are our own. Emails: charly.porcher@georgetown.edu, ecmorale@princeton.edu, and fujiwara@princeton.edu.

1 Introduction

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

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

We obtain four main results. First, workers generally only have coarse information on location-specific wages. However, those living in areas with better internet access or larger populations are better informed. Second, our estimates of the migration elasticity to expected wages are three times larger than those obtained using common estimation procedures, which place stronger assumptions on workers' information, whereas our migration cost estimates are 21% lower on average. Third, migration rates would increase significantly if workers had full information on location-specific wages. Fourth, researchers assuming workers' information is better than it truly is overestimate the welfare gains from reductions in migration costs.

Our baseline analysis is based on a static model that incorporates expectations on wages, migration costs, amenities, and prices, as well as idiosyncratic preferences, as drivers of workers' location choices.[1] We impose no restriction on wage expectations beyond assuming these are rational; as a result, expectations on wages may vary across workers with different information sets, which we treat as potentially heterogeneous across every worker and unobserved to the researcher. Conversely, for each destination, we assume expectations on migration costs, amenities, and prices are common across workers sharing a prior location of residence,

[1] In Appendix G, we extend our analysis to models with sunk migration costs and forward-looking workers.

but allow them to vary flexibly across these worker groups, and impose no restriction on their rationality. Thus, while wage expectations give rise to individual unobserved heterogeneity, expected migration costs, amenities, and prices are captured by destination-specific effects that may vary across workers depending on their prior location.[2] A second source of unobserved heterogeneity in workers' choices is the presence of idiosyncratic preferences, which we assume are independent across destinations and follow a type I extreme value distribution.

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

Key for our procedure is a new type of moment inequality that we call *bounding* inequality. To derive it, we compare workers' expected utility in any two locations in their choice set, obtaining as a result a conditional moment *equality* that depends on a concave function of the worker's expected utility difference in those two locations.[3] The conditioning set in this moment equality is a covariate vector assumed to belong to the worker's information set. From this equality, we derive moment *inequalities* by bounding the concave function from above by its tangent at any point. Importantly, the resulting inequalities are linear in the worker's expected utility difference between the two locations. We then substitute workers' unobserved wage expectations with the ex post realized wages. While this introduces workers' expectational errors in the moment, the rational expectations assumption implies these are mean zero conditional on any variable in the worker's information set. This property of expectational errors, combined with the linearity of the bounding inequality, leaves this one unaffected. The resulting moment inequality thus depends on the difference in the fixed effects between the two locations being compared, as well as on the utility difference that arises from any wage variation between both locations.[4]

[2] Our model may be estimated separately for worker groups defined by gender, race, or education level, allowing thus expected amenities, migration costs, and prices to vary flexibly across those worker groups.

[3] To derive our inequalities, we require the relative probability of choosing the two locations to be convex in the worker's expected utility difference between them. This requirement holds in multinomial logit models, but also in nested logit models if the locations being compared belong to the same nest. Thus, our moment inequality procedure may be applied to a generalized version of our baseline model in which idiosyncratic preferences are allowed to be correlated across nests of destinations.

[4] More generally, our inequality will depend on the utility difference coming from any choice characteristic (e.g., amenities) that is both observed by the researcher and not absorbed by the included fixed effects.

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

We show theoretically that our inequalities provide bounds on all parameters when the researcher correctly specifies a subset of workers' information sets, and point identify them when such subset coincides with the true information sets. Point identification is thus achieved precisely when maximum likelihood estimators are consistent, implying no loss of identification power may be incurred in this case when using our inequalities.[5] When the researcher only observes a proper subset of workers' information sets, we show in simulations that the maximum likelihood estimates are not only biased but also often outside of the bounds defined by our inequalities. Our simulations also illustrate that, when the researcher misspecifies the content of the worker's information set by assuming that a variable belongs to it when it truly does not, the identified set defined by our moment inequalities may be empty. We use this result to test how accurate workers' wage information is.[6]

We employ our estimator to study internal migration in Brazil. We use data from the *Relação Anual de Informações Sociais* (RAIS), which has information on the wage and the sector and region of work of all formal workers. We estimate our model for the population of white male workers aged 25-64 with at least a high school degree.[7] We define a labor market as a sector-region pair, and study the information workers in our population have on market-by-period wage shifters. These shifters account for all demand and supply factors having a

[5]See corollaries 1 and 3 in Section 3 for more details.

[6]In some misspecified models, the identified set is non-empty while not including the true parameter value. See Molinari (2020) for a discussion of this phenomenon, and Andrews and Kwon (2024) and Kaido and Molinari (2024) for inference procedures in partially identified models robust to misspecification.

[7]These are, respectively, the largest race, gender, and education categories in RAIS.

common impact on the wages of all workers in our population in a market and period. To estimate these shifters, we regress wages on sector-by-region-by-period fixed effects (which equal our shifters of interest) while controlling for worker-by-sector fixed effects (which account for unobserved worker cross-sectoral comparative advantage) and time-varying worker characteristics (e.g., sector-specific experience) with sector-specific coefficients.

Our analysis yields four conclusions. First, workers face substantial information frictions. We reject the common assumption that workers have perfect information on prevalent wages in every labor market. Furthermore, when exploring how finely workers can classify markets on the basis of the previous year's wage shifter, we conclude they can only classify markets into four bins. In particular, we cannot reject that workers can classify each market as being in the top 25% by its previous year's sector-by-region wage shifter, in the 50-75% bracket, in the 25-50% bracket, or in the bottom 25%, but we reject that *every* worker can classify *every* market according to finer partitions. Concurrently, we find that workers' wage information is heterogeneous and that geography plays a key role in driving that heterogeneity. Specifically, we cannot reject that workers in regions with better internet access or larger populations have more precise wage information, or that all workers have more accurate information about wages in markets that are geographically close to their location of residence.

Second, relative to our moment inequalities, estimators common in the migration literature yield smaller estimates of the migration wage elasticity and larger migration cost estimates. Specifically, our approach yields a 95% confidence interval for the elasticity of migration to expected wages centered at 1.5, and does not include the Poisson Pseudo-Maximum Likelihood (PPML) estimate of 0.5. In addition, our migration cost estimates (measured in utility terms) are centered around values 21% lower than the PPML estimates. Therefore, in our setting, standard assumptions on the worker information set drive the researcher to overestimate the role that non-wage factors play in determining workers' location choices.[8]

Third, we quantify workers' individual welfare gains from improved wage information. The 95% confidence interval for the welfare change that results from giving the average worker perfect wage information is $[2.5\%, 4.1\%]$. Importantly, this welfare change is partly driven by an increase in migration. When perfectly informed about wages, the average worker changes locations in two consecutive years with probability between 9 and 14%; when they can only discern quartiles of lagged wage shifters, this probability is between 4 and 7%. Importantly, this relationship between wage information and migration probabilities is not due to workers being too pessimistic about wages in locations other than the place of residence, as workers in our model are always rational, regardless of the information to which they have access.

[8]The PPML estimator we implement is consistent if all workers employed in the same sector have a common information set. The difference between the PPML and moment inequality estimates is thus compatible with our finding that information sets are heterogeneous by worker location.

Fourth, migration costs and information frictions interact in rich ways when determining workers' migration rates. As a result, model-implied welfare gains from reductions in migration costs are sensitive to the researcher's assumptions on workers' information. The 95% confidence interval for the welfare change that results from a 10% decline in migration costs is $[4.2\%, 5.6\%]$ if the worker is fully informed about current sector-by-region wage shifters, but only $[2.2\%, 2.9\%]$ if they only observe the quartiles of the previous year's shifters.

Our paper is related to three strands of the literature. First, it relates to work studying workers' mobility within countries. Our static model incorporates location-specific idiosyncratic preferences and fixed migration costs as Tombe and Zhu (2019) and Morten and Oliveira (2024). In the dynamic extension to our model, we further allow for forward-looking workers and one-time migration costs as Kennan and Walker (2011). Our contribution is to show how to estimate static and dynamic migration models without fully specifying workers' information, and to quantify the impact on model estimates and counterfactual predictions of misspecifying workers' information sets. However, our analysis so far has two limitations: it does not allow workers' location decisions to depend on unobserved wage determinants, and all our counterfactual predictions are partial equilibrium.[9]

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

Third, our paper is related to studies using choice data to identify agents' preferences when their expectations of choice characteristics are rational but unobserved. In the absence of measures of agents' expectations (Manski, 2004), it is common to assume that the researcher observes agents' full information sets (Manski, 1991). A recent approach allows the content of these sets to be partly unobserved by the researcher, but rules out heterogeneity in such content between agents of the same observable type (Traiberman, 2019). Building on Pakes (2010), Ho and Pakes (2014), and Pakes et al. (2015), we allow information sets to vary across agents in unobservable ways, and use moment inequalities to partially identify agents' preferences in models, static or dynamic, with large choice sets and choice-specific fixed

[9] See Fan et al. (2023) for work incorporating flexible beliefs in a model à la Caliendo et al. (2019).

effects.[10] We combine our bounding inequality with the odds-based inequality in Dickstein and Morales (2018) and Dickstein et al. (2023) to obtain tighter bounds than if each of these two inequalities were used separately. When applying our estimator to the study of dynamic models, we combine the bounding inequality with the Euler approach in Morales et al. (2019).

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

2 Model of Migration with Incomplete Information

We model the choice of location for workers in a population defined, e.g., by their demographic characteristics and prior location. Workers in this population are partitioned into S types defined, e.g., by their sector of employment. While all parameters may vary freely across populations, we assume parameters do not vary across types. As shown in Appendix F, the worker's type may be endogenous and chosen simultaneously with their location.

We index types by s or r and workers by i or j within a type. Defining a variable y^l_{is} that equals one if worker i of type s chooses location l (and zero otherwise), we assume

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

where $\mathbb{1}\{A\}$ is an indicator function that equals 1 if A is true, $\mathcal{U}^l_{is} \in \mathbb{R}$ denotes the worker's utility of choosing l, $\mathcal{J}_{is} \in \mathbb{R}^{d_{is}}$ with $d_{is} \geqslant 0$ is the worker's information set, and $\mathbb{E}[\cdot|\mathcal{J}_{is}]$ is a conditional expectation operator reflecting the worker's beliefs.[11]

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

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

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

$$\mathcal{U}^l_{is} = u^l_{is} + \varepsilon^l_{is}, \tag{3a}$$

[10]For a different treatment of unobserved information sets, see the work on Bayes correlated equilibria; e.g., Bergemann and Morris (2013, 2016), Bergemann et al. (2022) and Magnolfi and Roncoroni (2023).

[11]Equation (1) assumes a common choice set for all workers. The estimator described in Section 3 can be adapted to allow for heterogeneous choice sets, if these are partly observed by the researcher and independent of workers' information sets and location-specific utilities.

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

where w_{is}^l is the natural logarithm of the nominal wage worker i of type s would earn if they chose location l and α captures the relative importance of wages in workers' utility. The terms κ^l and ε_{is}^l are the common and idiosyncratic components of all other determinants of utility. For simplicity, we refer to κ^l as location l's amenity, although it will also account for location-specific log prices and, depending on the definition of the population of study, for other determinants of workers' preferences such as migration costs.[12]

Third, defining $\varepsilon_{is} = (\varepsilon_{is}^1, \ldots, \varepsilon_{is}^L)$, we assume that

$$\varepsilon_{is} \subseteq \mathcal{J}_{is}, \tag{4}$$

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

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

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

with $\mathcal{W}_{is}$ including all elements of $\mathcal{J}_{is}$ other than ε_{is}; i.e., $\mathcal{W}_{is} = \mathcal{J}_{is} \backslash \{\varepsilon_{is}\}$. The first equality in equation (5) imposes that a type-s worker has at least as much information as any worker of a different type r about the differences in the wage a type-s worker would earn in different locations. The second equality imposes that the worker's expected wage difference between locations l and l' only depends on the expected difference between type-specific terms. Finally, the third equality imposes that, once we condition on all other elements of the worker's information set, idiosyncratic preferences do not help the worker forecast wages.[13]

The first equality in equation (5) naturally holds if all workers in the population have the same information; i.e., if $\mathcal{J}_{is} = \mathcal{J}_{jr}$ for any i, j, s, and r. It also holds if workers know more about their type-specific wage differences than workers of a different type.[14] Importantly, this equality does not restrict the variation in information across workers of the same type.

The second equality in equation (5) is imposed by data limitations. Our moment inequality procedure permits to flexibly model the information workers have on payoff-relevant variables whose ex-post (or realized) value the researcher either observes or can consistently

[12]When workers differing in their prior location are classified into different populations, they may differ in the value of κ^l in any l and, thus, these parameters will account for origin-by-destination migration costs.

[13]Generally, for any variables x_{is}^l and $x_{is}^{l'}$, we define $\Delta x_{is}^{ll'} \equiv x_{is}^l - x_{is}^{l'}$.

[14]When types correspond to sectors of employment, equation (5) imposes that, e.g., real estate workers know more about differences across locations in real estate wages than healthcare workers, and vice versa.

estimate. Generally, one cannot estimate, for every worker and location, a wage component that is location- and worker-specific. Hence, we must impose that workers ignore their own idiosyncratic location-specific wage shifters when making their location choices. In contrast, equation (5) does not impose any assumption on the information workers have about worker-by-type or type-by-location wage shifters. Specifically, when types correspond to sectors, we impose no assumption on the information workers have about their own sectoral comparative advantage or about sector-by-location specific shocks driving labor demand or supply.[15]

The third equality in equation (5) is an implication of the exogeneity of idiosyncratic shocks often assumed in discrete choice models.[16]

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

$$F_\varepsilon(\varepsilon_{is}|\mathcal{W}_{is}, \mathcal{J}_{jr}) = F_\varepsilon(\varepsilon_{is}|\mathcal{W}_{is}) = F_\varepsilon(\varepsilon_{is}) = \exp\Big(-\sum_{l=1}^{L}\exp(-\varepsilon^l_{is})\Big), \tag{6}$$

for any worker i of type s and worker j of type r. The first equality imposes that a worker's idiosyncratic preferences are independent of all other workers' information sets, including their own idiosyncratic preferences. The second equality imposes that a worker's idiosyncratic preferences are independent of all other elements of their own information sets. The third equality imposes that ε^l_{is} is *iid* across locations and follows a type I extreme value distribution with location parameter equal to zero and scale parameter equal to one.

Equations (1) to (6) are the only model assumptions we impose. Hence, not only do we allow for unobserved heterogeneity in workers' information sets and, as a result, in workers' wage expectations, but we also leave the wage data-generating process unrestricted. Furthermore, equations (2) to (5) imply $\mathbb{E}[\mathcal{U}^{l'}_{is}|\mathcal{J}_{is}] = \kappa^l + \alpha\mathbb{E}[w^l_{is}|\mathcal{W}_{is}] + \varepsilon^l_{is}$ and, thus, we can interpret κ^l as capturing the expectation that all workers in the population have on amenities in l. Consequently, we allow workers to have irrational expectations on amenities, but restrict these expectations to be common across workers in the population. Conversely, equation (2) restricts workers' wage expectations to be rational, but the flexible modeling of information sets allows expected wages in any given location to differ across all workers.

We assume the researcher observes a random sample of workers by type. For all workers, the researcher observes the location choice $y_{is} = (y^1_{is}, \ldots, y^L_{is})$. Additionally, for all types, the researcher observes wage shifters $w_s = (w^1_s, \ldots, w^L_s)$ and a vector $z_s = (z^1_s, \ldots, z^L_s)$, with z^l_s a

[15] By assuming that unobserved (to the researcher) wage shifters are unknown to workers when choosing locations, we rule out the selection mechanism in Roy (1951). An alternative is to follow the procedure in Section 8.2 in Dickstein and Morales (2018) and allow workers to choose locations based on unobserved wage shifters; computational reasons would then force us to limit the number of parameters in the wage equation.

[16] Our model allows for any correlation pattern between $\kappa = (\kappa^1, \ldots, \kappa^L)$ and the wage shifters $w_s = (w^1_s, \ldots, w^L_s)$ of any type s. It is thus consistent with location-specific amenities and wages being correlated.

potential predictor of w_s^l. Alternatively, w_s may not be observed but consistently estimated. We do not assume the researcher observes $\mathcal{W}_{is}$ for any sampled worker.

Only differences between the elements of κ are identified. We thus normalize $\kappa^1 = 0$. The goal of estimation is to recover α and $(\kappa^2, \ldots, \kappa^L)$, and to learn about workers' information. We denote by $\theta \equiv (\theta_\alpha, \theta_2, \ldots, \theta_L)$ the parameter vector with true value $\theta^* \equiv (\alpha, \kappa^2, \ldots, \kappa^L)$. To infer workers' information, we test the null hypothesis that, for a given set of locations, certain wage predictors belong to the information set of all workers in a given group.

3 Estimation Through Moment Inequalities

If the number of choices L is large, θ will be high dimensional. Common moment inequality inference procedures rely on inverting a test at each point in a grid covering the parameter space, complicating their applicability in models with large parameter vectors. We propose a two-step procedure that circumvents these computational challenges and produces a confidence interval for each element of θ individually. In the first step, we compute a confidence interval for θ_α using inequalities that difference out the parameters $\theta_2, \ldots, \theta_L$. In the second step, for each $l = 2, \ldots, L$, we derive inequalities that depend only on θ_α and θ_l, which we combine with the first-step confidence interval for θ_α to obtain a confidence interval for θ_l.[17]

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

3.1 Second-Step Moment Inequalities

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

Both the bounding and the odds-based inequality exploit the same implication of the model described in Section 2. Specifically, equation (1) implies that, for any worker i of type s and any two locations l and l', it holds that

$$(y_{is}^l + y_{is}^{l'})(\mathbb{1}\{\mathbb{E}[\mathcal{U}_{is}^l - \mathcal{U}_{is}^{l'} | \mathcal{J}_{is}] \geqslant 0\} - y_{is}^l) = 0. \tag{7}$$

This equation indicates that, for any worker i of type s who chooses location l or location l'

[17] If u_{is}^l included worker-by-location covariates other than the wage w_{is}^l (e.g., worker-by-location amenities), the first-step inequalities could also be used to compute bounds on the coefficients on those covariates.

