Drought Responses in the Western United States:
Household Location Choices and Housing Market Feedback

Mingzhou Shen[1], Qin Fan[2], Karen Fisher-Vanden[1], Douglas H. Wrenn[1], and Shan Zuidema[3]

1 Department of Agricultural Economics, Sociology, and Education, College of Agricultural Sciences, Pennsylvania State University;

2 Department of Economics, Craig School of Business, California State University, Fresno;

3 Earth Systems Research Center, University of New Hampshire.

Selected Paper prepared for presentation at the 2024 Agricultural & Applied Economics Association Annual Meeting, New Orleans, LA; July 28-30, 2024

Drought Responses in the Western United States: Household Location Choices and Housing Market Feedback

Mingzhou Shen[1], Qin Fan[2], Karen Fisher-Vanden[1], Douglas H. Wrenn[1], and Shan Zuidema[3]

1 Department of Agricultural Economics, Sociology, and Education, College of Agricultural Sciences, Pennsylvania State University;

2 Department of Economics, Craig School of Business, California State University, Fresno;

3 Earth Systems Research Center, University of New Hampshire.

May 14, 2024

Abstract

As climate change accelerates, the Western U.S. is projected to experience an increase in both the frequency and intensity of extreme droughts. This looming crisis underscores the need to better understand potential adaptive responses from households. In this study, we employ a spatial equilibrium framework to examine the influence of drought-induced water shortage on household location choices within counties across the Western U.S. Our findings are multifaceted: First, households experience reduced utility when residing outside their birth state, with significant preference heterogeneity regarding relocation. Second, water shortage impacts household location choices by reducing utility and raising rents, prompting households to move to locations where higher incomes offset these risks. Third, households are willing to pay $0.18 to avoid an additional gallon of unmet water demand. Lastly, counties experiencing water shortage have a more inelastic housing supply. Our study highlights the critical role of water scarcity in shaping population distribution and economic behaviors in the Western U.S.

Keywords: Residential Sorting; Water Shortage; Housing Market

JEL codes: Q54, R21, R23, R31, R38

1. Introduction

Since 2000, the Western United States has been experiencing some of the driest conditions on record. Particularly affected is the southwestern region, which is enduring a megadrought—the driest 22-year period in 1,200 years (Maxwell and Soulé 2009; King et al. 2024). Climate change has exacerbated this aridity, leading to more frequent, prolonged, and severe droughts (Zhang et al. 2021; Dannenberg et al. 2022; Yuan et al. 2023). Consequently, these droughts have become one of the most devastating natural hazards in the U.S., often resulting in annual economic losses totaling billions of dollars (Zhou, Leng and Peng 2018). At the same time, a significant demographic shift is occurring, with populations migrating from the Northeast and Midwest to the Western and Southern states (Maley and Hawkins 2018)—regions grappling with heightened drought risks. This paper examines whether these relocations are coincidental or a reflection of deliberate decision-making, investigating how households weigh wages, rents, and moving costs against drought risks.

To capture this sorting process, we estimate a spatial equilibrium model of household location choice. Our theoretical framework is informed by Roback (1982), while our empirical approach follows Diamond (2016). We use data from the 2020 5-year American Community Survey (ACS), focusing on heads of household and their spouses aged 19-64, with at least one member employed full-time at the time of the survey. Our analysis includes a choice set of 83 counties in the U.S. West. To quantify drought risks, we use a novel unmet water demand metric developed from the University of New Hampshire's water balance model, reflecting a 10-year average (2010-2019) of water deficit within each location.

Following the methodology established by Berry, Levinsohn and Pakes (1995), we estimate a two-stage structural model. In the first stage, we recover estimates of household preference heterogeneity with respect to moving costs and a set of mean indirect utility values for each county. The second stage involves decomposing this utility and estimating a housing supply equation using a simultaneous generalized method of moments (GMM) estimator. Specifically, in our utility equation, we instrument for wages and rents to derive estimates of the marginal utility of income, water shortage, and amenities. Simultaneously, in the housing supply equation, the model instruments for housing demand to identify the inverse elasticity of housing supply in response to geographic constraints, land-use regulations, and water shortage. The instruments employed in the GMM model are constructed based on the shift-share theory developed by Bartik (1991).

Results from the first stage of our model emphasize the importance of controlling for moving costs. We find that households generally experience a decrease in utility when residing outside their birth state. Moreover, there is significant heterogeneity in preferences related to relocation. Specifically, households with kids and those led by single or female heads exhibit a lower propensity to move from their home state compared to their counterparts. In contrast, household heads who are college-educated, over the age of 30, white, and renters show a higher likelihood of relocating. In the second stage, our utility equation results reveal that while an increase in real income generally improves household utility, this is offset by water shortage. This finding is consistent with the Rosen-Roback spatial equilibrium prediction that households require higher compensation to remain in locations with increased disaster risks—real income must rise to maintain utility constant as disaster risk escalates. Furthermore, our housing supply equation suggests that water shortage tends to drive up rents and results in a more inelastic housing supply, after controlling for geographic constraints and land-use regulations.

Using our marginal utility estimates from the second stage, we calculate that households are willing to pay $0.18 to avoid an additional gallon of unmet water demand per year. Regarding extreme weather events, only days with extreme precipitation are significant, with households willing to pay $XXX for each additional day with more than one inch of rain. Additionally, we estimate housing supply elasticity for each location, utilizing coefficients derived from our housing supply equation. Across counties in the West, the population-weighted average elasticity of supply is estimated at 0.98, with an unweighted average of 1.01. Counties like Whatcom, Imperial, Davis, Solano, and Marin feature the most inelastic markets, whereas Deschutes, Clark (Nevada), Washington, Yakima, and Clark (Oregon) counties have more elastic supplies. These elasticity estimates suggest that increasing water scarcity could significantly influence future housing market stability and demographic trends.

Our paper makes significant contributions to the existing body of literature. Building on Tiebout (1956), which explores how individuals with heterogeneous preferences choose living environments that best suit their needs, recent studies have estimated individual or household preferences for various spatial amenities (Bayer and Timmins 2007; Walsh 2007; Allen Klaiber and Phaneuf 2010). These include factors such as extreme weather or natural disasters (Timmins 2007; Fan and Bakkensen 2022; Bakkensen and Ma 2020; Wrenn 2023). However, these studies often assume a fixed housing supply or rely on a constant-elasticity assumption. Beyond wages, rents play a crucial role in households' decision-making processes regarding where to live. As heterogeneous households sort themselves across different locations, their collective choices significantly impact housing prices, particularly in regions with inelastic housing supplies where high demand can lead to sharp increases in rents. From a utility perspective, while certain locations may offer higher wages, the advantages could be offset by higher rents. If rents become too high, households may start relocating to more affordable areas, thereby exerting downward pressure on rents until there is no further incentive to move. To fully understand the interplay between labor and housing markets, it is essential to analyze both simultaneously within the framework of spatial equilibrium. Building on (Diamond 2016) analysis of workers' sorting patterns across U.S. cities, which includes a housing supply equation, our model extends this framework to incorporate the impact of water shortage on housing supply. As droughts become increasingly frequent in the Western U.S., drought-induced water shortage can restrict housing development either by limiting the availability of developable land (Cremades et al. 2021) or through regulatory measures such as water-related building moratoria (Shen, Fisher-Vanden and Wrenn 2023). Overall, our work offers valuable insights into demographic shifts and population adaptations to escalating drought risks in the West.

The remainder of the paper is organized as follows. Section 2 details the data and presents descriptive statistics for our variables. Section 3 describes the theoretical framework of our spatial equilibrium model. Section 4 explains our identification strategy for estimating the empirical models. Section 5 presents our main results, and Section 6 concludes.

2. Data

To create our county-level dataset for the Western U.S., we draw on multiple sources. Below, we outline each data source followed by a detailed explanation of how we construct the final dataset.

2.1. Demographic Data and Choice Set

We use the 1% microdata sample from the 2020 (2016-2020) 5-year ACS, administered by the United States Census Bureau. This dataset provides comprehensive individual and household-level information, including wages, housing prices, household composition, state of birth, and household location at the Public Use Micro Area (PUMA) level. Our analysis focuses on a choice set of 83 counties in the Western U.S., identified in the 2020 sample. Figure 1 maps the spatial distribution of these counties, covering the major populated areas of the West and representing 87% of the region's total population.