(for whom $y^l_{is}+y^{l'}_{is}=1$), they would choose l if and only if their expected utility of choosing l is larger than that of choosing l'; i.e., $y^l_{is}=1$ if and only if $\mathbb{E}[\mathcal{U}^l_{is}-\mathcal{U}^{l'}_{is}|\mathcal{J}_{is}]\geqslant 0$. As equation (7) holds for every worker, it must also hold for the average worker with a particular information set $\mathcal{W}_{is}$ who effectively chooses l or l'. Equations (2) to (6) then imply the following *equality*:

$$\mathbb{E}\left[\frac{\exp(\mathbb{E}[u^l_{is}-u^{l'}_{is}|\mathcal{W}_{is}])}{1+\exp(\mathbb{E}[u^l_{is}-u^{l'}_{is}|\mathcal{W}_{is}])}-y^l_{is}\middle|\mathcal{W}_{is},y^l_{is}+y^{l'}_{is}=1\right]=0,$$

which we can rewrite as

$$\mathbb{E}[y^{l'}_{is}+y^l_{is}(-\exp(-\mathbb{E}[u^l_{is}-u^{l'}_{is}|\mathcal{W}_{is}]))|\mathcal{W}_{is}]=0, \tag{8}$$

where $\exp(-\mathbb{E}[u^l_{is}-u^{l'}_{is}|\mathcal{W}_{is}])$ equals the probability of choosing l' relative to that of choosing l. This equality cannot be used to identify θ due to the weak restrictions our model imposes on the content of $\mathcal{W}_{is}$ for any worker i of type s. However, as shown in sections 3.1.1 and 3.1.2, the convexity of $\exp(-x)$ in x can be exploited to derive inequalities that do not depend on $\mathcal{W}_{is}$ and provide non-trivial bounds on θ.[18] Specifically, the bounding inequalities in Section 3.1.1 exploit the fact that any convex function is bounded from below by any first-order approximation to it, regardless of the approximation point. Conversely, the odds-based inequalities in Section 3.1.2 exploit the fact that, according to Jensen's inequality, the expectation of a convex function is larger than the function of the corresponding expectation.

3.1.1 Bounding Moment Inequalities

Given equation (8) and the convexity of $\exp(-x)$ in x, we can use the first-order approximation to this function around any point $e^{ll'}_{is}$ to derive the following moment inequality:

$$\mathbb{E}[y^{l'}_{is}-y^l_{is}\exp(-e^{ll'}_{is})(1+e^{ll'}_{is}-\mathbb{E}[u^l_{is}-u^{l'}_{is}|\mathcal{W}_{is}])|\mathcal{W}_{is}]\geqslant 0. \tag{9}$$

As this inequality is linear in $\mathbb{E}[u^l_{is}-u^{l'}_{is}|\mathcal{W}_{is}]$, the rationality of workers' expectations implies:

$$\mathbb{E}[y^{l'}_{is}-y^l_{is}\exp(-e^{ll'}_{is})(1+e^{ll'}_{is}-(u^l_{is}-u^{l'}_{is}))|\mathcal{W}_{is}]\geqslant 0. \tag{10}$$

Finally, given any $z_s\subseteq\mathcal{W}_{is}$, the Law of Iterated Expectations implies the following inequality:

$$\mathbb{E}[y^{l'}_{is}-y^l_{is}\exp(-e^{ll'}_{is})(1+e^{ll'}_{is}-(u^l_{is}-u^{l'}_{is}))|z_s]\geqslant 0, \tag{11}$$

[18] The inequalities in sections 3.1.1 and 3.1.2 can be derived if the relative probability of choosing locations l and l' is convex (or concave) in the agent's expected utility difference between them: the relative probability of choosing l' over choosing l becomes more (or less) sensitive to $\mathbb{E}[u^l_{is}-u^{l'}_{is}|\mathcal{W}_{is}]$ as it increases. This property holds in multinomial logit models; also in nested logit models when l and l' belong to the same nest.

which no longer depends on the unobserved information set $\mathcal{W}_{is}$ of any worker. The remainder of this section explains how we use this inequality to derive bounds on $\theta_2, \ldots, \theta_L$.

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

$$\mathbb{m}^{ll'}(z_s, \Delta\theta_{ll'}) \equiv \mathbb{E}[y_{is}^{l'} - y_{is}^{l} \exp(-e_{is}^{ll'})(1 + e_{is}^{ll'} - (\Delta\theta_{ll'} + \alpha\Delta w_s^{ll'}))|z_s], \tag{12}$$

which equals the left-hand side of equation (11) but written as a function of $\Delta\theta_{ll'}$. Theorem 1 establishes a property of this moment when $\Delta\theta_{ll'} = \Delta\kappa^{ll'}$.[19]

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

The proof of Theorem 1 is in Appendix A.1. This theorem implies that, if its conditions hold, the set of values of $\Delta\theta_{ll'}$ for which

$$\mathbb{m}^{ll'}(z_s, \Delta\theta_{ll'}) \geqslant 0 \tag{13}$$

includes $\Delta\kappa^{ll'}$ regardless of the value of z_s, of the locations l and l' being compared, and of how the approximation points $e_{is}^{ll'}$ are chosen. However, which other values of $\Delta\theta_{ll'}$ satisfy equation (13) depends on these approximation points. Appendix B.1 shows that the set of values of $\Delta\theta_{ll'}$ that satisfy the inequality in equation (13) is minimized when:

$$e_{is}^{ll'} = \Delta\theta_{ll'} + \alpha\mathbb{E}[\Delta w_s^{ll'}|z_s, y_{is}^{l} = 1]. \tag{14}$$

To provide intuition on how the inequality implied by equations (12) to (14) may be used to partially identify $\Delta\theta_{ll'}$, we show in Appendix B.1 that this inequality can be written as

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

As this inequality holds for any two locations, we can swap the identity of l and l' and obtain

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

Equations (15) and (16) provide bounds on the amenity difference $\Delta\theta_{ll'}$. These bounds are based on the relative probability with which workers with information on z_s choose l over l',

[19]Note that, given the normalization $\kappa^1 = 0$, it holds that $\Delta\theta_{l1} = \kappa^l$.

$\mathbb{E}[y_{is}^l|z_s]/\mathbb{E}[y_{is}^{l'}|z_s]$. However, the mapping between relative choice probabilities and amenity differences is not straightforward, as a location l may be preferred over a location l' not only because of a higher amenity value in l but also because of higher expected wages in l. That is, to infer $\Delta\theta_{ll'}$ from the relative probability of choosing l over l' one must first net out the effect of the expected wage difference between l and l'. The challenge in doing so is that this expectation depends on the unobserved information set $\mathcal{W}_{is}$ and, thus, is also unobserved. The inequality in equations (12) to (14) uses the fact that one can bound this expectation as $\mathbb{E}[\Delta w_s^{ll'}|z_s, y_{is}^{l'} = 1] \leqslant \mathbb{E}[\Delta w_s^{ll'}|\mathcal{W}_{is}] \leqslant \mathbb{E}[\Delta w_s^{ll'}|z_s, y_{is}^l = 1]$. Using these bounds, one obtains the inequalities in equations (15) and (16). These inequalities are generally strict; thus, $\Delta\theta_{ll'}$ is typically partially identified. Corollary 1 describes a case where $\Delta\theta_{ll'}$ is point identified.

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

We prove Corollary 1 in Appendix B.2. Corollary 1 strengthens the assumptions in Theorem 1 by requiring that $\mathbb{E}[\Delta w_s^{ll'}|z_s] = \mathbb{E}[\Delta w_s^{ll'}|\mathcal{W}_{is}]$. That is, by requiring that z_s contains all the information worker i of type s has about the wage difference between locations l and l'. In this case, equations (15) and (16) jointly hold only when $\Delta\theta_{ll'}$ equals its true value $\Delta\kappa^{ll'}$. Importantly, this is true regardless of how precise workers' wage expectations are.

3.1.2 Odds-based Moment Inequalities

As discussed in Dickstein et al. (2023), equation (8), the convexity of $\exp(-x)$ in x, and the rationality of workers' expectations imposed in equation (2) implies the following inequality:

$$\mathbb{E}[y_{is}^l \exp(-(u_{is}^l - u_{is}^{l'})) - y_{is}^{l'}|\mathcal{W}_{is}] \geqslant 0. \tag{17}$$

Thus, given any $z_s \subseteq \mathcal{W}_{is}$, the Law of Iterated Expectations implies

$$\mathbb{E}[y_{is}^l \exp(-(u_{is}^l - u_{is}^{l'})) - y_{is}^{l'}|z_s] \geqslant 0. \tag{18}$$

The remainder of this section explains how we use this inequality to derive bounds on $\Delta\theta_{ll'}$ for any l and l'. For any two locations l and l' and a random vector z_s, define the moment

$$\mathbb{m}_o^{ll'}(z_s, \Delta\theta_{ll'}) \equiv \mathbb{E}[y_{is}^l \exp(-(\Delta\theta_{ll'} + \alpha\Delta w_s^{ll'})) - y_{is}^{l'}|z_s], \tag{19}$$

which equals the left-hand side of equation (18) but written as a function of the unknown parameter $\Delta\theta_{ll'}$. Theorem 2, from Dickstein et al. (2023), establishes a key property of this moment when evaluated at $\Delta\theta_{ll'} = \Delta\kappa^{ll'}$.

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

The proof of Theorem 2, which follows Dickstein et al. (2023), is in Appendix A.2. This theorem implies that, if its conditions hold, the set of values of $\Delta\theta_{ll'}$ that satisfies

$$\mathbb{m}_o^{ll'}(z_s, \Delta\theta_{ll'}) \geqslant 0 \tag{20}$$

includes $\Delta\kappa^{ll'}$. Appendix B.3 shows that this inequality can be rewritten as

$$\frac{\mathbb{E}[y_{is}^l|z_s]}{\mathbb{E}[y_{is}^{l'}|z_s]}\mathbb{E}[(\exp(\alpha\Delta w_s^{ll'}))^{-1}|z_s, y_{is}^l = 1] \geqslant \exp(\Delta\theta_{ll'}). \tag{21}$$

By swapping the identity of locations l and l' in equation (21), we obtain the inequality

$$\frac{\mathbb{E}[y_{is}^l|z_s]}{\mathbb{E}[y_{is}^{l'}|z_s]}(\mathbb{E}[\exp(\alpha\Delta w_s^{ll'})|z_s, y_{is}^{l'} = 1])^{-1} \leqslant \exp(\Delta\theta_{ll'}). \tag{22}$$

Equations (21) and (22) identify bounds on $\Delta\theta_{ll'}$. To gain intuition on these bounds, consider the case in which $z_s = \mathcal{W}_{is}$, allowing to rewrite equations (21) and (22) as

$$\frac{\mathbb{E}[y_{is}^l|z_s]}{\mathbb{E}[y_{is}^{l'}|z_s]}(\mathbb{E}[\exp(\alpha\Delta w_s^{ll'})|z_s])^{-1} \leqslant \exp(\Delta\theta_{ll'}) \leqslant \frac{\mathbb{E}[y_{is}^l|z_s]}{\mathbb{E}[y_{is}^{l'}|z_s]}\mathbb{E}[(\exp(\alpha\Delta w_s^{ll'}))^{-1}|z_s]. \tag{23}$$

Due to Jensen's inequality, these inequalities generally only partially identify $\Delta\theta_{ll'}$. However, the following corollary describes a case where these bounds coincide, point identifying $\Delta\theta_{ll'}$.

Corollary 2 *Assume equations (1) to (6) hold. Then, the bounds in equations (21) and (22) imply* $\Delta\theta_{ll'} = \Delta\kappa^{ll'}$ *if* $\Delta w_s^{ll'} = \mathbb{E}[\Delta w_s^{ll'}|\mathcal{W}_{is}]$.

We prove Corollary 2 in Appendix B.4. Corollary 2 strengthens the assumptions in Theorem 2 by requiring that workers be fully informed about $\Delta w_s^{ll'}$, $\Delta w_s^{ll'} = \mathbb{E}[\Delta w_s^{ll'}|\mathcal{W}_{is}]$. In this case, the inequality in equation (20) only holds if $\Delta\theta_{ll'}$ equals its true value $\Delta\kappa^{ll'}$. Thus, the odds-based inequality point identifies $\Delta\theta_{ll'}$ if workers forecast wages without error. This is true regardless of the extent to which the researcher observes workers' information sets.

3.1.3 Combining Bounding and Odds-based Moment Inequalities

There are advantages in using our bounding inequality jointly with the odds-based inequality in Dickstein et al. (2023). As stated in theorems 1 and 2, the identified sets defined by these two types of inequality always contain the true parameter value. However, as exemplified by corollaries 1 and 2, these identified sets may differ. As a result, there may be gains from combining both types of inequality in estimation; that is, the intersection of both identified sets may be smaller than each of them individually.

The identified set defined by the odds-based inequality increases in the relevance of the error affecting workers' expectations. Intuitively, these inequalities are convex in the agent's expectational error. Consequently, Jensen's inequality implies that the odds-based moment at any parameter value is larger in expectation the larger the relevance of the expectational error, making the resulting moment inequality weaker. Thus, if workers are poorly informed about wages, the identified set defined by the odds-based inequality will be large, including parameter values quite distinct from the true one. Importantly, this is the case even if the researcher observes the worker's true information set. Conversely, the identified set defined by the bounding inequality is not affected by the worker's expectational error. Intuitively, the bounding inequalities are linear in this error, making it irrelevant for the identified set defined by these inequalities. Thus, even if workers are poorly informed about wages, the resulting identified set will be tight as long as the researcher observes the key variables used by workers to form their expectations. Moreover, as indicated in Corollary 1, the bounding inequality point identifies the parameter of interest if the researcher observes the worker's information set. This is important because full observability of the worker's information set is generally necessary for standard estimators (e.g., maximum likelihood) to be consistent; thus, a researcher that uses bounding moment inequalities in her analysis will not suffer from any loss of identification power whenever those standard estimators are consistent.

The identified set defined by the bounding inequality increases in the relevance of the variables that enter the worker's information set but the researcher does not observe. Intuitively, the optimal approximation point in equation (14) is the best approximation to the worker's true expected utility difference between locations l and l' that the researcher can build with the variables she observes. The larger the mismatch between this point and the agent's true expected utility difference between l and l', the weaker the bounding inequality is.[20] Conversely, the identified set defined by the odds-based inequality is not affected by the extent to which the worker's information set is observed by the researcher. Even if workers have access to information the researcher cannot observe, the identified set defined by the odds-based inequality will be tight if the worker's true information set allows them to forecast wages with little error. Moreover, as indicated in Corollary 2, the odds-based inequality point identifies the parameter of interest if workers do not experience expectational errors.

In sum, the bounding inequality being robust to workers' expectational errors and the odds-based inequality being robust to the presence of unobserved variables that belong to

[20]More specifically, the larger the distance between $e_{is}^{ll'}$ and $\mathbb{E}[u_{is}^{l} - u_{is}^{l'}|\mathcal{W}_{is}]$, the worse the first-order approximation is and, as a result, the larger the value of the left-hand side of the inequality in equation (9).

the worker's true information set implies there often are gains from simultaneously using both types of inequality. We illustrate this in our simulation in Section 4. In settings such as ours in which the choice set is large and the worker's utility depends on choice-specific fixed effects, the bounding inequality has the extra advantage that it is linear in the utility difference between any two locations; see equation (11). As shown in Section 3.2, this allows to combine multiple bounding inequalities in order to difference out the choice-specific fixed effects, making our two-step estimation procedure feasible.

3.2 First-Step Moment Inequalities

Similarly to how we derive the bounding inequality in equation (11), we derive

$$\mathbb{E}[y_{jr}^{l'}(y_{is}^{l'} - y_{is}^{l}\exp(-e_{is}^{ll'})(1 + e_{is}^{ll'} - (u_{is}^{l} - u_{is}^{l'})))|z_s, z_r] \geqslant 0, \tag{24a}$$
$$\mathbb{E}[y_{is}^{l}(y_{jr}^{l} - y_{jr}^{l'}\exp(-e_{jr}^{l'l})(1 + e_{jr}^{l'l} - (u_{jr}^{l'} - u_{jr}^{l})))|z_s, z_r] \geqslant 0, \tag{24b}$$

for any locations l and l', types s and r, random vectors z_s and z_r, and approximation points $(e_{is}^{ll'}, e_{jr}^{l'l}) \subseteq \mathcal{J}_{is} \cup \mathcal{J}_{jr}$. Loosely speaking, the moment in equation (24a) compares the utility of choosing l over l' for a worker i of type s, and that in equation (24b) compares the utility of choosing l' over l for a worker j of type r. If the approximation points in these two moments coincide, the sum of the inequalities in equations (24a) and (24b) results in

$$\mathbb{E}[y_{is}^{l}y_{jr}^{l} + y_{is}^{l'}y_{jr}^{l'} - y_{is}^{l}y_{jr}^{l'}\exp(-e_{isjr}^{ll'})(2 + 2e_{isjr}^{ll'} - \theta_\alpha(\Delta u_{is}^{ll'} + \Delta u_{jr}^{l'l}))|z_s, z_r] \geqslant 0, \tag{25}$$

where $e_{isjr}^{ll'}$ denotes the common approximation point. As all workers, regardless of their type, have a shared valuation of the amenities in every location, the amenity differences between locations l and l' cancel when adding both moments in equation (24); thus, we can substitute $\Delta u_{is}^{ll'} + \Delta u_{jr}^{l'l} = \alpha(\Delta w_s^{ll'} + \Delta w_r^{l'l})$ in equation (25).[21] The remainder of this section explains how we use the inequality in equation (25) to derive bounds on θ_α. For any locations l and l', types s and r, vectors z_s and z_r, and scalar variable $e_{isjr}^{ll'}$, we define the moment

$$\begin{gathered}\mathbb{M}^{ll'}(z_s, z_r, \theta_\alpha) \equiv \\ \mathbb{E}[y_{is}^{l}y_{jr}^{l} + y_{is}^{l'}y_{jr}^{l'} - y_{is}^{l}y_{jr}^{l'}\exp(-e_{isjr}^{ll'})(2 + 2e_{isjr}^{ll'} - \theta_\alpha(\Delta w_s^{ll'} + \Delta w_r^{l'l}))|z_s, z_r],\end{gathered} \tag{26}$$

which equals the left-hand side of equation (25) but written as a function of θ_α. Theorem 3 establishes a property of the moment in equation (26) when evaluated at $\theta_\alpha = \alpha$.

[21] Ho and Pakes (2014) apply a similar strategy to difference out choice-specific fixed effects in a model without individual-by-choice idiosyncratic preferences.

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

The proof of Theorem 3 is in Appendix A.3. This theorem implies that, if its conditions hold, the set of values of θ_α for which

$$\mathbb{M}^{ll'}(z_s, z_r, \theta_\alpha) \geqslant 0 \tag{27}$$

includes α regardless of (z_s, z_r), of the locations l and l', and of how the approximation points $e^{ll'}_{isjr}$ are chosen. However, the set values of θ_α other than α that satisfy equation (27) depends on these approximation points. Appendix B.5 shows this set is minimized when:

$$e^{ll'}_{isjr} = \theta_\alpha \mathbb{E}[0.5(\Delta w^{ll'}_s + \Delta w^{l'l}_r)|z_s, z_r, y^l_{is} y^{l'}_{jr} = 1]. \tag{28}$$

As shown in Appendix B.5, equations (27) and (28) jointly imply that

$$\frac{\mathbb{E}[y^l_{is} y^{l'}_{jr}|z_s, z_r]}{\mathbb{E}[0.5(y^l_{is} y^l_{jr} + y^{l'}_{is} y^{l'}_{jr})|z_s, z_r]} \leqslant \exp(\theta_\alpha \mathbb{E}[0.5(\Delta w^{ll'}_s + \Delta w^{l'l}_r)|z_s, z_r, y^l_{is} y^{l'}_{jr} = 1]). \tag{29}$$

This inequality yields a lower bound on θ_α if its right-hand side is increasing in this parameter, and vice versa. Intuitively, if type-s workers are likely to choose location l whereas type-r workers are likely to choose l', as represented by a high value of the ratio on the left-hand side of equation (29), and type-s workers expect their wage to be higher in l than in l', and vice versa, as represented by a positive value of $\mathbb{E}[0.5(\Delta w^{ll'}_s + \Delta w^{l'l}_r)|z_s, z_r, y^l_{is} y^{l'}_{jr} = 1]$, then θ_α cannot be too low. Conversely, if both worker types are still likely to choose locations l and l', respectively, but now type-s workers expect their wage to be lower in l than in l', and vice versa, as represented by a negative value of $\mathbb{E}[0.5(\Delta w^{ll'}_s + \Delta w^{l'l}_r)|z_s, z_r, y^l_{is} y^{l'}_{jr} = 1]$, then θ_α cannot be too high. Generally, the parameter θ_α is partially identified. The following corollary describes when moment inequalities of the type in equation (29) point identify θ_α.

Corollary 3 *Assume equations* (1) *to* (6) *hold,* $e^{ll'}_{isjr} \subseteq \mathcal{J}_{is} \cup \mathcal{J}_{fr}$*,* $(z_s, z_r) \subseteq \mathcal{W}_{is} \cup \mathcal{W}_{jr}$*,* $\Delta\kappa^{ll'} = 0$ *and* $\mathbb{E}[\Delta w^{ll'}_s|z_s] = \mathbb{E}[\Delta w^{ll'}_s|\mathcal{W}_{is}] = \mathbb{E}[\Delta w^{l'l}_r|z_r] = \mathbb{E}[\Delta w^{l'l}_r|\mathcal{W}_{jr}]$*. Then, moment inequalities of the type in equation* (29) *can point identify* θ_α*.*

The proof of Corollary 3 is in Appendix B.6. To understand this corollary, it is useful to compare it to Corollary 1. The conditions listed in Corollary 3 are more restrictive. The reason is that, unless these conditions are satisfied, there is a loss of identification power coming from having to impose the common approximation point in equation (28) on each of the inequalities in equation (24). When the conditions in Corollary 3 are satisfied, the approximation point in equation (28) is optimal for each of the two inequalities in equation (24) (i.e., it coincides with the approximation point in equation (14) for both moments) and thus no loss of identification power occurs when deriving the inequality in equation (25).

3.3 Using the Inequalities for Estimation

For estimation, we use a set of unconditional moment inequalities that we derive from the conditional ones introduced in sections 3.1 and 3.2. We describe here how we derive unconditional bounding moment inequalities from the conditional ones described in Section 3.1.1. In practice, when computing a confidence interval for the amenity term $\Delta\theta_{ll'}$, we combine these unconditional bounding inequalities with unconditional odds-based moment inequalities derived in a similar way from the conditional ones in Section 3.1.2. In Appendix B.7, we describe how we rely on the conditional inequalities in Section 3.2 to derive the unconditional inequalities we use to compute a confidence interval for θ_α.

We implement the following steps to derive $k = 1, \ldots, K$ unconditional moment inequalities from the conditional one described in Section 3.1.1. First, we choose a scalar $\Delta z_s^{ll'} \subseteq z_s$ that is correlated with $\Delta w_s^{ll'}$ and that we will use in all K inequalities. Second, for each k, we choose a subset $[\underline{z}_k, \overline{z}_k]$ of the support of $\Delta z_s^{ll'}$ and an integer d_k. We then build

$$\mathbb{E}[(y_{is}^{l'} - y_{is}^{l} \exp(-e_{is}^{ll'})(1 + e_{is}^{ll'} - (\Delta\theta_{ll'} + \alpha\Delta w_s^{ll'})))g_k(\Delta z_s^{ll'})] \geqslant 0, \tag{30}$$

where the term in parenthesis coincides with the moment function in equation (12) and

$$g_k(\Delta z_s^{ll'}) = \mathbb{1}\{\underline{z}_k < \Delta z_s^{ll'} \leqslant \overline{z}_k\}|\Delta z_s^{ll'}|^{d_k}. \tag{31}$$

The inequality in equation (30) is implied by that in equation (13) regardless of the choice of predictor $\Delta z_s^{ll'}$, interval limits $\underline{z}_k$ and $\overline{z}_k$, and exponent d_k. In practice, we fix a $q \in \mathbb{N}$ and map $\underline{z}_k$ and $\overline{z}_k$ to consecutive elements of the vector of q-quantiles of the distribution of $\Delta z_s^{ll'}$ across types and location pairs. E.g., if $q = 2$ and $d_k = 0$ for all k, we use $K = 2$ inequalities with instruments $g_1(\Delta z_s^{ll'}) = \mathbb{1}\{\Delta z_s^{ll'} \leqslant \text{med}(\Delta z_s^{ll'})\}$ and $g_2(\Delta z_s^{ll'}) = \mathbb{1}\{\text{med}(\Delta z_s^{ll'}) > \Delta z_s^{ll'}\}$, which split all observations depending on whether $\Delta z_s^{ll'}$ is above or below median. If $q = 4$, similar instruments will split observations according to the quartile $\Delta z_s^{ll'}$ belongs to. More generally, the larger q is, the larger the number of unconditional inequalities we use.

In practice, computing a confidence interval (CI) for $\Delta\theta_{ll'}$ requires computing first a CI for θ_α. To compute a 95% CI for $\Delta\theta_{ll'}$, we first compute 96% CIs for $\Delta\theta_{ll'}$ conditional on each value of θ_α in a 99% CI for this parameter. We denote these as $\hat{\Theta}^l_{.96}(\theta_\alpha)$. We then compute the 95% CI for θ_l as the union of $\hat{\Theta}^l_{.96}(\theta_\alpha)$ for each value of θ_α in its 99% CI. All our CIs are computed following the moment selection procedure in Andrews and Soares (2010).[22]

4 Properties of Moment Inequalities: Simulation

This section uses simulations to illustrate properties of the moment inequalities introduced in Section 3. Three insights emerge. First, consistent with theorems 1 to 3, when the conditions in those theorems are satisfied, we obtain intervals that contain the true parameter values even when workers imperfectly forecast wages and the researcher imperfectly observes agents' information sets. Second, when the researcher misspecifies agents' information sets, the maximum likelihood (ML) estimator is inconsistent and often not within the bounds defined by our moment inequalities. Third, our inequalities may yield empty confidence sets when the researcher incorrectly assumes that certain covariates belong to agents' information sets, demonstrating the potential of our estimator to test for the true content of such sets.

4.1 Simulation Set-up

Workers choose between three locations $l = \{1, 2, 3\}$ according to the model in Section 2. We simulate data for $6,000,000$ workers, each of them of a different type, allowing us to index observations by s.[23] We set the wage coefficient to $\alpha = 1$ and the location-specific amenities to $\kappa^1 = \kappa^2 = 0$ and $\kappa^3 = 1$. While the estimator described in Section 3 is valid for any stochastic process for wages and any specification of workers' information sets, we need to set out these model aspects to generate the model-implied choice for all sampled workers.

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

$$w_s^l = z_{1s}^l + z_{2s}^l + z_{3s}^l, \tag{32}$$

with z_{ks}^l independent across l, s, and k, and distributed uniformly with mean μ_k^l and support of length $2\sigma_k$.[24] We assume worker s observes (z_{1s}^l, z_{2s}^l) for all l. Consequently, for all s and l,

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

[23] We do not aim to illustrate the statistical properties of our inference procedure (see Andrews and Soares, 2010) but to characterize the bounds defined by the inequalities introduced in Section 3. To this end, the large number of workers is useful, as it limits the impact of simulation noise on our results. We set $L > 2$ to illustrate that our inequalities apply to multinomial settings; larger values of L result in tighter bounds for θ_α (as the number of inequalities of the type in equation (27) increases in the number of possible location pairs) at the expense of more time needed to compute CIs for the larger number of parameters $(\theta_2, \ldots, \theta_L)$.

[24] We set $\mu_1^l = \mu_3^l = 0$ for all l and $(\mu_2^1, \mu_2^2, \mu_2^3) = (0, -0.5, -1)$; thus, mean wages decline in order from $l = 1$ to $l = 3$. In terms of the dispersion, we set $\sigma_2 = 4$ and present results for different values of σ_1 and σ_3.

$\mathbb{E}[w_s^l|\mathcal{W}_s] = z_{1s}^l + z_{2s}^l$, and z_{3s}^l equals the expectational error worker s makes when forecasting w_s^l. Thus, the larger the value of σ_3, the larger the variance of the expectational error.

We assume the researcher observes (y_s^l, w_s^l, z_{2s}^l) for every worker s and location l. Therefore, for all s and l, z_{1s}^l is a variable used by worker s when forming their expectations about w_s^l but not observed by the researcher. Thus, the larger the value of σ_1, the larger the role in workers' expectations of variables that are unobserved to the researcher.

4.2 Simulation Results

Table 1 presents the main simulation results. We consider cases that differ in the value of σ_1 and σ_3. The former determines the dispersion in z_{1s}^l and, thus, the relevance of unobserved (to the researcher) variables that belong to the worker's information set. The latter determines the dispersion in z_{3s}^l and, thus, the relevance of payoff-relevant variables that workers do not observe. Section 4.2.1 discusses our CIs for the wage coefficient θ_α, displayed in Table 1 in the column labeled *First Step* and computed using the inequalities introduced in Section 3.2. Section 4.2.2 describes CIs for θ_2 and θ_3, displayed in the columns labeled *Second Step* and computed using the inequalities introduced in Section 3.1. Appendix C.1 describes the unconditional moment inequalities we use in this simulation exercise.

4.2.1 Confidence Intervals for the Wage Coefficient

Cases 1 and 2 share the feature that $\sigma_1 = 0$ and, thus, the researcher observes the agent's full information set—as $z_{1s}^l = 0$ for every l, the agent's information set only includes z_{2s}^l, which is the wage predictor used by the researcher. In these cases, we obtain CIs for θ_α that are tight around its true value. This is related to two aspects of our setting. First, we build first-stage inequalities separately for each pair of locations; thus, as $\kappa^1 = \kappa^2$ in our simulation, the inequality with location indices equal $l = 1$ and $l' = 2$ (or vice versa) verifies that $\Delta\kappa^{ll'} = 0$. Furthermore, when building the inequality corresponding to any location pair l and l', we match each worker s with a worker r such that $\mathbb{E}[\Delta w_s^{ll'}|\Delta z_{2s}^{ll'}]$ is close to $\mathbb{E}[\Delta w_r^{l'l}|\Delta z_{2r}^{ll'}]$. As a result, consistent with Corollary 3, the inequalities used to compute the bounds on θ_α in Table 1 are close to point identifying this parameter whenever $\mathbb{E}[w_s^l|\mathcal{W}_s] = \mathbb{E}[w_s^l|\Delta z_{2s}^{ll'}]$ for all l and s. This extra condition is met precisely when $\sigma_1 = 0$.[25]

Cases 3 and 4 share the feature that $\sigma_1 > 0$. Thus, the researcher only observes part of the agent's information set—the true information set is (z_{1s}^l, z_{2s}^l) for every l, but the researcher

[25] In Table C.3 in Appendix C.4, we present results for other values of $(\kappa^1, \kappa^2, \kappa^3)$. The CI for θ_α is tight if $\Delta\kappa^{ll'} = 0$ for at least two locations l and l'. In Table C.1 in Appendix C.2, we show that the CI becomes wider as the mean difference between $\mathbb{E}[\Delta w_s^{ll'}|z_{2s}]$ and $\mathbb{E}[\Delta w_r^{l'l}|z_{2r}]$ for matched workers s and r increases.

Table 1: Simulation Results - Moment Inequality Confidence Intervals

Case	σ_1	σ_3	z_s^l	First Step	Second Step		
				θ_α	Mom. Ineq.	θ_2	θ_3
1	0	0	z_{2s}^l	[1 , 1.02]	Bounding	[0 , 0]	[1 , 1]
					Odds-based	[0 , 0]	[1 , 1]
					Both	[0 , 0]	[1 , 1]
2	0	1	z_{2s}^l	[1 , 1.01]	Bounding	[0 , 0]	[1 , 1]
					Odds-based	[-0.33 , 0.32]	[0.68 , 1.33]
					Both	[0 , 0]	[1 , 1]
3	1	0	z_{2s}^l	[0.82 , 1.29]	Bounding	[-0.31 , 0.31]	[0.70 , 1.30]
					Odds-based	[0 , 0]	[1 , 1.01]
					Both	[0 , 0]	[1 , 1.01]
4	1	1	z_{2s}^l	[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^l	[0.87 , 0.87]	Bounding	[-0.05 , -0.01]	[0.85 , 0.88]
					Odds-based	$\varnothing$	$\varnothing$
					Both	$\varnothing$	$\varnothing$

The true parameter values are $\alpha = 1$, $\kappa^2 = 0$, and $\kappa^3 = 1$. The column θ_α contains 95% CIs based on the estimator described in Section 3.2. The columns θ_2 and θ_3 contain 95% CIs based on the estimators described in Section 3.1. The rows labeled *Bounding* use the inequalities in Section 3.1.1; those labeled *Odds-based* use the inequalities in Section 3.1.2; and those labeled *Both* combine both inequalities. CIs are computed using the moment selection procedure in Andrews and Soares (2010). See Appendix C.1 for more details.

only observes z_{2s}^l. By relying on only a subset of the agent's information set when building the moment inequalities, the CI for θ_α becomes wider.

In case 5, the researcher wrongly assumes the agent has perfect information on wages. The resulting CI is very tight without including the true parameter value; it only includes $\theta_\alpha = 0.87$. This illustrates a problematic situation for the researcher using moment inequalities, as such researcher may wrongly conclude that $\alpha = 0.87$ and that the inequalities are very tight around the truth (see Molinari, 2020; Andrews and Kwon, 2024). In our setting, as shown in Table C.2 in Appendix C.3, the resulting confidence set becomes empty as we increase the number of instruments used to form our inequalities.

4.2.2 Confidence Intervals for Amenities

Consistently with Corollary 1, the bounding inequalities point identify θ_2 and θ_3 when the agent's information set is fully observed by the researcher; i.e., when $\sigma_1 = 0$, as in cases 1 and 2. When $\sigma_1 > 0$, as in cases 3 and 4, there are variables the agent knows but the researcher does not observe and, as a result, the CIs built using the bounding inequalities alone include parameter values in addition to the true ones. As a comparison of cases 1 and 2, or cases 3 and 4, illustrates, this holds regardless of how large the agent's expectational errors are.[26]

The length of the CIs on θ_2 and θ_3 defined by the odds-based inequalities increases in the importance of the worker's expectational errors; i.e., increases in σ_3, as a comparison of cases 1 and 2, or cases 3 and 4, illustrates. When workers make no expectational errors (i.e., when $\sigma_3 = 0$) and the first-step CI for θ_α equals its true value α, the odds-based inequalities point identify the amenity parameters θ_2 and θ_3, as predicted by Corollary 2. When $\sigma_3 = 0$ but the CI for θ_α includes values other than α, the odds-based moment inequalities may still point identify the amenities (as in cases 1 and 3), but will not do so always.[27]

Since the bounding inequalities are insensitive to agents' expectational errors (i.e., insensitive to σ_3) and the odds-based inequalities are partially insensitive to agents having information the researcher does not observe (i.e., partially insensitive to σ_1), there are advantages from combining both types of inequalities in estimation (see Section 3.1.3). Cases 2 and 3 show that, when either $\sigma_1 = 0$ or $\sigma_3 = 0$, combining bounding and odds-based inequalities point identifies θ_2 and θ_3, although, when considered in isolation, neither of these two inequalities point identifies these parameters in both cases.

Case 4 is likely the most empirically relevant: the agent's information set is partly unobserved (i.e., $\sigma_1 > 0$) and the agent predicts wages with error (i.e., $\sigma_3 > 0$). Our estimator still yields CIs that contain the true parameter values. In this particular case, the odds-based inequalities are redundant: the combined CIs are larger than those obtained from the bounding inequalities alone.

Case 5 shows the bounding inequalities may fail to produce empty confidence intervals when the researcher wrongly assumes workers have complete information. Conversely, the odds-based inequalities alone, or when used jointly with the bounding inequalities, produce empty CIs for θ_2 and θ_3 even when the CI for θ_α is non-empty.

4.2.3 Maximum Likelihood Estimates

Table 2 reports ML estimates. Given a wage predictor z_s^l, we compute the ML estimator of $(\theta_\alpha, \theta_2, \theta_3)$ assuming that z_s^l is all the information worker s has on w_s^l; that is,

$$\underset{(\theta_\alpha,\theta_2,\theta_3)}{\operatorname{argmax}} \left\{ \sum_{s=1}^{S} \sum_{l=1}^{3} \mathbb{1}\{y_s^l = 1\} \ln \left(\frac{\exp(\theta_l + \theta_\alpha \mathbb{E}[w_s^l | z_s^l])}{\sum_{l'=1}^{3} \exp(\theta_{l'} + \theta_\alpha \mathbb{E}[\Delta w_s^{l'} | z_s^{l'}])} \right) \right\} \qquad \text{with } \theta_1 = 0. \tag{33}$$

[26] All CIs in Table 1 are computed using the approximation points in equations (14) and (28). In Table C.4 in Appendix C.5, we show the CIs become wider when using other approximation points. How we compute these points thus does not affect the validity of our inequalities, but it may affect how tight these are.

[27] For example, in unreported results, we observe that the confidence intervals for θ_2 and θ_3 defined by the odds-based inequalities include values other than the true ones when $\sigma_1 = 2$ and $\sigma_3 = 0$.

Table 2: Simulation Results - Maximum Likelihood Estimator

Case	σ_1	σ_3	z_s^l	α	κ^2	κ^3
1	0	0	z_{2s}^l	1	0	1
2	0	1	z_{2s}^l	1	0	1
3	1	0	z_{2s}^l	0.91	0	0.92
4	1	1	z_{2s}^l	0.91	0	0.92
5	0	1	w_s^l	0.87	-0.03	0.87

The true parameter values are $\alpha = 1$, $\kappa^2 = 0$, and $\kappa^3 = 1$. Estimates are computed according to equation (33). Given the large sample size, unreported standard errors are always smaller than 0.001.

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

A comparison of the estimates in tables 1 and 2 yields three conclusions. First, when the ML estimator identifies the true parameter values (as in cases 1 and 2), the CIs defined by the bounding inequalities either include only the true parameter values (as it is the case for θ_2 and θ_3) or are very tight around them (as it is the case for θ_1). Thus, very little identification power is lost when using our moment inequality estimator instead of the ML estimator in those cases in which the latter is consistent. Second, when the ML estimator does not identify the true parameter values (as in cases 3 to 5), our moment inequality estimator still yields CIs that contain those values. Moreover, as illustrated by case 3, the CIs produced by our moment inequality estimator may not include the corresponding ML estimates of some of the parameters; e.g., for θ_2 and θ_3.