In our study, we consider the head-of-household to be the primary decision-maker. Our focus is on heads of household and their spouses aged 19-64, with at least one member employed full-time at the time of the survey. The demographic profiles of these heads are detailed in Table 1. To capture moving costs, we define *MovingCost* as a binary variable: it is set to one if a household's residing county is outside its birth state and zero otherwise. This definition is based on the hypothesis that relocating from familial roots to areas with different climates, job markets, and cultural norms can incur significant psychological costs (Bayer, Keohane and Timmins 2009; Klaiber 2014; Diamond 2016; Fan and Bakkensen 2022; Wrenn 2023).

2.2. Water Shortage Data

Our novel water shortage index is calculated using the water balance model. In the Western U.S., water demands are regulated by a water rights system in each water management area (WMA). Water rights are granted to users based on the order in which their claims were made. Often, water rights are overallocated, meaning more rights have been allocated than the available water in the system. This leaves some junior rights holders without water during shortages.

We collect water rights data across all sectors and construct cumulative curves for each water management area in the West. Our water balance model uses the actual water availability and demands from all sectors to determine a threshold—the cutoff date—beyond which rights holders may not receive water allocations. An earlier cutoff date indicates a more severe water shortage. However, because water rights might be systematically established earlier in some regions than in others, cutoff dates are not directly comparable across different areas. Consequently, for regions with a specified cutoff date, we calculate the log of the physically unmet water demand in the agricultural sector, weighted by population, as a proxy for the severity of water shortage for that WMA. This data is then aggregated to the county level. Figure 2 illustrates the spatial distribution of water shortage, with California, Arizona, and Denver experiencing the most acute conditions.

2.3. Weather Data

We source our weather data from the PRISM Climate Group (PRISM Climate Group, 2020). Since 1981, PRISM has been providing daily temperature and precipitation records for the contiguous U.S., with a granular 4km grid resolution. Each grid point's value is interpolated from nearby weather stations. For our analysis, we extract daily maximum and minimum temperatures, as well as precipitation data for each county over a ten-year retrospective period (2011-2020). From this data, we then calculate the average annual figures for the following variables: *HotDay* (number of days with a maximum temperature above 90 °F), *ColdDay* (number of days with a minimum temperature below 32 °F), and *ExtPrecip* (number of days with total precipitation above one inch).

Table 2 shows the summary statistics of these extreme weather variables. Notable variances are evident across counties. For instance, southern counties generally experience more extreme hot days, in contrast to the cooler conditions observed in northwestern counties. Specifically, Yuma County in Arizona consistently recorded the highest number of hot days. On the other hand, several counties in Washington had no days that reached the hot day threshold. Regarding extreme precipitation, Humboldt County in California led with the most days recorded, while Imperial County, located at the southernmost tip of California, recorded the fewest days of extreme precipitation.

2.4. Geographic Constraints

Geographic constraints refer to the proportion of land within a county that is unsuitable for residential development, as characterized by (Saiz 2010). These undevelopable areas consist of two main components. The first includes wetlands, lakes, and other internal water bodies, which we quantify using satellite-based geographic land use data from the United States Geological Survey (USGS). The second component encompasses terrains within each county with slopes exceeding 15 percent. We delineate these topographical features using the USGS Digital Elevation Model on a 90-square meter cell grid. Utilizing ArcGIS Pro, we calculate the shares of both internal water bodies and steep terrains for each county, summing them to determine the total share of undevelopable land per county. Table 2 presents the summary statistics for our geographic data, indicating an average of 21% undevelopable land across the counties.

2.5. Wharton Residential Land Use Regulation Index

We utilize the 2018 Wharton Residential Land Use Regulation Index (WRLURI), developed by Gyourko et al. (2021), as our benchmark for assessing land-use regulations. This comprehensive dataset comprises 12 subindexes, each evaluating different aspects of the regulatory landscape across approximately 2,500 U.S. communities. These subindexes highlight the influence of key stakeholders, including local councils, state legislatures, and local citizens, in shaping the regulatory environment. They also account for specific regulations such as permitting caps, density restrictions, affordable housing mandates, and impact fees. By employing a simple factor analysis, the authors aggregate and standardize these subindexes into a single index, WRLURI18, which reflects the overall regulatory stringency at the census place level. To fully capture local land-use regulations, we aggregate WRLURI18 to the county level. The summary statistics of this county-level index are presented in Table 2, where a lower (or higher) index value indicates a more lenient (or stringent) regulatory environment, with an average value of 0.47.