5 Empirical Application

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

5.1 Data

Our main data source is the *Relação Anual de Informações Sociais* (RAIS), an administrative dataset that includes information on workers and establishments in the Brazilian formal labor market. We use the establishment's location (microregion, the closest equivalent to a commuting zone in the Brazilian administrative map) and sector (industry) to define labor markets, and we measure workers' annual wages. By using wages aggregated over the year, our wage measure includes information on work hours as well as on the number of days during the year each worker was employed.

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

To ensure we observe a large number of individuals per market, we focus on 1,000 labor markets consisting of all combinations of the 50 microregions (out of 558) and 20 sectors (out of 51) with the largest total employment reported in RAIS. We obtain the data we use in estimation by randomly sampling one million individuals per year among those employed in the 1,000 labor markets of interest. Appendix D provides more details on the RAIS data and the construction of our sample, and reports summary statistics on migration rates.[28]

5.2 Estimation of Model Parameters

In Section 5.2.1, we detail the implementation of the moment inequality estimator. In Section 5.2.2, we discuss our estimates and compare them to those obtained using other estimators.

5.2.1 Implementation of Moment Inequalities

We estimate the parameters of the model described in Section 2, with the type s of each worker defined by their sector. While we assume that the wage coefficient α is common to all sampled workers, we let the location-specific fixed effects in the vector κ vary by year t

[28] Our moment inequalities are valid if the researcher observes a subset of the markets workers choose from; see footnote 11. The informal sector and all microregion-sector pairs not included in the analysis may thus still belong to the worker's choice set. However, our estimates are based on how workers located in one of the 50 largest microregions compare two regions among these 50. Thus, when interpreting our estimates, one should think mostly of urban migrants evaluating other high-density regions. However, nothing prevents performing a different analysis with rural areas or other demographic groups.

and the worker's prior location. Thus, we accommodate for unobserved expectations about migration costs, amenities, and price levels that may vary over time and between workers with different locations of origin. For simplicity, we refer to these origin-by-destination-by-year effects as *amenities.*

Key variables entering our moment inequalities are the differences between any two locations l and l' in the location-by-sector-by-time wage shifters w^l_{st}. To estimate these shifters, we use data on observed wages. Specifically, we assume the log wage a worker would have obtained if employed in a particular sector, region, and year can be expressed as:

$$w^{sl}_{it} = w^l_{st} + \underbrace{\alpha^s_i + \beta^s_e exp^s_{it} + \beta^s_{ee}(exp^s_{it})^2 + \beta^s_a age_{it} + \beta^s_{aa} age^2_{it}}_{\text{sector-specific skill}} + \nu^{sl}_{it}, \tag{34}$$

where exp^s_{it} denotes the number of years of employment of worker i in sector s prior to year t, age_{it} denotes the age of worker i at t, and ν^{sl}_{it} is an unobserved term. Thus, in addition to the labor market-specific term w^l_{st}, we allow wages to depend on a worker-by-sector-by-year term and a residual that varies by sector, location, worker, and year. We model the worker-by-sector-by-year term as the sum of an individual-by-sector fixed effect and a function of the worker's age and sector-specific experience.[29]

Given equation (34), it holds that, for any locations l and l', $\Delta w^{ll'}_{it} = \Delta w^{ll'}_{st} + \Delta\nu^{ll'}_{it}$ and, thus, equation (5) implies that $\mathbb{E}[\nu^{sl}_{it}|\mathcal{J}_{it}] = 0$. We impose no assumption on the information workers have on shifters w^l_{st}, which account for supply and demand factors that impact the wages of all sampled workers in a labor market, or on their sector-specific skill. In particular, at the time of choosing their labor market of employment for a year t, we allow workers to have information on a worker-by-sector specific term that the researcher cannot observe; that is, on the term α^s_i. This selection of *sectoral* labor markets based on unobserved determinants of wages is important in, e.g., Dix-Carneiro (2014). In contrast, we impose the assumption that the selection of *local* labor markets is not driven by worker-by-location terms unobserved to the researcher. This assumption is consistent with the findings in Kennan and Walker (2011), who allow for a permanent worker-specific location-match component in wages, but conclude that the estimated effect of this component is negligible.[30]

In addition to having a measure of wage shifters $w_{st} = (w^1_{st}, \ldots, w^L_{st})$, which we denote in

[29] We measure workers' sector-specific experience using information from 1993 onward. As all coefficients in equation (34) are indexed by s, we estimate them by running separate regressions for each of the 20 sectors in our analysis. These regressions are run on our sample prior to extracting the set of 1 million individuals per year that we use in our moment inequality estimation. The median R^2 across these regressions is 0.83.

[30] Some location choices may be driven by worker-by-location-by-year shocks; e.g., job offers. We conjecture one can substitute the assumption $\mathbb{E}[\nu^{sl}_{it}|\mathcal{J}_{it}] = 0$ by the assumption $\mathbb{E}[\nu^{sl}_{it}|\mathcal{J}_{it}] = \nu^{sl}_{it}$ and still derive inequalities compatible with our flexible treatment of workers' information on w^l_{st}. Deriving these inequalities requires generalizing to a multinomial setting the estimator in Section 8.2 of Dickstein and Morales (2018).

the following simply as *wages*, our inequalities require a wage predictor z_{st}^l for each sector, region, and year. To build these predictors, we implement the following procedure. First, we fix $b \in \mathbb{N}$ and calculate the vector of b-quantiles of the distribution of w_{st-1}^l across sectors and locations. Second, we build as many "bins" as intervals can be constructed using two consecutive elements of this vector of quantiles. Third, we identify the bin to which each market belongs. Fourth, and finally, we compute z_{st}^l as the average wage in year $t-1$ across all labor markets that belong to the same bin as w_{st-1}^l. Note that, as b increases, the wage predictors become closer on average to the $t-1$ wages. Thus, a larger b can be interpreted as workers having more precise wage information. Specifically, if $b = 2$, our wage predictor assumes workers can determine if lagged wages in a labor market are above or below the median. If $b = 4$, workers can determine the quartile of the $t-1$ wage distribution to which a labor market belongs. As $b \to \infty$, z_{st}^l and w_{st-1}^l coincide and, thus, workers know every market's lagged wages. We provide additional implementation details in Appendix E.1.

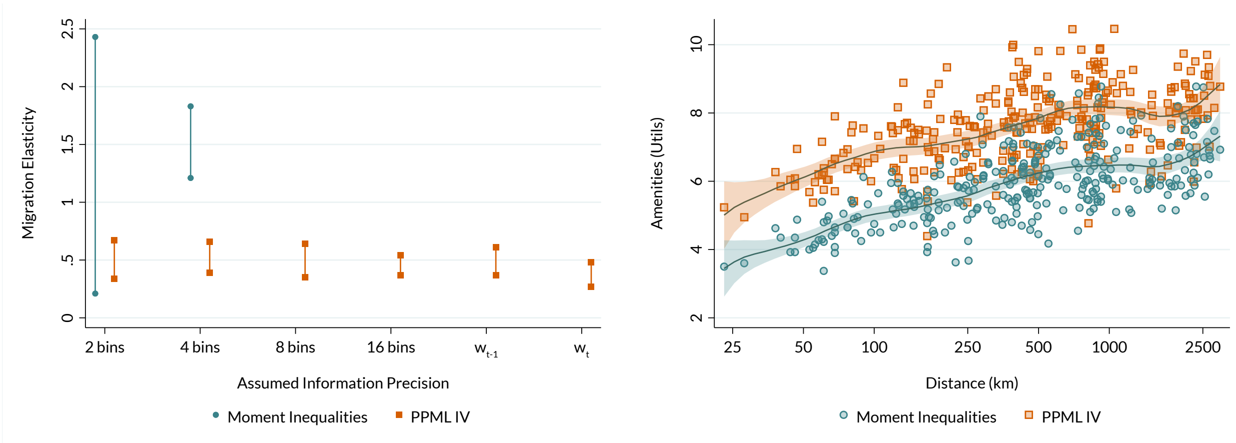
5.2.2 Estimation Results

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

For comparison, we also include 95% CIs computed using a two-step PPML-IV estimator (see Artuç and McLaren, 2015). As discussed in Appendix E.3, this estimator yields point estimates of α at the expense of assuming that all workers in the same sector in a period t (regardless of their location of residence) have the same information set and, consequently, the same wage expectations. This is a stronger assumption than the one required for our moment inequalities to bound α, which requires the researcher to specify a (possibly different) variable that belongs to every worker's information set, but does not restrict the additional information each worker may have, which may vary flexibly across workers and

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

(a) Migration Elasticity (95% conf. int.) (b) Amenities (95% conf. int. midpoint)



Panel (a) reports 95% CIs for α. The blue circles delimit the moment inequality CIs. The absence of circles for certain cases reflects that these CIs are empty. The orange squares mark the PPML-IV CIs. In panel (b), the blue circles indicate the midpoints of the moment inequality 95% CI for κ^l_{nt}, for $t = 2011$, and the orange squares indicate the PPML-IV estimates. The fit lines are kernel-weighted local polynomial estimates, with the shaded area representing 95% CIs.

labor markets.[31] The PPML-IV estimator yields CIs for the wage coefficient that generally do not overlap with the main CI generated by our moment inequalities: while the PPML-IV estimator yields CIs between 0.3 and 0.6, the lower bound in our preferred moment inequality CI is 1.21. Relative to our moment inequality estimator, the PPML-IV estimator thus underestimates the value workers assign to the expected monetary returns of migration.

Second-step estimates: amenities. Panel (b) in Figure 1 illustrates the moment inequality estimates of κ^l_{nt} for $t = 2011$ and all origin n and destination l locations in our sample. Specifically, this panel displays midpoints of the 95% moment inequality CI for each amenity term κ^l_{nt}.[32] For comparison, we also display PPML-IV estimates of these amenities. Although we estimate each parameter κ^l_{nt} without imposing any restriction on their variability, our estimates tend to increase in the distance between locations n and l, consistently with these parameters accounting for migration costs in our model. The differences in levels between the PPML-IV and the moment inequality estimates are substantial, the latter being on average 21% smaller than the former. Moreover, if we convert migration costs into their log-wage equivalents by dividing them by the estimates of the wage coefficient α obtained

[31] E.g., the moment inequality CI that uses quartiles of lagged wages as wage predictor is valid if workers, even within the same sector, location, and period, have different information, as long as all workers can at least classify labor markets into quartiles on the basis of lagged market wages.

[32] Appendix E.2 displays the corresponding CIs. We report estimates for the 292 origin-destination pairs with enough observed migration events to yield both PPML-IV and moment inequality estimates.

by each estimation method, we find that the moment inequality estimates are 74% smaller. In sum, estimation procedures commonly used in the migration literature yield estimates of migration costs or, more generally, of the relative importance of non-wage variables in migration decisions that are significantly larger than our moment inequality estimates.

5.3 Tests of Information Heterogeneity

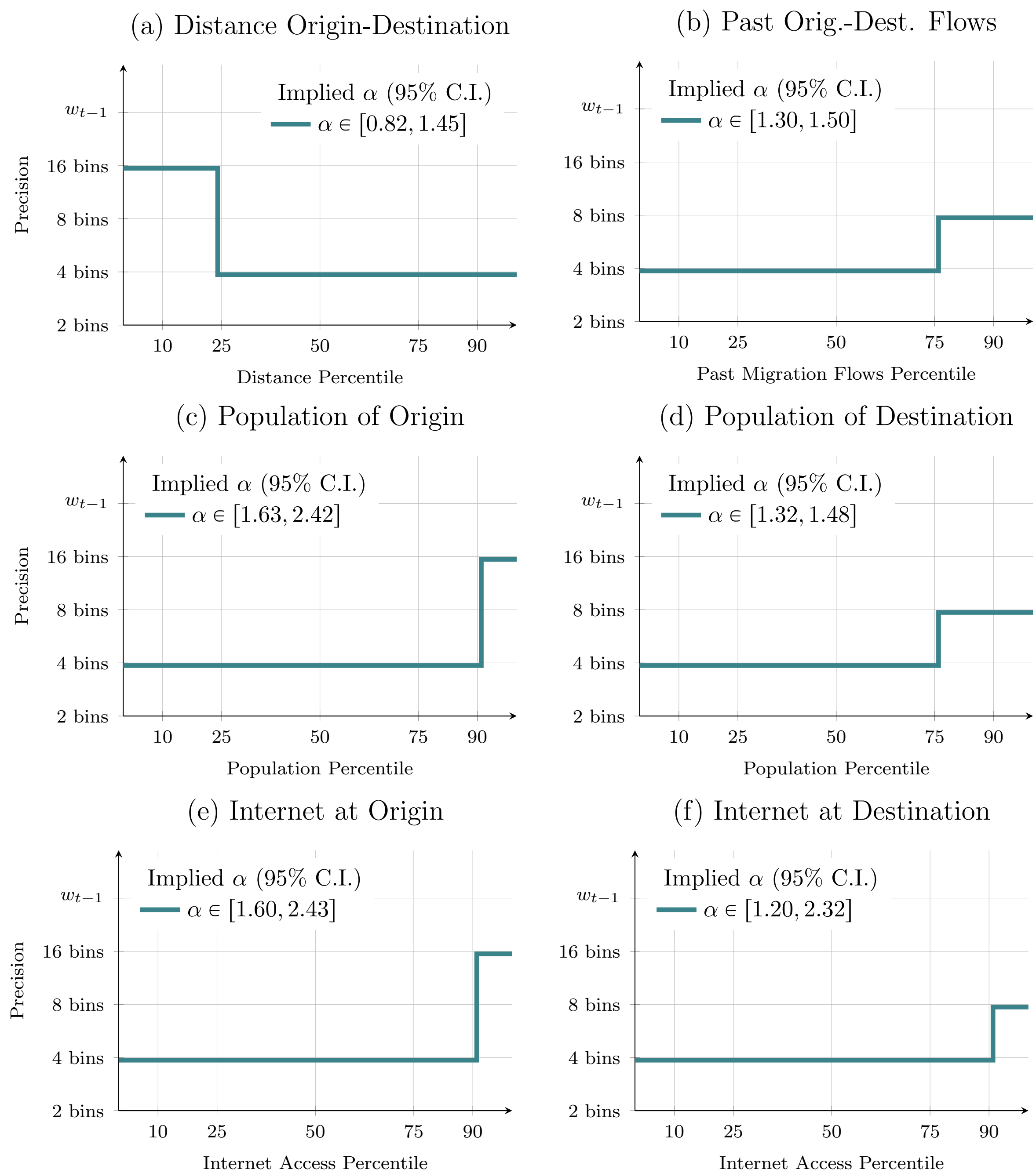
As panel (a) in Figure 1 shows, the moment inequality 95% CI for α is nonempty when we assume workers know the quartile to which lagged wages belong, but is empty when we assume *all* workers can classify *all* markets into eight (or more) bins. However, it is possible that some workers are more informed than others or that workers have more information about some markets than others.

We explore here whether observed migration choices are consistent with certain workers having more precise information about wages in some, but not all, markets. We do so by checking whether the moment inequality 95% CI for α remains non-empty when we assume that a group of workers has extra wage information on a group of labor markets. Specifically, we consider worker groups defined by the population and internet penetration in their prior location of residence, and groups of labor markets defined by their distance and past migration flows from the worker's location, and by their population and internet penetration.[33]

We implement the same testing procedure for each dimension of heterogeneity. First, we classify workers (or markets) into six intervals delimited by the 10th, 25th, 50th, 75th, and 90th percentiles of the distribution of workers (or markets) along the corresponding dimension. We then order these intervals according to the direction along which we hypothesize information may be more precise. For example, we classify workers into intervals depending on the internet penetration in the location of residence, and order these from higher to lower internet penetration. Consistently with our finding in panel (a) in Figure 1, we start from a baseline information set according to which all workers can classify all markets into quartiles, and test whether workers in the first interval can further classify markets into eight bins. In practice, this translates into wage predictors with different levels of precision in different combinations of workers and markets, affecting the values of the instruments in equation (31) and thus changing our inequalities. If the resulting 95% CI for α is empty, we reject that assumption and end the testing. If it is not empty, we increase the level of precision to 16 bins on that first interval and perform a new test. Calling B_j the maximum precision level

[33] The distance between markets equals the geodesic distance between their centroids. Past migration flows are measured as the total number of workers recorded in RAIS as having migrated between any two locations in the three years prior to our sample period (1999-2001). The population of each location is computed as the total employment in RAIS in 1999-2001. The measure of internet penetration in each location equals the average share of households with broadband internet access between 2007 and 2011. See Appendix D.

Figure 2: Testing for Heterogeneous Information Sets



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

tested and not rejected for the jth interval, the next iteration maintains B_j on that interval and searches for the maximum level of precision in the interval $j+1$, up to precision B_j.

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

5.4 Counterfactuals

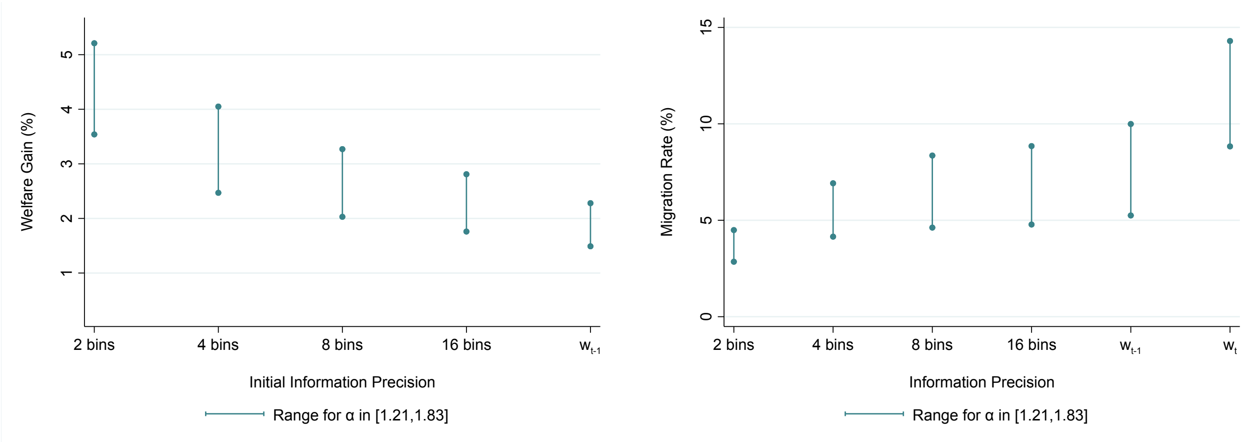
To illustrate the relevance of the estimates described above, we quantify the impact of changes in workers' information sets and migration costs on their location choices and expected utility. We study workers' individual responses, omitting the impact that changes in information and costs may have on migration flows through changes in wages and prices.

While we could compute the moment inequality estimates above without imposing any assumption on the stochastic process of wages, specifying this element of the model is needed to determine workers' migration choices. Hence, to compute the results in this section, we assume wages follow an AR(1) process with sector- and location-specific drifts, and estimate this process using observed wages in our sample of 50 regions and 20 sectors over 2002-2011. We find wages are strongly serially correlated, with a persistence estimate of 0.93.

Additionally, while our estimation procedure makes it computationally feasible to obtain CIs for all model parameters, computing model predictions that account for uncertainty in

Figure 3: Effects of Providing Full Information About Wages

(a) Welfare Changes (b) Migration Rates



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

all parameter estimates is costly, as it requires building a multidimensional grid that spans all CIs and evaluating our model at each point in that grid. Instead, we consider all points in the CI for the wage coefficient α, but calibrate amenities by regressing the midpoint of the moment inequality CIs (see panel (b) in Figure 1) on a constant and logarithmic distance.

We compute all model implications for a set of one million individuals randomly drawn from the 2002 empirical distribution of workers across the 50 locations and 20 sectors in our sample. We simulate these workers' choices during 2002-2011 for 100 simulation draws of the wage process. We then report average outcomes across all workers and simulations.

5.4.1 Changes in Information Sets

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

Panel (a) in Figure 3 shows that the gains from improving workers' information can be substantial. For workers whose initial information only allows them to determine whether lagged market wages are above or below the median, welfare gains are between 3.5 and 5.2%, with the highest gains in this interval corresponding to the model that sets the wage coefficient α at the highest value within its 95% CI. The gains remain significant for workers who were initially better informed. Even if all workers observed perfectly lagged wages, a hypothesis we reject in panel (a) in Figure 1, the gains from observing contemporaneous wages would still range between 1.5 and 2.3%.

Panel (b) in Figure 3 reports migration rates for workers with different information. Migration rates increase steeply in the precision of the worker's wage information. They are below 5% for workers with the coarsest information set we consider, and between 9 and 14% when information is complete.[34] While it is hardly surprising that better-informed workers have higher welfare, as better-informed workers choose more often the location that yields higher ex post utility, it is not obvious that better-informed workers will have higher migration rates. Workers in our model are rational, and thus, when acquiring additional wage information, expected wages go up for certain workers and locations and down for others in such a way that average expectations do not change. The reason why average migration rates change while average expectations do not is that a worker's migration probabilities are a nonlinear function of the worker's expectations.

5.4.2 Reducing Migration Costs

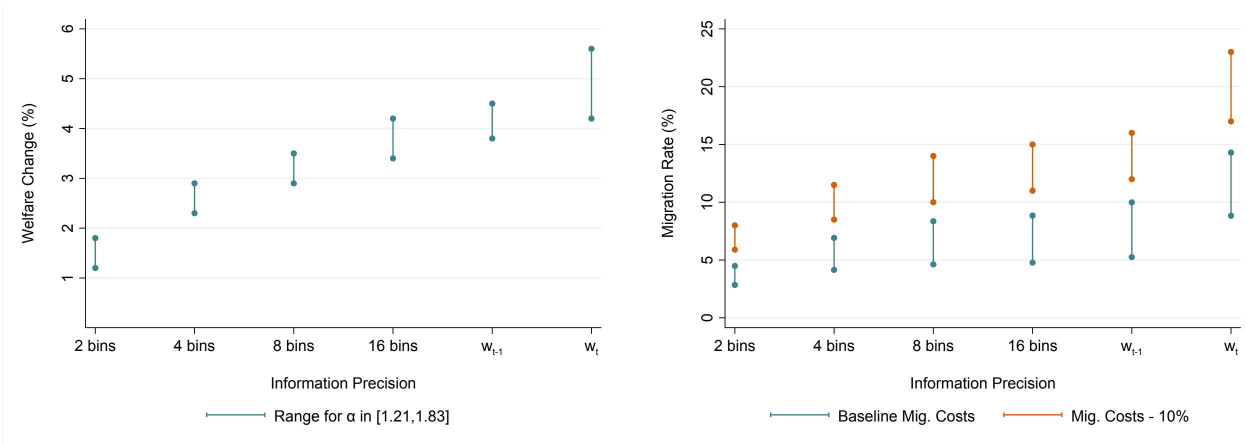
As shown in Bryan and Morten (2019) and Morten and Oliveira (2024), reducing physical barriers to geographic mobility is an important policy lever to alleviate spatial misallocation. However, the benefits of reducing migration costs may depend on whether agents are well-informed about the economic opportunities in different regions. In this section, we evaluate how the effect of reductions in migration costs depends on workers' wage information. Specifically, for several information sets, we compute the predictions of our model for a 10% reduction in our calibrated migration costs.

Panel (a) in Figure 4 reveals that the welfare gains from a 10% reduction in migration costs increase with the precision of workers' wage expectations. When workers are fully informed, the welfare gains range from 4.2 to 5.7%, depending on the estimate of α. When workers can only discern whether lagged wages in a location are above or below their median, the same reduction in migration costs only yields 1.2 to 1.8% welfare gains.

[34] When assuming all workers know the quartile to which lagged wages belong, which is the strongest informational assumption we tested without rejecting it (see Section 5.2.2), our model predicts a migration rate between 4.2 and 6.9%, which includes the average migration rate of 6.8% observed in the sample; see Appendix D.2. This illustrates that our estimated model fits the baseline migration elasticity in the data.

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

(a) Welfare Changes (b) Migration Rates



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

Panel (b) in Figure 4 illustrates the increases in migration rates from reducing migration costs at each information level. Migration rates increase significantly for all information levels, and more so in relative terms for workers with a lower level of information precision. However, those larger increases in mobility have a higher rate of mistakes when the information precision is low, leading to the lower welfare gains reported in panel (a).

6 Conclusion

We introduce a new moment inequality method to measure the impact of migration costs and information frictions on workers' location decisions. Our method allows workers' information sets to be unobserved by the researcher and to vary flexibly between workers, and migration costs to vary flexibly across pairs of origin and destination locations. Applying our method to data on the internal migration of formal workers in Brazil, we obtain four main results. First, workers have heterogeneous information on location-specific wages. In particular, gravity forces play an important role in determining the precision of workers' wage information. Second, accounting for this rich heterogeneity in information sets alters the mapping from observed location choices and wages to workers' preferences. More specifically, our wage preference estimates are three times larger than those obtained using common estimation procedures, and our migration cost estimates are, on average, 21% lower. Third, our estimated model predicts that providing wage information to workers results in increases in both migration rates and welfare. Fourth, relative to a setting in which workers have perfect information on destination characteristics, policies that reduce migration costs by, for example, improving transportation infrastructure are less effective in improving worker welfare when workers are imperfectly informed about destination characteristics.

The two-step moment inequality estimator we introduce may be used more generally to estimate multinomial discrete choice models when the choice set is large, payoffs are parameterized with choice-specific fixed effects, and information sets are unobserved to the researcher and potentially heterogeneous across any two agents. This type of model may be suitable to study, for example, student decisions over which schools to apply to, or patient decisions over which hospital to attend.

References

Akerman, Anders, Edwin Leuven, and Magne Mogstad, "Information Frictions, Internet and the Relationship between Distance and Trade," *American Economic Journal: Applied Economics*, 2022, *14* (1), 133–163. [29]

Andrews, Donald W. K. and Soonwoo Kwon, "Misspecified Moment Inequality Models: Inference and Diagnostics," *Review of Economic Studies*, 2024, *91* (3), 45–76. [3, 20]

Andrews, Donald W.K. and Gustavo Soares, "Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection," *Econometrica*, 2010, *78*, 119–157. [18, 20]

Andrews, Isaiah, Jonathan Roth, and Ariel Pakes, "Inference for Linear Conditional Moment Inequalities," *The Review of Economic Studies*, 2023, *90* (6), 2763–2791. [18]

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. [25]

Baseler, Travis, "Hidden Income and the Perceived Returns to Migration: Experimental Evidence from Kenya," *American Economic Journal: Applied Economics*, 2023, *15* (4), 321–352. [5]

Beine, Michel, Simone Bertoli, and Jesús Fernández-Huertas Moraga, "A Practitioners' Guide to Gravity Models of International Migration," *The World Economy*, 2016, *39* (4), 496–512. [29]

Bergemann, Dirk and Stephen Morris, "Robust Predictions in Games With Incomplete Information," *Econometrica*, 2013, *81* (4), 1251–1308. [6]

_ and _ , "Bayes Correlated Equilibrium and the Comparison of Information Structures in Games," *Theoretical Economics*, 2016, *11* (2), 487–522. [6]

_ , Benjamin Brooks, and Stephen Morris, "Counterfactuals with Latent Information," *American Economic Review*, 2022, *112* (1), 343–368. [6]

Bergman, Peter, Eric Chan, and Adam Kapor, "Housing Search Frictions: Evidence from Detailed Search Data and a Field Experiment," *mimeo*, May 2020. [5]

_ , Raj Chetty, Stefanie DeLuca, Nathaniel Hendren, Lawrence F Katz, and Christopher Palmer, "Creating Moves to Opportunity: Experimental Evidence on Barriers to Neighborhood Choice," *mimeo*, August 2023. [5]

Bryan, Gharad and Melanie Morten, "The Aggregate Productivity Effects of Internal Migration: Evidence from Indonesia," *Journal of Political Economy*, 2019, *127* (5), 2229–2268. [31]

_ , 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. [5]

Bugni, Federico A., Ivan A. Canay, and Xiaoxia Shi, "Inference for Functions of Partially Identified Parameters in Moment Inequality Models," *Quantitative Economics*, 2016, *8* (1), 1–38. [18]

Caliendo, Lorenzo, Maximiliano Dvorkin, and Fernando Parro, "Trade and Labor Market Dynamics: General Equilibrium Analysis of the China Trade Shock," *Econometrica*, 2019, *87* (3), 741–835. [5]

De la Roca, Jorge and Diego Puga, "Learning by Working in Big Cities," *Review of Economic Studies*, January 2017, *84* (1), 106–142. [29]

Dickstein, Michael J. and Eduardo Morales, "What do Exporters Know?," *The Quarterly Journal of Economics*, 2018, *133* (4), 1753–1801. [6, 8, 24]

_ , Jihye Jeon, and Eduardo Morales, "Patient Costs and Physicians' Information," *mimeo*, 2023. [36, 3, 6, 9, 12, 13]

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

Fan, Jingting, Sungwan Hong, and Fernando Parro, "Learning and Expectations in Dynamic Spatial Economies," *mimeo*, 2023. [5]

Glaeser, Edward L. and David C. Maré, "Cities and Skills," *Journal of Labor Economics*, 2001, *19* (2), 316–342. [29]

Ho, Katherine and Ariel Pakes, "Hospital Choices, Hospital Prices, and Financial Incentives to Physicians," *American Economic Review*, 2014, *104* (12), 3841–3884. [5, 15]

Kaido, Hiroaki and Francesca Molinari, "Information Based Inference in Models with Set-Valued Predictions and Misspecification," *mimeo*, January 2024. [3]

_ , _ , and Jorg Stoye, "Confidence Intervals for Projections of Partially Identified Parameters," *Econometrica*, 2019, *4* (87), 1397–1432. [18]

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

Kennan, John and James R. Walker, "The Effect of Expected Income on Individual Migration Decisions," *Econometrica*, 2011, *79* (1), 211–251. [5, 24]

Magnolfi, Lorenzo and Camilla Roncoroni, "Estimation of Discrete Games with Weak Assumptions on Information," *The Review of Economic Studies*, 2023, *90* (4), 2006–2041. [6]

Manski, Charles F., "Nonparametric Estimation of Expectations in the Analysis of Discrete Choice Under Uncertainty," in William Barnett, James Powell, and George Tauchen, eds., *Nonparametric and Semiparametric Methods in Econometrics and Statistics*, Cambridge: Cambridge University Press, 1991. [5]

_ , "Measuring Expectations," *Econometrica*, 2004, *5* (9), 1329–1376. [5]

Molinari, Francesca, "Microeconometrics with Partial Identification," in Steven N. Durlauf, Lars Peter Hansen, James J. Heckman, and Rosa L. Matzkin, eds., *Handbook of Econometrics, Volume 7A*, Elsevier, 2020, pp. 355–486. [3, 20]

Morales, Eduardo, Gloria Sheu, and Andrés Zahler, "Extended Gravity," *The Review of Economic Studies*, 2019, *86* (6), 2668–2712. [6]

Morten, Melanie and Jaqueline Oliveira, "The Effects of Roads on Trade and Migration: Evidence from a Planned Capital City," *American Economic Journal: Applied Economics*, 2024, *16* (2). [5, 31]

Munshi, Kaivan, "Social Networks and Migration," *Annual Review of Economics*, 2020, *12* (Volume 12, 2020), 503–524. [29]

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]

Roy, A. D., "Some Thoughts on the Distribution of Earnings," *Oxford Economic Papers, New Series*, 1951, *3* (2), 135–146. [8]

Tombe, Trevor and Xiaodong Zhu, "Trade, Migration and Productivity: A Quantitative Analysis of China," *American Economic Review*, 2019, *109* (5), 1843–1872. [5]

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

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

A Proofs

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

A.1 Proof of Theorem 1

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

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

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

$$(y^l_{is} + y^{l'}_{is})(\mathbb{1}\{\Delta\kappa^{ll'} + \alpha\mathbb{E}[\Delta w^{ll'}_s | \mathcal{W}_{is}] + \Delta\varepsilon^{ll'}_{is} \geqslant 0\} - y^l_{is}) = 0. \tag{A.1}$$

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

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

which implies the following moment equality

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

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

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

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

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

Consider the alternative moment,

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

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

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

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

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

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

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

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

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

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

A.2 Proof of Theorem 2

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

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

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

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

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

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

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

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

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

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

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

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

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

A.3 Proof of Theorem 3

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

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

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

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

which, after some algebra, implies

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

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

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

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

$$\mathbb{E}[y^{l'}_{is}y^{l'}_{jr}+y^l_{is}y^{l'}_{jr}\exp(-e^{ll'}_{is})(-(1+e^{ll'}_{is})+\Delta\kappa^{ll'}+\alpha\mathbb{E}[\Delta w^{ll'}_s|\mathcal{W}_{is}])|\mathcal{W}_{is},\mathcal{W}_{jr}]\geqslant 0. \tag{A.7}$$

Consider the alternative moment,

$$\mathbb{E}[y^{l'}_{is}y^{l'}_{jr}+y^l_{is}y^{l'}_{jr}\exp(-e^{ll'}_{is})(-(1+e^{ll'}_{is})+\Delta\kappa^{ll'}+\alpha\Delta w^{ll'}_s)|\mathcal{W}_{is},\mathcal{W}_{jr}]. \tag{A.8}$$

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

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

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

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

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

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

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

$$\mathbb{E}[y^{l'}_{is}y^{l'}_{jr}+y^l_{is}y^{l'}_{jr}\exp(-e^{ll'}_{is})(-(1+e^{ll'}_{is})+\Delta\kappa^{ll'}+\alpha\Delta w^{ll'}_s)|\mathcal{W}_{is},\mathcal{W}_{jr}]\geqslant 0. \tag{A.9}$$

This moment inequality is one of the two that we will combine to obtain that in equation (27). To obtain the second moment inequality, we start from

$$y_{is}^{l}(y_{jr}^{l}+y_{jr}^{l'})(\mathbb{1}\{\Delta\kappa^{l'l}+\alpha\mathbb{E}[\Delta w_{r}^{l'l}|\mathcal{W}_{jr}]+\Delta\varepsilon_{jr}^{l'l}\geqslant 0\}-y_{jr}^{l'})=0, \tag{A.10}$$

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

$$\mathbb{E}[y_{is}^{l}y_{jr}^{l}+y_{is}^{l}y_{jr}^{l'}\exp(-e_{jr}^{l'l})(-(1+e_{jr}^{l'l})+\Delta\kappa^{l'l}+\alpha\Delta w_{r}^{l'l})|\mathcal{W}_{is},\mathcal{W}_{jr}]\geqslant 0. \tag{A.11}$$

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

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

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

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

Charly Porcher
Georgetown University

Eduardo Morales
Princeton University

Thomas Fujiwara
Princeton University

B Additional Derivations

Appendix B.1 derives expressions appearing in Section 3.1.1. Appendix B.2 proves Corollary 1. Appendix B.3 derives expressions appearing in Section 3.1.2. Appendix B.4 proves Corollary 2. Appendix B.5 derives expressions appearing in Section 3.2. Appendix B.6 proves Corollary 3. Appendix B.7 shows how we use inequalities to identify the parameter θ_α.

B.1 Second-Step Bounding Inequalities: Additional Derivations

Derivation of equation (14). Consider points $e^{ll'}_{is}$ such that $e^{ll'}_{is} = h^{ll'}(z_s, \Delta\theta_{ll'})$ for some function $h^{ll'}(\cdot)$. We compute the function in equation (14) by finding the value of $h^{ll'}(z_s, \Delta\theta_{ll})$ that minimizes the moment in equation (12) at each value of $\Delta\theta_{ll'}$. Specifically, given z_s and $\Delta\theta_{ll'}$, the first-order condition of the moment in equation (12) with respect to $h^{ll'}(z_s, \Delta\theta_{ll})$ is

$$\mathbb{E}[y^l_{is}(h^{ll'}(z_s, \Delta\theta_{ll'}) - (\Delta\theta_{ll'} + \alpha\Delta w^{ll'}_s))|z_s] = 0,$$

or, equivalently, $\mathbb{E}[h^{ll'}(z_s, \Delta\theta_{ll'}) - (\Delta\theta_{ll'} + \alpha\Delta w^{ll'}_s)|z_s, y^l_{is} = 1] = 0$. Solving for $h^{ll'}(z_s, \Delta\theta_{ll'})$, we obtain the expression in equation (14).

Derivation of equation (15). Equations (12) to (14) imply the following inequality

$$\mathbb{E}[y^{l'}_{is} - y^l_{is}\exp(-(\Delta\theta_{ll'} + \alpha\mathbb{E}[\Delta w^{ll'}_s|z_s, y^l_{is} = 1]))(1 - \alpha(\Delta w^{ll'}_s - \mathbb{E}[\Delta w^{ll'}_s|z_s, y^l_{is} = 1]))|z_s] \geqslant 0.$$

We can simplify this inequality as

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

or equivalently,

$$\mathbb{E}[y^{l'}_{is}|z_s] \geqslant \exp(-(\Delta\theta_{ll'} + \alpha\mathbb{E}[\Delta w^{ll'}_s|z_s, y^l_{is} = 1]))$$

$$\times\,(\mathbb{E}[y^l_{is}|z_s] - \alpha(\mathbb{E}[\Delta w^{ll'}_s|z_s, y^l_{is} = 1]\mathbb{E}[y^l_{is}|z_s] - \mathbb{E}[\Delta w^{ll'}_s|z_s, y^l_{is} = 1]\mathbb{E}[y^l_{is}|z_s])).$$

Eliminating terms that cancel each other, we obtain the inequality in equation (15); i.e.,

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

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

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

Rearranging terms, we obtain the inequality in equation (16).