2.6. Data Construction

In this section, we outline the process of transforming raw data into the final estimation sample used in our spatial equilibrium model. We start by using the Census Tract to PUMA Relationship File, provided by the Census Bureau, to link observations lacking a recognizable COUNTYFP to specific census tracts and subsequently to counties. Next, we use STATEFIP and COUNTYFP codes to create a unique FIPS code identifier for each county, allowing us to accurately determine the current residence of each individual when both codes are available. We then restrict our sample to individuals residing in the Western U.S. Additionally, we calculate the number of kids below and above 19 in each household and retain only the household heads and their spouses, ensuring our data focuses on traditional housing units and excludes roommates and other unrelated individuals. Our analysis further focuses on age constraints, targeting working-age individuals between 19 to 64 years old—adults likely past high school and not yet of retirement age. We also exclude non-U.S. born individuals who may face mobility restrictions, such as visa limitations. To concentrate on the most mobile demographic, we exclude farming households and those with businesses. Moreover, we retain only those employed full-time, defined as working at least 30 hours per week and 48 weeks per year, and earning wages. We also drop households lacking home value data or not paying rent, as both wages and rents play a role in the budget constraint. Lastly, we exclude individuals living in group quarters, further refining our sample for analysis.

In our model, a budget constraint incorporates both wages and rents, for which we run separate regressions to derive their respective values. We start by using census microdata to create household wage estimates. For each individual in a county, a wage value is assigned. To mitigate bias from households endogenously sorting themselves across locations based on economic opportunities, we adopt the method developed in Dahl (2002), adjusting wages for non-random sorting. Specifically, for each individual earning a non-zero wage, we regress the log of real wages ($2000), on race, gender, age, a household head indicator, educational attainment, and a set of control functions that account for non-random sorting across states by education status. These regressions are performed separately for each county. Using the coefficients obtained from these regressions, we calculate wage estimates for individuals by zeroing out the control functions. The calculated wages are then aggregated at the household level to determine average household wages for each county. Due to separate regressions for each of the 83 counties, we summarize the findings by presenting the mean values of the coefficients and their standard errors in Appendix Table A1, instead of individual regression outputs. A more detailed description of the wage adjustment methodology is provided in Appendix—Wages and Rents.

To create our county-level rent estimates, we also use micro census data on rents and house values. Each household in the county is associated with an actual rent value for households that rent, or an estimated housing price for homeowners. To create our final household-level rent values, we combine actual rents with a user cost value of rent for homeowners. This user cost adjustment translates home prices into imputed rents that capture what a homeowner would pay if they were renting from themselves, adjusting for the benefits and costs of homeownership. We then follow previous research and hedonically adjust rent values for compositional differences in housing across space (Bayer et al. 2009; Fan, Klaiber and Fisher-Vanden 2016). We regress the log of real monthly rent ($2000) on a matrix of variables capturing the characteristics of each house and a county fixed effect that measures the price of housing services. This produces a quality-adjusted price of housing services in each county, which serves as our measure of rents. A full explanation of the estimation process for equation 2 is given in Appendix—Wages and Rents. Table A2 providing regression results for the model along with average rent values.

Our final dataset is completed by integrating a series of computed variables for each county, including wages, rents, water shortage, extreme weather events, geographical constraints, and land-use regulations. This county-level dataset forms the foundation for estimating our spatial equilibrium model and generating all economic variables within the model.

3. Spatial Equilibrium Model

We adopt the spatial equilibrium framework developed by Diamond (2016), incorporating two main components: the labor market and the housing market. Each of these components will be explored in detail in the subsequent discussion.