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

Equations (12) to (14) imply the following moment inequality:

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

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

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

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

$$\mathbb{E}[\mathbb{E}[y^l_{is}|\mathcal{W}_{is}]\exp(-(\Delta\theta_{ll'} + \alpha\mathbb{E}[\Delta w^{ll'}_s|z_s, y^l_s = 1])) - \mathbb{E}[y^{l'}_{is}|\mathcal{W}_{is}]|z_s] \geqslant 0. \tag{B.2}$$

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

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

Given equation (6), we can further rewrite

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

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

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

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

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

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

$$\exp(\Delta\kappa^{ll'} - \Delta\theta_{ll'}) - 1 \geqslant 0 \qquad \Leftrightarrow \qquad \Delta\kappa^{ll'} \geqslant \Delta\theta_{ll'}. \tag{B.3}$$

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

$$\exp(\Delta\kappa^{l'l} - \Delta\theta_{l'l}) - 1 \geqslant 0 \qquad \Leftrightarrow \qquad \Delta\kappa^{ll'} \leqslant \Delta\theta_{ll'}. \tag{B.4}$$

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

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

Derivation of equation (21). Equations (19) and (20) imply the following inequality

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

We can rewrite this inequality as

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

Rearranging terms, we obtain the expression in equation (21); i.e.,

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

Derivation of equation (22). Swapping the indices l and l' in equation (21), we obtain

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

Rearranging terms, we immediately obtain the inequality in equation (22).

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

Equations (19) and (20) imply the following moment inequality:

$$\mathbb{E}[y_{is}^{l}\exp(-(\Delta\theta_{ll'}+\alpha\Delta w_{s}^{ll'}))-y_{is}^{l'}|z_s]\geqslant 0. \tag{B.5}$$

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

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

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

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

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

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

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

$$(\exp(\Delta\kappa^{ll'}-\Delta\theta_{ll'})-1)\mathbb{E}[(\sum_{l''=1}^{L}\exp(\Delta\kappa^{l''l'}+\alpha\Delta w_s^{l''l'}))^{-1}|z_s]\geqslant 0. \tag{B.6}$$

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

$$\exp(\Delta\kappa^{ll'}-\Delta\theta_{ll'})-1\geqslant 0 \qquad \Leftrightarrow \qquad \Delta\kappa^{ll'}\geqslant\Delta\theta_{ll'}. \tag{B.7}$$

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

$$\exp(\Delta\kappa^{l'l}-\Delta\theta_{l'l})-1\geqslant 0 \qquad \Leftrightarrow \qquad \Delta\kappa^{ll'}\leqslant\Delta\theta_{ll'}. \tag{B.8}$$

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

B.5 First-Step Moment Inequalities: Additional Derivations

Derivation of equation (28). Consider points $e_{isjr}^{ll'}$ such that $e_{isjr}^{ll'}=h^{ll'}(z_s,z_r,\theta_\alpha)$ for some function $h^{ll'}(\cdot)$. We compute the function in equation (28) by finding the value of $h^{ll'}(z_s,z_r,\theta_\alpha)$ that minimizes the moment in equation (26) at each value of θ_α. Specifically, given z_s, z_r, and θ_α, the first-order condition of the moment in equation (26) with respect to $h^{ll'}(z_s,z_r,\theta_\alpha)$

is

$$\mathbb{E}[y^l_{is}y^{l'}_{jr}(2h^{ll'}(z_s,z_r,\theta_\alpha)-\theta_\alpha(\Delta w^{ll'}_s+\Delta w^{l'l}_r))|z_s,z_r]=0,$$

or, equivalently, $\mathbb{E}[2h^{ll'}(z_s,z_r,\theta_\alpha)-\theta_\alpha(\Delta w^{ll'}_s+\Delta w^{l'l}_r)|z_s,z_r,y^l_{is}y^{l'}_{is'}=1]=0$. Solving for $h^{ll'}(z_s,z_r,\theta_\alpha)$, we obtain the solution in equation (28).

Derivation of equation (29). Equations (26) to (28) imply the following inequality

$$\mathbb{E}[y^l_{is}y^l_{jr}+y^{l'}_{is}y^{l'}_{jr}-y^l_{is}y^{l'}_{jr}\exp(-\theta_\alpha\mathbb{E}[0.5(\Delta w^{ll'}_s+\Delta w^{l'l}_r)|z_s,z_r,y^l_{is}y^{l'}_{jr}=1])\times \\ (2-\theta_\alpha((\Delta w^{ll'}_s+\Delta w^{l'l}_r)-\mathbb{E}[\Delta w^{ll'}_s+\Delta w^{l'l}_r|z_s,z_r,y^l_{is}y^{l'}_{jr}=1]))|z_s,z_r]\geqslant 0,$$

or, equivalently,

$$\mathbb{E}[y^l_{is}y^l_{jr}+y^{l'}_{is}y^{l'}_{jr}|z_s,z_r]\geqslant\mathbb{E}[y^l_{is}y^{l'}_{jr}\exp(-\theta_\alpha\mathbb{E}[0.5(\Delta w^{ll'}_s+\Delta w^{l'l}_r)|z_s,z_r,y^l_{is}y^{l'}_{jr}=1])\times \\ (2-\theta_\alpha((\Delta w^{ll'}_s+\Delta w^{l'l}_r)-\mathbb{E}[\Delta w^{ll'}_s+\Delta w^{l'l}_r|z_s,z_r,y^l_{is}y^{l'}_{jr}=1]))|z_s,z_r].$$

Using the LIE, we can rewrite this inequality as

$$\mathbb{E}[y^l_{is}y^l_{jr}+y^{l'}_{is}y^{l'}_{jr}|z_s,z_r]\geqslant\mathbb{E}[\exp(-\theta_\alpha\mathbb{E}[0.5(\Delta w^{ll'}_s+\Delta w^{l'l}_r)|z_s,z_r,y^l_{is}y^{l'}_{jr}=1])\times \\ (2-\theta_\alpha((\Delta w^{ll'}_s+\Delta w^{l'l}_r)-\mathbb{E}[\Delta w^{ll'}_s+\Delta w^{l'l}_r|z_s,z_r,y^l_{is}y^{l'}_{jr}=1]))|z_s,z_r,y^l_{is}y^{l'}_{jr}=1]\times \\ \mathbb{E}[y^l_{is}y^{l'}_{jr}|z_s,z_r]\geqslant 0.$$

Simplifying this expression, and rearranging, we obtain the expression in equation (29).

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

Since $z_s\subseteq\mathcal{W}_{is}$ and $z_r\subseteq\mathcal{W}_{jr}$ according to Corollary 3, we rewrite equation (29) as

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

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

$$\frac{\mathbb{E}[\mathbb{E}[y^l_{is}|\mathcal{W}_{is}]\mathbb{E}[y^{l'}_{jr}|\mathcal{W}_{jr}]|z_s,z_r]}{\mathbb{E}[0.5(\mathbb{E}[y^l_{is}|\mathcal{W}_{is}]\mathbb{E}[y^l_{jr}|\mathcal{W}_{jr}]+\mathbb{E}[y^{l'}_{is}|\mathcal{W}_{is}]\mathbb{E}[y^{l'}_{jr}|\mathcal{W}_{jr}])|z_s,z_r]}\leqslant \\ \exp(\theta_\alpha\mathbb{E}[0.5(\Delta w^{ll'}_s+\Delta w^{l'l}_r)|z_s,z_r,y^l_{is}y^{l'}_{jr}=1]). \tag{B.9}$$

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 then rewrite the inequality in equation (B.9) 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

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

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

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

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

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

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

B.7 Using Inequalities for Estimation of the Wage Parameter

We describe here how we use the inequalities in Section 3.2 to compute a CI for θ_α. The inequality in equation (26) is specific to locations l and l' and conditions on the vectors z_s and z_r. Given that conditional moment inequality, we implement the following steps to derive $k = 1, \ldots, K$ unconditional moment inequalities. First, we choose scalars $\Delta z_s^{ll'} \subseteq z_s$ and $\Delta z_r^{l'l} \subseteq z_r$ that are correlated with $\Delta w_s^{ll'}$ and $\Delta w_r^{l'l}$, and that we use in all K inequalities. Second, for each k, we choose a subset $[\underline{z}_{ks}, \overline{z}_{ks}]$ of the support of $\Delta z_s^{ll'}$, a subset $[\underline{z}_{kr}, \overline{z}_{kr}]$ of the support of $\Delta z_r^{l'l}$, and a exponent $d_k \in \mathbb{Z}$. Given these choices, we build the inequality

$$\mathbb{E}[(y_{is}^{l}y_{jr}^{l} + y_{is}^{l'}y_{jr}^{l'} - y_{is}^{l}y_{jr}^{l'}\exp(-e_{isjr}^{ll'})(2 + 2e_{isjr}^{ll'} - \theta_{\alpha}(\Delta w_{s}^{ll'} + \Delta w_{r}^{l'l}))) \times g_k(\Delta z_s^{ll'}, \Delta z_r^{l'l})] \geqslant 0, \tag{B.10}$$

where the term in parenthesis coincides with the moment function in equation (26) and

$$g_k(\Delta z_s^{ll'}, \Delta z_r^{l'l}) = \mathbb{1}\{\underline{z}_{ks} < \Delta z_s^{ll'} \leqslant \overline{z}_{ks}\}\mathbb{1}\{\underline{z}_{kr} < \Delta z_r^{l'l} \leqslant \overline{z}_{kr}\}(|\Delta z_s^{ll'}||\Delta z_r^{l'l}|)^{d_k}.$$

Effectively, for a pre-specified $q \in \mathbb{N}$, we choose limits $\underline{z}_{ks}$ and $\overline{z}_{ks}$ that correspond to consecutive elements of the vector of q-quantiles of the distribution of $\Delta z_s^{ll'}$ across all types and location pairs. We do the same for the limits $\underline{z}_{kr}$ and $\overline{z}_{kr}$. Concerning the exponent d_k, we set it to either -1, 0, or 1. For example, when $q = 2$ and $d_k = 0$ for all k, the number of inequalities of the type in equation (30) is $K = 4$, and the corresponding instruments are

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

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

C Additional Simulation Results

In Appendix C.1, we describe the inequalities we use to obtain the results in Table 1. In Appendix C.2, we explore alternative ways of building the first-step inequalities. In Appendix C.3, we present estimates analogous to those in Table 1, but using a larger set of instruments. In Appendix C.4, we explore the robustness of the results in Table 1 to different values of amenities. In Appendix C.5, we compare the estimates in Table 1 to alternative estimates computed using approximations points other than those in equations (14) and (28).

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

First step. Given a predictor z_s^l of the wage level w_s^l in every location l, a pair of locations l and l', and a pair of indices k and k' determining the instruments we use, we compute the

CIs for θ_α displayed in Table 1 using moment inequalities of the type:

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

where $r(s)$ indexes the type matched with type s. We use this moment inequality for every pair of locations in the set $\{(l, l'); l \in \{1, 2, 3\}, l' \in \{1, 2, 3\}, l \neq l'\}$ and every pair of instrument indices in the set $\{(k, k'); k \in \{1, 2\}, k' \in \{1, 2\}\}$. Thus, we use 24 inequalities to identify θ_α. We now describe in more detail how we choose the type $r(s)$ for each s; how we define the approximation points $e_s^{ll'}$ for each s; and how we define the instrument $g_k(\cdot)$ for $k \in \{1, 2\}$.

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

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

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

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

By selecting the type r according to equation (C.2), we aim to approximate the condition in equation (C.3) while taking into account that the sets $\mathcal{W}_s$ and $\mathcal{W}_r$ are generally not observed.

Second, in terms of the approximation point $e_s^{ll'}$ for each type s used in equation (C.1), we build on equation (28) and impose

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

where, as indicated above, e.g., $\hat{\mathbb{E}}[\Delta w_s^{ll'}|\Delta z_s^{ll'}, y_s^l = 1]$ is the predicted value of $\Delta w_s^{ll'}$ computed using a linear regression of $\Delta w_s^{ll'}$ on $\Delta z_s^{ll'}$ estimated on the subset of observations with $y_s^l = 1$. Equation (28) indicates that the optimal approximation point $e_s^{ll'}$ is a function of, for example, the conditional expectation of $\Delta w_s^{ll'}$ given $\Delta z_s^{ll'}$ and $y_s^l = 1$; in equation (C.4), we compute instead a linear regression. In unreported results, we have instead approximated nonparametrically the conditional expectation of $\Delta w_s^{ll'}$ given $\Delta z_s^{ll'}$ and $y_s^l = 1$, obtaining very similar results to those attained when using linear regressions instead.

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

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

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

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

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

$$\sum_{s=1}^{S}(y_s^l - y_s^1 \exp(-e_s^{1l})(1 + e_s^{1l} + (\theta_l + \check{\theta}_\alpha \Delta w_s^{l1})))g_k(\Delta z_s^{1l}) \geqslant 0; \tag{C.6b}$$

or the following odds-based moment inequalities

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

$$\sum_{s=1}^{S}(y_s^1 \exp(\theta_l + \check{\theta}_\alpha \Delta w_s^{l1}) - y_s^l)g_k(\Delta z_s^{1l}) \geqslant 0; \tag{C.7b}$$

or both the bounding and odds-based moment inequalities jointly. The instrument function $g_k(\cdot)$ in equations (C.6) and (C.7) is defined as in equation (C.5). The points e_s^{l1} and e_s^{1l} entering the inequalities in equation (C.6) are

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

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

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

Differences across cases. In cases 1 to 4 in Table 1, the wage predictor z_s^l introduced in equations (C.1) to (C.8) equals the shifter z_{2s}^l in equation (32). In case 5, z_s^l equals w_s^l. In this case, equation (C.2) simplifies to $|\Delta w_s^{ll'} - \Delta w_{r(s)}^{l'l}| \leqslant \tau$; equation (C.4) simplifies to $e_s^{ll'} = \theta_\alpha 0.5(\Delta w_s^{ll'} + \Delta w_{r(s)}^{l'l})$; and equation (C.8) simplifies to $e_s^{l1} = \theta_l + \check{\theta}_\alpha \Delta w_s^{l1}$ and $e_s^{1l} = -e_s^{l1}$.

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

In Table C.1, we present CIs for θ_α computed according to equations (C.1) to (C.5), and show how these CIs vary as we change the value of τ entering equation (C.2). Table C.1 shows that the 95% CI for θ_α becomes wider as we increase the value of τ.

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

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

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

C.3 Two-step Moment Inequalities with Additional Instruments

In Table C.2, we present CIs analogous to those in Table 1 with the only difference that, instead of using the instrument function in equation (C.5), we use the following instrument function

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

where $Q_{25}(\Delta z_s^{ll'})$, $Q_{50}(\Delta z_s^{ll'})$, and $Q_{75}(\Delta z_s^{ll'})$ respectively denote the percentiles 25, 50, and 75 of the distribution of $\Delta z_s^{ll'}$ across all types s and pairs of locations l and l'. As a comparison of the results in tables 1 and C.2 illustrates, the CIs for θ_α, θ_2, and θ_3, become tighter when we swap the instrument function in equation (C.5) for the more detailed function in equation (C.9). This change in instrument functions results in an increase in the number of moment inequalities we use in our estimation.

Table C.2: Simulation Results - Confidence Intervals With Additional Instruments

Case	σ_1	σ_3	z_s^l	First Step	Second Step		
				θ_α	Mom. Ineq.	θ_2	θ_3
1	0	0	z_{2s}^l	[1 , 1.02]	Bounding	[0 , 0]	[1 , 1]
					Odds-based	[0 , 0]	[1 , 1]
					Both	[0 , 0]	[1 , 1]
2	0	1	z_{2s}^l	[1 , 1.01]	Bounding	[0 , 0]	[1 , 1]
					Odds-based	[-0.33 , 0.32]	[0.68 , 1.33]
					Both	[0 , 0]	[1 , 1]
3	1	0	z_{2s}^l	[0.91 , 1.15]	Bounding	[-0.31 , 0.31]	[0.70 , 1.30]
					Odds-based	[0 , 0]	[1 , 1.01]
					Both	[0 , 0]	[1 , 1.01]
4	1	1	z_{2s}^l	[0.91 , 1.19]	Bounding	[-0.31 , 0.31]	[0.70 , 1.30]
					Odds-based	[-0.32 , 0.32]	[0.68 , 1.33]
					Both	[-0.31 , 0.31]	[0.70 , 1.31]
5	0	1	w_s^l	$\varnothing$	Bounding	$\varnothing$	$\varnothing$
					Odds-based	$\varnothing$	$\varnothing$
					Both	$\varnothing$	$\varnothing$

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

C.4 Amenity Differences Across All Locations

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

Table C.3: Simulation Results - Confidence Intervals With Amenity Differences

Case	σ_1	σ_3	z_s^l	$(\kappa^1, \kappa^2, \kappa^3)$	1st Step
					θ_α
2	0	1	z_{2s}^l	$(0, 0, 1)$	[1 , 1.01]
2	0	1	z_{2s}^l	$(0, 0.5, 1)$	[0.92 , 1.09]
2	0	1	z_{2s}^l	$(0, 0, 2)$	[1 , 1.02]
2	0	1	z_{2s}^l	$(0, 1, 2)$	[0.73 , 1.29]
2	0	1	z_{2s}^l	$(0, 0, 3)$	[1 , 1.02]
2	0	1	z_{2s}^l	$(0, 1.5, 3)$	[0.55 , 1.49]

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

C.5 Estimator with Alternative Approximation Points

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

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

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

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

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

Table C.4: Simulation Results - Confidence Intervals With Alternative Approximation Points

σ_1	σ_3	z_s^l	Mom. Ineq.	θ_α	θ_2	θ_3
Panel A: Confidence Intervals Using Approximation Points in Eqs. (C.4) *and* (C.8)						
0	0	z_{2s}^l	Bounding	[1 , 1.02]	[0 , 0]	[1 , 1]
0.25	0	z_{2s}^l	Bounding	[0.98 , 1.04]	[-0.02 , 0.02]	[0.98 , 1.02]
0.50	0	z_{2s}^l	Bounding	[0.95 , 1.06]	[-0.09 , 0.08]	[0.92 , 1.08]
0.75	1	z_{2s}^l	Bounding	[0.89 , 1.16]	[-0.18 , 0.18]	[0.82 , 1.18]
1	0	z_{2s}^l	Bounding	[0.82 , 1.29]	[-0.31 , 0.31]	[0.70 , 1.30]
Panel B: Confidence Intervals Using Approximation Points in Eqs. (C.10) *and* (C.11)						
0	0	z_{2s}^l	Bounding	[1 , 1.02]	[0 , 0]	[1 , 1]
0.25	0	z_{2s}^l	Bounding	[0.98 , 1.04]	[-0.02 , 0.02]	[0.98 , 1.02]
0.50	0	z_{2s}^l	Bounding	[0.94 , 1.09]	[-0.09 , 0.08]	[0.92 , 1.08]
0.75	1	z_{2s}^l	Bounding	[0.88 , 1.24]	[-0.20 , 0.20]	[0.81 , 1.21]
1	0	z_{2s}^l	Bounding	[0.81 , 1.50]	[-0.41 , 0.42]	[0.63 , 1.46]

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

Relative to the approximation points in equations (C.4) and (C.8), those in equations (C.10) and (C.11) have the advantage that they do not involve any estimation; in particular, they do not require computing predicted values from a linear regression. Consequently, they do not affect the validity of standard moment inequality inference procedures. However, as a comparison of both panels in Table C.4 illustrates, the use of the approximation points in equations (C.10) and (C.11) results in CIs that are larger than when the approximations points in equations (C.4) and (C.8) are used. The difference between both sets of CIs increases in the value of σ_1.

D Data and Summary Statistics

D.1 Data Sources and Sample Construction

The RAIS data. Our main data source is the *Relação Anual de Informações Sociais* (RAIS), an administrative dataset maintained by Brazil's Ministry of Labor. It includes the universe of formal employment spells in the private and public sectors. Individual workers are identified by government-issued identification numbers (PIS/PASEP and CPF), allowing us to track them as they change employers. For all spells observed between 1993 and 2011, we use information on their start and end dates, average monthly wage, number of work hours stipulated in the contract, 2-digit sector (according to the *Classificação Nacional de Atividades Econômicas*, CNAE), and information on the worker's gender, age, race, and education level. All information is reported by the employers.

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

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

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

Sample restrictions and sampling. We limit our data to workers aged 25 to 64, as these are the workers most likely to have completed their education and not yet retired. The sample period is 2002-2011. We use 1993-2001 information to measure each worker's experience in each sector and microregion. To limit our data to workers with a sufficiently close labor relationship with the formal sector, we restrict our sample to workers observed at RAIS for at least seven years in the sample period. We also restrict our sample to workers with similar demographic characteristics; specifically, we focus on workers with at least a high school degree identified as male and white. For computational reasons, we focus on a sample of 10 million worker-year pairs. To ensure we observe a large enough number of individuals per market, we focus on 1,000 labor markets consisting of all combinations of the 50 microregions (out of 558) and 20 sectors (out of 51) with the largest total employment reported in RAIS. We then obtain our sample by randomly sampling 1 million individuals per sample year among those employed in the 1,000 labor markets of interest.

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

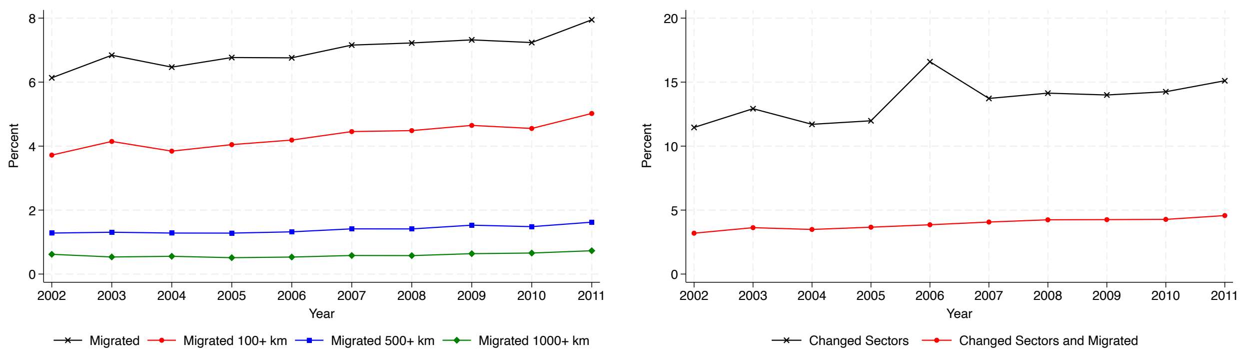
D.2 Summary Statistics: Migration

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

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

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

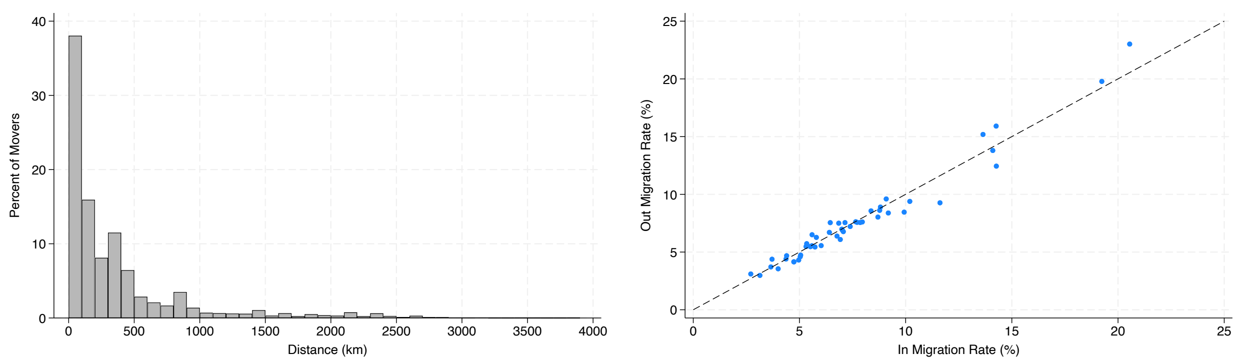
(a) Migration Rates by Year (b) Sector and Microregion Changes



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

Figure D.2: Migration Patterns

(a) Histogram: Migration Distance (b) Out- vs. In-Migration (sample)



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

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

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

E Appendix to Empirical Analysis

E.1 Implementation of Moment Inequalities

Approximation points. Computing the moments in equations (12) and (26) requires specifying the approximation points $e^{ll'}_{is}$ and $e^{ll'}_{isjr}$, respectively. Equations (14) and (28) provide functional forms for these approximation points that, according to corollaries 1 and 3, result in second-step bounding inequalities and first-step inequalities, respectively, that can point identify the parameters of interest. Furthermore, as discussed in Appendix sections B.1 and B.5, the approximation points in equations (14) and (28) yield the tightest identified sets among all approximation points in a family of functions described in detail in those sections. The expressions in equations (14) and (28), however, depend on the expectation of specific wage differences conditional on the wage predictor used to build the corresponding inequality. Since we ignore the true value of those expectations, we must approximate them in some way. In our empirical application, we use approximation points that are simple to compute and that, importantly, do not depend on any regression estimate.[35]

Specifically, we compute the approximation points entering the bounding moment inequalities in equation (12) as

$$e^{ll'}_{is} = \psi_0 + \psi_1 \Delta z^{ll'}_{st}, \tag{E.1}$$

where ψ_0 and ψ_1 are constants the researcher chooses. In practice, we use simultaneously in estimation several different bounding inequalities of the type in equation (12) computed using

[35]This stands in contrast with the approach we follow in our simulation setting when computing the CIs reported in Table 1, where we approximate the conditional expectations entering equations (14) and (28) using linear regressions; see Appendix C.1 for more details. While the large number of observations we use in our simulation setting implies that any noise in the linear regression estimates will have a minimal impact in our estimated CIs, this may not be true in our empirical application. Thus, to simplify the computation of the moment inequality CIs in our application, we restrict ourselves to using approximation points that are not functions of prior estimates.

the approximation points in equation (E.1) with different values of the constants ψ_0 and ψ_1. Similarly, we compute the approximation points entering the first-step moment inequalities in equation (26) as

$$e_{is}^{ll'} = \psi_2 0.5(\Delta z_{st}^{ll'} + \Delta z_{rt}^{l'l}), \tag{E.2}$$

where ψ_2 is a constant. In practice, we also use simultaneously in estimation several different first-step moment inequalities of the type in equation (26) computed using the approximation points in equation (E.2) with different values of the constant ψ_2.[36]

Aggregating across pairs of locations in the first-step moment inequalities. The moment inequality in equation (26) holds for any two locations l and l'. As we consider 50 possible locations in our empirical application, there is a very large number of potential location pairs. Instead of using a correspondingly large number of different moment inequalities of the type in equation (26), we use a smaller number of inequalities that aggregate across location pairs. Our choice of which location pairs to combine is guided by Corollary 3. This corollary indicates that a requisite for the inequality in equation (26) to point identify the wage parameter θ_α is that the locations l and l' being compared offer the same amenity level in the population of reference; that is, using the notation in our empirical application, $\kappa_{nt}^{l} - \kappa_{nt}^{l'} = 0$. Enforcing this condition is infeasible as these amenity levels are continuous variables that are only estimated later in our estimation procedure. However, as κ_{nt}^{l} accounts for migration costs in our setting, it is reasonable to expect it will vary with the distance between locations n and l. Thus, we hypothesize that locations l and l' that are at a similar distance to an origin n are more likely to have similar amenity levels from the perspective of workers located in n and, when combined in the context of the inequality in equation (26), should yield smaller identified sets. Consequently, we form the sample analogue of the moment in equation (26) aggregating only across location pairs l and l' for which the difference between the distance from n to l and the distance from n to l' is in the lower tercile of all pairwise differences in distance to n. Given such location pairs, we form the moment function in equation (26) by further aggregating across all sector pairs s and r, and across worker pairs within those sectors.

Instrument vectors. Given a wage predictor z_{st}^{l} for every l, s, and t, we construct mo-

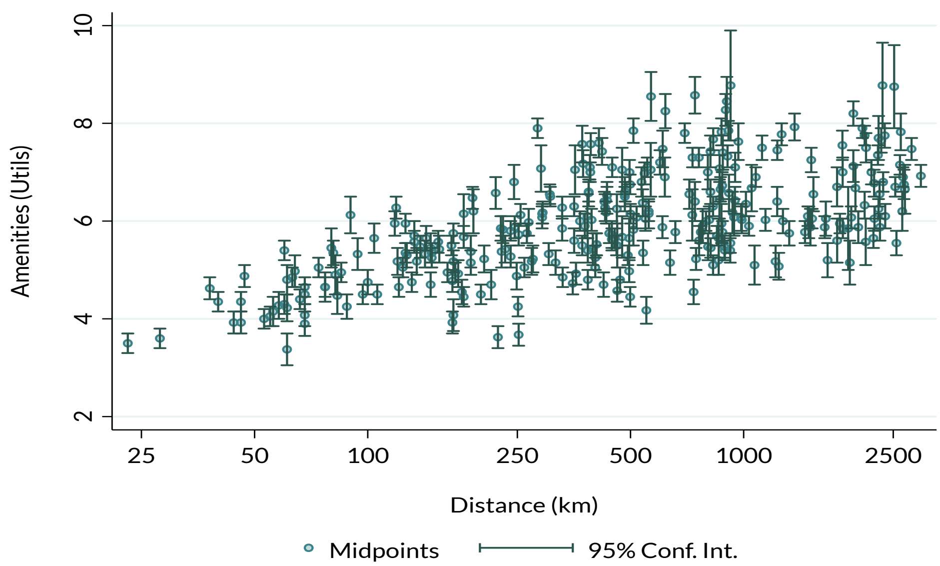
[36] The formulas for the approximation points in equations (E.1) and (E.2) are similar to those used to compute the simulation results discussed in Appendix C.5. However, the formulas in equations (E.1) and (E.2) differ from those in equations (C.10) and (C.11) in that they do not depend on the structural parameters to estimate, but on constants chosen by the researcher. The fact that the researcher can explore the identifying power of a large set of possible values for the constants (ψ_0, ψ_1, ψ_2) has the advantage of generating tighter CIs, and the computational disadvantage that the number of moment inequalities used in estimation increases in the set of values of (ψ_0, ψ_1, ψ_2) the researcher uses.

ment inequalities using instruments of the type in equation (31) for every $d_k \in \{-1, 1\}$ and $[\underline{z}_k, \overline{z}_k] \in \{[-\infty, Q_{25}(\Delta z^{ll'}_{st})], [Q_{25}(\Delta z^{ll'}_{st}), Q_{50}(\Delta z^{ll'}_{st})], [Q_{50}(\Delta z^{ll'}_{st}), Q_{75}(\Delta z^{ll'}_{st})], [Q_{75}(\Delta z^{ll'}_{st}), \infty]\}$, where $Q_q(\Delta z^{ll'}_{st})$ denotes the percentile q of the distribution of $\Delta z^{ll'}_{st}$. Thus, we use 8 different instrument vectors.

E.2 Additional Result

Figure E.1 provide the CIs of estimated amenities κ^l_{nt}, described in Section 5.2.2.

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



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

E.3 PPML-IV Estimator

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

$$M^l_{nst} = \frac{\exp\left(\alpha\mathbb{E}[w^l_{st}|\mathcal{J}_{st}] - \kappa^l_{nt}\right)}{\sum_k \exp\left(\alpha\mathbb{E}[w^k_{st}|\mathcal{J}_{st}] - \kappa^k_{nt}\right)} L_{nst-1}$$

$$
\begin{aligned}
&= \exp\left(\alpha \mathbb{E}[w_{st}^{l} | \mathcal{J}_{st}] - \kappa_{nt}^{l} + \Gamma_{nst}\right) \\
&= \exp\left(\alpha (w_{st}^{l} - \nu_{st}^{l}) - \kappa_{nt}^{l} + \Gamma_{nst}\right) \\
&= \exp\left(\Lambda_{st}^{l} + \Psi_{nt}^{l} + \Gamma_{nst}\right), \qquad \text{(E.3)}
\end{aligned}
$$

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

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

Using information on $\{M_{nst}^{l}\}_{n=1,l=1}^{L,L}$ for a period t, sector s, and L origin and destination locations, the procedure in Artuç and McLaren (2015) recovers estimates of α and $\{\kappa_{nt}^{l}\}_{n=1,l=1}^{L,L}$ in three steps. First, it computes PPML estimates of $\{\Lambda_{st}^{l}\}_{l=1}^{L}$, $\{\Gamma_{nst}\}_{n=1}^{L}$ and $\{\Psi_{nt}^{l}\}_{n=1,l=1}^{L,L}$ using the expression in the last line in equation (E.3). Second, under the assumption that a variable z_{st}^{l} correlated with w_{st}^{l} belongs to the information set $\mathcal{J}_{st}$, it computes an IV estimate of α by regressing $\hat{\Lambda}_{st}^{l}$ on w_{st}^{l} using z_{st}^{l} as an instrument, with $\hat{\Lambda}_{st}^{l}$ the first-step estimate of Λ_{st}^{l}. Third, it recovers $\hat{\kappa}_{nt}^{l} = -\hat{\Psi}_{nt}^{l}$ for every origin n and destination l.

F Model with Endogenous Worker Types - Sectors

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

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

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

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

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

The assumption in equation (4) applies directly to the model with endogenous worker types. The assumption in equation (5) extends naturally to the model considered here.

Specifically, for any sectors s and r, locations l and l', and worker indices i and j, it holds:

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

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

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

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

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

G Extension: Dynamic Model of Location Choice

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

G.1 Theoretical Framework

Defining a dummy variable y_{ist}^l that equals one if worker i of type s chooses l at t, we assume

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

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

$$\mathcal{V}_{ist}^l = v_{ist}^l + \varepsilon_{ist}^l, \tag{G.2a}$$

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

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

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

Equation (G.2a) splits the choice-specific value function into the idiosyncratic component ε^l_{ist} and a variable v^l_{ist} that equation (G.2b) defines as the sum of four terms. First, the migration costs between locations n and l, modeled as a function of observed covariates x^l_{nt} and a vector of parameters β. Second, a location- and period-specific term λ^l_t, which captures a location's amenities and (log) price index. Third, the wage component αw^l_{ist}. Fourth, the product of the discount factor δ and a variable $\mathcal{V}^{(lt)}_{ist+1}$ that, according to equation (G.3), equals the worker's period-$t+1$ value function conditional on choosing alternative l at period t.[37]

Defining $\lambda_t = (\lambda^1_t, \ldots, \lambda^L_t)$ and $\varepsilon_{ist} = (\varepsilon^1_{ist}, \ldots, \varepsilon^L_{ist})$, we assume that

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

Thus, when making choices at t, workers know the vectors of contemporaneous idiosyncratic preferences ε_{ist} and amenities λ_t. Equation (G.4) does not restrict the information workers have about wages $w_{ist'} = (w^1_{ist'}, \ldots, w^L_{ist'})$ for any $t' \geqslant t$ or amenities $\lambda_{t'}$ for any $t' > t$.

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

$$\mathcal{J}_{ist'} \perp\!\!\!\perp y_{ist} | \mathcal{J}_{ist}. \tag{G.5}$$

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

Defining $\Delta v^{ll'}_{ist} \equiv v^l_{ist} - v^{l'}_{ist}$, we impose that for any period t, locations l and l', types s and r, and workers i and j that share a common prior location n,

$$\mathbb{E}[\Delta v^{ll'}_{ist} | \mathcal{J}_{ist}, \mathcal{J}_{jrt}] = \mathbb{E}[\Delta v^{ll'}_{ist} | \mathcal{J}_{ist}] = \mathbb{E}[\Delta v^{ll'}_{ist} | \mathcal{W}_{ist}]. \tag{G.6}$$

The first equality imposes that every worker has at least as much information as any other worker of a different type r with whom it shares prior location n about differences in their own location-specific value functions. The second equality imposes that, once we condition on all other elements of the information set of worker i of type s at period t, the idiosyncratic preferences in ε_{ist} do not contain any information on $\Delta v^{ll'}_{ist}$ for any two locations l and l'.[38]

[37] A comparison of equations (3) and (G.2) shows that, at the expense of assuming $\delta = 0$, the static model allows for a more flexible specification of migration costs, which may vary freely between locations and periods.

[38] The variable $\Delta v^{ll'}_{ist}$ depends on the worker's future choices, which will depend on $\varepsilon_{ist'}$ for $t' > t$; thus, equation (G.6) will generally not hold unless ε_{ist} is independent over time.

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

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

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

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

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

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

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

G.2 Estimation With Moment Inequalities

We provide a two-step estimator. In the first step, we compute a confidence set for (α, β) using inequalities that difference out the amenity term λ_t^l for all l and t. In the second step, for each $l = 2, \ldots, L$ and sample period t, we derive inequalities that depend only on α, β, and λ_t^l, and combine these inequalities with the confidence set for (α, β) to compute a confidence interval for λ_t^l. We denote by $(\theta_\alpha, \theta_\beta)$ the parameter vector with true value (α, β), and by $\Theta_{(\alpha,\beta)}$ the set of possible values of $(\theta_\alpha, \theta_\beta)$. We denote by θ_t^l the parameter with true value λ_t^l and by Θ_t^l the set of possible values of θ_t^l. In Appendix G.2.1, we discuss the estimation of $\{\theta_t^l\}_{l,t}$. In Appendix G.2.2, we describe the estimation of $(\theta_\alpha, \theta_\beta)$.