3.1. Labor Market

Each household i, represented by a head-of-household, chooses to live in a county j that provides them with the highest level of utility. Specifically, household inelastically supplies one unit of labor in exchange for a wage, W_j. Using this wage, they purchase a national good, C, at price P, and a local good, H_j, at price R_j, assuming that they exhibit Cobb-Douglas preferences over local and national goods. Additionally, the household gains utility from consuming a bundle of these goods, as well as from local attributes, which include our water shortage S_j, and other amenities, $\mathbf{A}_j$, and paying moving costs, M_{ij}. As described earlier, we define M_{ij} as a binary variable that takes the value of one if the household resides in the same state where the head-of-household was born, and zero otherwise. The corresponding utility and budget constraint are:

$$\begin{aligned} &\max_{C,H_j} \quad \ln(C^{1-\gamma}) + \ln(H_j^{\gamma}) + S_j + \mathbf{A}_j + M_{ij} \\ &\text{s.t.} \quad PC + R_j H_j \leq W_j, \end{aligned} \tag{1}$$

where γ signifies a household's relative preference for the local good. Under the Cobb-Douglas preference framework, γ represents the share of expenditure allocated to the local goods bundle. Following Wrenn (2023)'s paper, we assume that γ remains consistent across all counties, fixing this parameter's value at 0.2 for all models. By solving the utility maximization problem in equation 1, we can express each household's optimal locational decision via the following indirect utility function. The detailed derivation from equation 1 to equation 2 is provided in Appendix.

$$\begin{aligned} V_{ij} &= \ln(\frac{W_j}{P}) - \gamma \ln(\frac{R_j}{P}) + S_j + \mathbf{A}_j + M_{ij} \\ &= w_j - \gamma r_j + S_j + \mathbf{A}_j + M_{ij}, \end{aligned} \tag{2}$$

where $w_j = \ln(\frac{W_j}{P})$ and $r_j = \ln(\frac{R_j}{P})$. The price of the national good, P, is measured by the urban consumer price index (CPI-U) index from the Bureau of Labor Statistics, and is quantified in real 2000 U.S. dollars. For empirical estimation of our model, we assume that household utility has an idiosyncratic error term, ε_{ij}. This error term captures unobservable attributes that are uncorrelated to wages, rents, local amenities, and moving costs. The resultant indirect utility function can be expressed as:

$$V_{ij} = \beta^I (w_j - \gamma r_j) + \beta^S S_j + \beta^A \mathbf{A}_j + \eta_j + \beta^M M_{ij} + \varepsilon_{ij}, \tag{3}$$

where term η_j captures all attributes that are unobservable to the researcher. We rearrange equation 3 to move all terms that vary only by location j into a new variable δ_j. Consequently, equation 3 can be reformulated as:

$$V_{ij} = \delta_j + \beta^M M_{ij} + \varepsilon_{ij}, \tag{4}$$

and

$$\delta_j = \beta^I (w_j - \gamma r_j) + \beta^S S_j + \beta^A \mathbf{A}_j + \eta_j, \tag{5}$$

where δ_j is the mean indirect utility value shared by all households in location j. Assuming idiosyncratic preferences ε_{ij} have an independent and identically distributed type I extreme value distribution, we obtain a multinomial logit model. This model has a closed form expression for the probability of household i choosing location j:

$$P(V_{ij} \geq V_{ik} \ \forall k \neq j) = \frac{\exp^{(\delta_j + \beta^M M_{ij})}}{\sum_{k=1}^{J} \exp^{(\delta_k + \beta^M M_{ik})}}, \tag{6}$$

with its log-likelihood function given by:

$$ll = \sum_{i=1}^{N} \sum_{j=1}^{J} Y_{ij} \ln P_{ij}, \tag{7}$$

where Y_{ij} is a dummy variable indicating whether household i chooses to live in location j.

We estimate equation 7 using maximum likelihood estimation (MLE) and a contraction-mapping algorithm. This process recovers preference coefficients of households associated with moving costs $\hat{\beta}^M$ and a unique vector of fixed effects $\hat{\delta}_j$. The contraction mapping iteratively updates $\hat{\delta}_j$ within the MLE framework until the observed and predicted population shares converge. This is in contrast to estimating these terms using gradient-based searches. Berry (1994) demonstrates that including a complete set of alternative specific constants results in predicted and observed population shares coinciding as a necessary condition for maximum likelihood estimation. These location-specific constants quantify the mean indirect utility for each location, relative to a reference location that is excluded from the analysis.

Also, we derive the average household's demand for local goods, denoted as housing demand HD_j :

$$HD_j = N_j \frac{\gamma W_j}{R_j}, \tag{8}$$

where N_j is the number of households of county j. This is obtained from summing the probabilities of all households choosing a specific county, as specified in equation 6. The housing expenditure, γW_j, maintains a constant proportion to wages.