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

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

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

with

$$\Delta\tilde{v}^{ll'}_{ist}(\Delta\theta^{ll'}_t) = \beta\Delta x^{ll'}_{nt} + \Delta\theta^{ll'}_t + \alpha\Delta w^{ll'}_{ist} + \delta\beta\sum_{l''=1}^{L} y^{(lt)l''}_{ist+1}(x^{l''}_{lt+1} - x^{l''}_{l't+1}). \tag{G.10}$$

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

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

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

To obtain the inequality $\tilde{\mathfrak{m}}^{ll'}(z_{st}, \Delta\lambda^{ll'}_t) \geqslant 0$, we first follow steps analogous to those taken to derive the static bounding inequality in equation (13).[39] In this way, we obtain

$$\mathbb{E}[y^{l'}_{ist} - y^{l}_{ist}\exp(-e^{ll'}_{ist})(1 + e^{ll'}_{ist} - (v^{l}_{ist} - v^{l'}_{ist}))|z_{st}] \geqslant 0. \tag{G.11}$$

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

$$v^{l}_{ist} - \tilde{v}^{l'}_{ist} = u^{l}_{ist} - u^{l'}_{ist} + \delta\beta\sum_{l''=1}^{L} y^{(lt)l''}_{ist+1}(x^{l''}_{lt+1} - x^{l''}_{l't+1}), \tag{G.12}$$

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

[39] While it may be feasible to use the odds-based moment inequalities introduced in Section 3.1.2 in the context of our dynamic model, we have not found a way of doing so.

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

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

$$\tilde{\mathbb{M}}^{ll'}(z_{st}, z_{rt}, \theta_\alpha, \theta_\beta) \equiv \tag{G.13}$$
$$\mathbb{E}[y^l_{ist}y^l_{jrt} + y^{l'}_{ist}y^{l'}_{jrt} - y^l_{ist}y^{l'}_{jrt}\exp(-e^{ll'}_{ijsrt})(2 + 2e^{ll'}_{ijsrt} - (\Delta\tilde{v}^{ll'}_{ist} + \Delta\tilde{v}^{l'l}_{jrt}))|z_s, z_r].$$

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

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

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

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

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

G.3 Proofs and Additional Details

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

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

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

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

$$(y^l_{ist} + y^{l'}_{ist})(\mathbb{1}\{\mathbb{E}[\Delta v^{ll'}_{ist}|\mathcal{W}_{ist}] + \Delta\varepsilon^{ll'}_{ist} \geqslant 0\} - y^l_{ist}) = 0, \tag{G.15}$$

where $\Delta v_{ist}^{ll'} = v_{ist}^{l} - v_{ist}^{l'}$ and $\Delta \varepsilon_{ist}^{ll'} = \varepsilon_{ist}^{l} - \varepsilon_{ist}^{l'}$. This equality holds for any worker i of any type s, any period t, and any two locations l and l'. Thus, computing a conditional expectation of both sides of this equality, we obtain

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

Following steps analogous to those described in Appendix A.1, we can derive the following moment inequality

$$\mathbb{E}[y_{ist}^{l'} + y_{ist}^{l}\exp(-e_{ist}^{ll'})(-(1 + e_{ist}^{ll'}) + \Delta v_{ist}^{ll'})|z_{st}] \geqslant 0. \tag{G.16}$$

This moment differs from that in equation (12) only in that the difference in the static utility between alternatives l and l' entering equation (12) (i.e., $\Delta\theta_{ll'} + \alpha\Delta w_{s}^{ll'}$) is substituted by the difference in the corresponding choice-specific value functions (i.e., $\Delta v_{ist}^{ll'}$).

The inequality in equation (G.16) is not immediately useful for the identification of the parameters of the dynamic model described in Appendix G.1. The term $\Delta v_{ist}^{ll'}$ depends on the optimal choices of worker i of type s in every period $t' > t$ both conditional on choosing l at period t (which matters for the value of v_{ist}^{l}) and conditional on choosing l' at period t (which matters for the value of $v_{ist}^{l'}$). To derive a moment inequality that can be used to partially identify the parameters of the model described in Appendix G.1, we follow the approach in Morales et al. (2019). Specifically, we substitute $\Delta v_{ist}^{ll'}$ by the variable

$$\Delta \tilde{v}_{ist}^{ll'} \equiv v_{ist}^{l} - \tilde{v}_{ist}^{l'}, \tag{G.17}$$

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

$$y_{ist'}^{(lt)} = (y_{ist'}^{(lt)1}, \ldots, y_{ist'}^{(lt)L}) \tag{G.18}$$

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

$$\begin{aligned} v_{ist}^{l} = \kappa_{nt}^{l} + \alpha w_{st}^{l} + \delta \sum_{l''=1}^{L} y_{ist+1}^{(lt)l''}(\kappa_{lt+1}^{l''} + \alpha w_{st+1}^{l''} + \varepsilon_{ist+1}^{l''}) \\ + \sum_{t'>t+1} \delta^{t'-t} \sum_{n'=1}^{L}\sum_{l''=1}^{L} y_{ist'-1}^{(lt)n'} y_{ist'}^{(lt)l''}(\kappa_{n't'}^{l''} + \alpha w_{st'}^{l''} + \varepsilon_{ist'}^{l''}), \end{aligned} \tag{G.19a}$$

$$v^{l'}_{ist} = \kappa^{l'}_{nt} + \alpha w^{l'}_{st} + \delta \sum_{l''=1}^{L} y^{(l't)l''}_{ist+1}(\kappa^{l''}_{l't+1} + \alpha w^{l''}_{st+1} + \varepsilon^{l''}_{ist+1}) \tag{G.19b}$$

$$+ \sum_{t'>t+1} \delta^{t'-t} \sum_{n'=1}^{L} \sum_{l''=1}^{L} y^{(l't)n'}_{ist'-1} y^{(l't)l''}_{ist'} (\kappa^{l''}_{n't'} + \alpha w^{l''}_{st'} + \varepsilon^{l''}_{ist'}),$$

$$\tilde{v}^{l'}_{ist} = \kappa^{l'}_{nt} + \alpha w^{l'}_{st} + \delta \sum_{l''=1}^{L} y^{(lt)l''}_{ist+1}(\kappa^{l''}_{l't+1} + \alpha w^{l''}_{st+1} + \varepsilon^{l''}_{ist+1}) \tag{G.19c}$$

$$+ \sum_{t'>t+1} \delta^{t'-t} \sum_{n'=1}^{L} \sum_{l''=1}^{L} y^{(lt)n'}_{ist'-1} y^{(lt)l''}_{ist'} (\kappa^{l''}_{n't'} + \alpha w^{l''}_{st'} + \varepsilon^{l''}_{ist'}).$$

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

$$\mathbb{E}[\Delta v^{ll'}_{ist}|\mathcal{W}_{ist}] \geqslant \mathbb{E}[\Delta \tilde{v}^{ll'}_{ist}|\mathcal{W}_{ist}]. \tag{G.20}$$

Equations (G.16) and (G.20) imply the following moment inequality

$$\mathbb{E}[y^{l'}_{ist} + y^{l}_{ist} \exp(-h^{ll'}_{ist}(z_{st}, \Delta\lambda^{ll'}_{t}))(-(1 + h^{ll'}_{ist}(z_{st}, \Delta\lambda^{ll'}_{t})) + \Delta\tilde{v}^{ll'}_{ist})|z_{st}] \geqslant 0. \tag{G.21}$$

Comparing the expressions for v^{l}_{ist} and $\tilde{v}^{l'}_{ist}$ in equations (G.19a) and (G.19c), we can write

$$\Delta\tilde{v}^{ll'}_{ist} = v^{l}_{ist} - \tilde{v}^{l'}_{ist} = (\kappa^{l}_{nt} - \kappa^{l'}_{nt}) + \alpha(w^{l}_{st} - w^{l'}_{st}) + \delta \sum_{l''=1}^{L} y^{(lt)l''}_{ist+1}(\kappa^{l''}_{lt+1} - \kappa^{l''}_{l't+1})$$

$$= \beta(x^{l}_{nt} - x^{l'}_{nt}) + (\lambda^{l}_{t} - \lambda^{l'}_{t}) + \alpha(w^{l}_{st} - w^{l'}_{st}) + \delta\beta \sum_{l''=1}^{L} y^{(lt)l''}_{ist+1}(x^{l''}_{lt+1} - x^{l''}_{l't+1}). \tag{G.22}$$

Combining equations (G.22) and (G.21), we obtain an inequality whose moment equals that in equation (G.9) for $\Delta\theta_{ll'} = \Delta\lambda^{ll'}$. Equations (G.21) and (G.22) thus imply Theorem 4. ■

G.3.2 Second-Step Bounding Inequalities: Additional Derivations

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

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

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

$$h^{ll'}_{ist}(z_{st}, \Delta\theta_{ll'}) = \mathbb{E}[\Delta\tilde{v}^{ll'}_{ist}|z_{st}, y^{l}_{ist} = 1], \tag{G.23}$$

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

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

$$y^{l'}_{jrt}(y^{l}_{ist} + y^{l'}_{ist})(\mathbb{1}\{\mathbb{E}[\Delta v^{ll'}_{ist}|\mathcal{W}_{ist}] + \Delta\varepsilon^{ll'}_{ist} \geqslant 0\} - y^{l}_{ist}) = 0. \tag{G.24}$$

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

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

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

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

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

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

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

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

As $-\exp(-x)$ is concave, we derive the following inequality given any scalar $e^{ll'}_{ijsrt}$,

$$\mathbb{E}[y^{l'}_{ist}y^{l'}_{jrt} + y^{l}_{ist}y^{l'}_{jrt}\exp(-e^{ll'}_{ijsrt})(-(1 + e^{ll'}_{ijsrt}) + \mathbb{E}[\Delta v^{ll'}_{ist}|\mathcal{W}_{ist}])|\mathcal{W}_{ist}, \mathcal{W}_{jrt}] \geqslant 0. \tag{G.25}$$

Let's consider the alternative moment

$$\mathbb{E}[y^{l'}_{ist}y^{l'}_{jrt} + y^{l}_{ist}y^{l'}_{jrt}\exp(-e^{ll'}_{ijsrt})(-(1 + e^{ll'}_{ijsrt}) + \Delta v^{ll'}_{ist})|\mathcal{W}_{ist}, \mathcal{W}_{jrt}]. \tag{G.26}$$

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

$$\mathbb{E}[\nu^{ll'}_{ist}|\mathcal{W}_{ist}, \mathcal{W}_{jrt}] = 0. \tag{G.27}$$

As $\mathcal{W}_{ist} \subseteq \mathcal{J}_{ist}$ and $\mathcal{W}_{jrt} \subseteq \mathcal{J}_{jrt}$, the LIE allows to write the moment in equation (G.26) as

$$\mathbb{E}[\mathbb{E}[y^{l'}_{ist}y^{l'}_{jrt} + y^{l}_{ist}y^{l'}_{jrt}\exp(-e^{ll'}_{ijsrt})(-(1 + e^{ll'}_{ijsrt}) + \mathbb{E}[\Delta v^{ll'}_{ist}|\mathcal{W}_{ist}] + \nu^{ll'}_{ist}))|\mathcal{J}_{ist}, \mathcal{J}_{jrt}]|\mathcal{W}_{ist}, \mathcal{W}_{jrt}].$$

Equation (G.1) implies $\mathbb{E}[y^{l}_{ist}y^{l'}_{jrt}|\mathcal{J}_{ist},\mathcal{J}_{jrt}] = y^{l}_{ist}y^{l'}_{jrt}$. Consequently, if $z_{st} \subseteq \mathcal{W}_{ist}$ and $z_{rt} \subseteq \mathcal{W}_{jrt}$, then $z_{st} \subseteq \mathcal{J}_{ist}$ and $z_{r} \subseteq \mathcal{J}_{jrt}$, and we can rewrite the moment in equation (G.26) as

$$\mathbb{E}[y^{l'}_{ist}y^{l'}_{jrt} + y^{l}_{ist}y^{l'}_{jrt}\exp(-e^{ll'}_{ijsrt})(-(1+e^{ll'}_{ijsrt}) + \mathbb{E}[\mathbb{E}[\Delta v^{ll'}_{ist}|\mathcal{W}_{ist}] + \nu^{ll'}_{ist}|\mathcal{W}_{ist},\mathcal{W}_{jrt}])|\mathcal{W}_{ist},\mathcal{W}_{jrt}],$$

and equation (G.27) implies we can rewrite the moment in equation (G.26) as

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

This moment is the same in equation (G.25), hence, we have shown:

$$\mathbb{E}[y^{l'}_{ist}y^{l'}_{jrt} + y^{l}_{ist}y^{l'}_{jrt}\exp(-e^{ll'}_{ijsrt})(-(1+e^{ll'}_{ijsrt}) + \Delta v^{ll'}_{ist})|\mathcal{W}_{ist},\mathcal{W}_{jrt}] \geqslant 0. \tag{G.28}$$

Following steps analogous to those we follow to derive the inequality in equation (G.21) from that in equation (G.16) (see Appendix G.3.1), we derive the following inequality from that in equation (G.28), where $\Delta\tilde{v}^{ll'}_{ist}$ is defined as in equation (G.22):

$$\mathbb{E}[y^{l'}_{ist}y^{l'}_{jrt} + y^{l}_{ist}y^{l'}_{jrt}\exp(-e^{ll'}_{ijsrt})(-(1+e^{ll'}_{ijsrt}) + \Delta\tilde{v}^{ll'}_{ist})|\mathcal{W}_{ist},\mathcal{W}_{jrt}] \geqslant 0. \tag{G.29}$$

This inequality is one of the two we combine to obtain the inequality that we use to bound the parameters θ_α and θ_β. To obtain the one, we start with the following expression

$$y^{l}_{ist}(y^{l}_{jrt} + y^{l'}_{jrt})(\mathbb{1}\{\mathbb{E}[\Delta v^{l'l}_{jrt}|\mathcal{W}_{jrt}] + \Delta\varepsilon^{l'l}_{jrt} \geqslant 0\} - y^{l'}_{jrt}) = 0. \tag{G.30}$$

Following the same steps we use to go from equation (G.24) to (G.29), we derive from equation (G.30) the following inequality

$$\mathbb{E}[y^{l'}_{ist}y^{l'}_{jrt} + y^{l}_{ist}y^{l'}_{jrt}\exp(-e^{ll'}_{ijsrt})(-(1+e^{ll'}_{ijsrt}) + \Delta\tilde{v}^{l'l}_{jrt})|\mathcal{W}_{ist},\mathcal{W}_{jrt}] \geqslant 0. \tag{G.31}$$

As the moments in equations (G.28) and (G.31) share the same function $g^{ll'}_{ijsrt}\colon \mathcal{Z}_{st}\times\mathcal{Z}_{rt}\times\Theta_{(\alpha,\beta)} \to \mathbb{R}$ and the same conditioning set, we add them to obtain:

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

with

$$\Delta\tilde{v}^{ll'}_{ist} = \beta(x^{l}_{nt} - x^{l'}_{nt}) + (\lambda^{l}_{t} - \lambda^{l'}_{t}) + \alpha(w^{l}_{st} - w^{l'}_{st}) + \delta\beta\sum_{l''=1}^{L} y^{(lt)l''}_{ist+1}(x^{l''}_{lt+1} - x^{l''}_{l't+1}),$$

$$\Delta\tilde{v}^{l'l}_{jrt} = \beta(x^{l'}_{nt} - x^{l}_{nt}) + (\lambda^{l'}_{t} - \lambda^{l}_{t}) + \alpha(w^{l'}_{rt} - w^{l}_{rt}) + \delta\beta \sum_{l''=1}^{L} y^{(lt)l''}_{ist+1}(x^{l''}_{l't+1} - x^{l''}_{lt+1}).$$

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

$$\mathbb{E}[y^{l}_{ist}y^{l}_{jrt} + y^{l'}_{ist}y^{l'}_{jrt} - y^{l}_{ist}y^{l'}_{jrt}\exp(-e^{ll'}_{ijsrt})(2 + 2e^{ll'}_{ijsrt} - (\Delta\tilde{v}^{ll'}_{ist} + \Delta\tilde{v}^{l'l}_{jrt}))|z_{st}, z_{rt}] \geqslant 0, \quad \text{(G.33)}$$

with

$$\begin{aligned}\Delta\tilde{v}^{ll'}_{ist} + \Delta\tilde{v}^{l'l}_{jrt} =& \alpha(w^{l}_{ist} - w^{l'}_{ist} + w^{l'}_{jrt} - w^{l}_{jrt}) \\ &+ \delta\beta \sum_{l''=1}^{L} y^{(lt)l''}_{ist+1}(x^{l''}_{lt+1} - x^{l''}_{l't+1} + x^{l''}_{l't+1} - x^{l''}_{lt+1}).\end{aligned} \quad \text{(G.34)}$$

Plugging equation (G.34) into equation (G.33), we obtain a moment inequality whose moment equals that in equation (G.13) when evaluated at $(\theta_\alpha, \theta_\beta) = (\alpha, \beta)$. Equations (G.33) and (G.34) thus imply Theorem 5. ■

G.3.4 First-Step Moment Inequalities: Additional Derivations

Derivation of optimal approximation points. Consider approximation points expressed as $e^{ll'}_{ijsrt} = g^{ll'}_{ijsrt}(z_{st}, z_{rt}, \theta_\alpha, \theta_\beta)$. We find the value of $g^{ll'}_{ijsrt}(z_{st}, z_{rt}, \theta_\alpha, \theta_\beta)$ that, given $z_{st} \in \mathcal{Z}_{st}$ and $z_{rt} \in \mathcal{Z}_{rt}$, minimizes the moment in equation (G.33) at each value of $(\theta_\alpha, \theta_\beta)$. Specifically, given z_{st}, z_{rt}, θ_α, and θ_β, the first-order condition of the moment in equation (G.33) with respect to the scalar $g^{ll'}_{ijsrt}(z_{st}, z_{rt}, \theta_\alpha, \theta_\beta)$ is

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

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

$$g^{ll'}_{ijsrt}(z_{st}, z_{rt}, \theta_\alpha, \theta_\beta) = \mathbb{E}[\Delta\tilde{v}^{ll'}_{ist} + \Delta\tilde{v}^{l'l}_{jrt}|z_{st}, y^{l}_{ist} = 1], \quad \text{(G.35)}$$

with $\Delta\tilde{v}^{ll'}_{ist} + \Delta\tilde{v}^{l'l}_{jrt}$ defined as in equation (G.34).

References

Andrews, Donald W.K. and Gustavo Soares, "Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection," *Econometrica*, 2010, *78*, 119–

157. [18, 20, 18, 20, 10, 11]

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. [25, 18, 19]

Dix-Carneiro, Rafael, "Trade Liberalization and Labor Market Dynamics," *Econometrica*, 2014, *82* (3), 825–885. [24, 13]

_ **and Brian K. Kovak**, "Trade Liberalization and Regional Dynamics," *American Economic Review*, 2017, *107* (10), 2908–2946. [13]

_ **and** _ , "Margins of Labor Market Adjustment to Trade," *Journal of International Economics*, 2019, *117*, 125–142. [13]

Felix, Mayara, "Trade, Labor Market Concentration, and Wages," *mimeo*, 2022. [13]

Morales, Eduardo, Gloria Sheu, and Andrés Zahler, "Extended Gravity," *The Review of Economic Studies*, 2019, *86* (6), 2668–2712. [6, 23, 25]

Szerman, Christiane, "The Labor Market Effects of Disability Hiring Quotas," *mimeo*, 2024. [13]