3.2. Housing Market

Our model for housing supply builds on the work developed in Saiz (2010) and is further extended by Diamond (2016). The empirical specification for the supply equation is as follows:

$$\begin{aligned} r_j &= \ln\left(\frac{R_j}{P}\right) = \kappa_j \ln(HD_j) + \varphi_j \\ \kappa_j &= \kappa + \kappa^{Geo} \exp(x_j^{Geo}) + \kappa^{Reg} \exp(x_j^{Reg}) + \kappa^S S_j \end{aligned} \tag{9}$$

where HD_j is the aggregate housing demand in county j. φ_j represents unobservables, such as the construction costs and other factors impacting house prices (rents) that are exogenous to housing demand (Gyourko and Saiz 2006; Diamond 2016). κ_j represents inverse housing supply elasticity, which is a function of the base inverse elasticity, geographic constraints, land-use regulations, and water shortage.

Geographic constraints, x_j^{Geo}, measure the share of land within a specific county that is unsuitable for development. This includes areas covered by wetlands, lakes, and other internal water bodies, as well as terrains with slopes exceeding a 15 percent (Saiz 2010). As the proportion of undevelopable land in a county increases, the elasticity of housing supply decreases. In other words, we would expect κ^{Geo} to be positive.

Similarly, our measure for local land-use regulations, x_j^{Reg}, is obtained from the Wharton survey data (Gyourko et al. 2021). Higher values in this index signify the enforcement of more restrictive real estate development policies, resulting in a less elastic housing supply. Consequently, we would also expect κ^{Reg} to be positive.

In addition to geographic constraints and land-use regulations, we further introduce a water shortage variable, S_j, into the housing supply equation. This variable retains its earlier definition: the average annual unmet water demand in county j. The intuition is that water shortage can lead to wildfire risks (Littell et al. 2016), and land degradation (Vicente-Serrano et al. 2015). These conditions make land less suitable for developments and may cause local governments to delay or halt new construction. As a result, housing supply is limited, thereby driving up rents.

In our framework, the total labor supply for a given location is determined by summing the probabilities of all households choosing that specific location. These probabilities, as outlined in equation 6, depend on how households weigh factors such as wages, rents, moving costs against drought risks when choosing between locations. We calculate the total housing demand for each location using the predicted probabilities from equation 8. The final step involves calculating housing prices and location choices, as specified in equation 9, to determine a new equilibrium where housing supply matches housing demand across all locations.

4. Econometric model

4.1. Estimation

We employ a two-stage procedure to estimate our spatial equilibrium model, aligning with methodologies previously established by Berry et al. (1995). In the first stage, we use MLE to reveal the heterogeneity in households' preferences regarding moving costs. Importantly, this stage computes the mean indirect utility for each county, shedding light on how households balance factors such as wages and rents against water shortage and other local amenities when making residential decisions. In the second stage, we decompose the estimated mean indirect utility and estimate a housing supply equation using a simultaneous GMM estimator, instrumenting for wages, rents, and housing demand.

The indirect utility function for household i, in county j is written as:

$$V_{ij} = \delta_j + \beta^M M_{ij} + \sum_{k=1}^{K} \beta^k HC_i^k M_{ij} + \varepsilon_{ij}, \tag{10}$$

and

$$\delta_j = \beta^I (w_j - \gamma r_j) + \beta^S S_j + \beta^A \mathbf{A}_j + \eta_j, \tag{11}$$

where we introduce an interaction term between k variables that describe household characteristics, HC_i^k, and the moving cost variable M_{ij}, to capture household preference heterogeneity with respect to relocation. In the first stage, we recover $\hat{\beta}^M$, $\hat{\beta}^K$, and $\hat{\delta}_j$, where $\hat{\beta}^M$ represents marginal utility associated with moving costs, and $\hat{\beta}^K$ corresponds to marginal utility related to moving costs compared to a reference group of households, both recovered through MLE. Meanwhile, $\hat{\delta}_j$ is a vector of mean indirect utility derived by contraction mapping.

Given that only relative utility matters, we normalize one of the J fixed effects to zero. Consequently, the model recovers only $J-1$ unique values.

In the second stage, we decompose mean indirect utility estimates into two components: one explained by observables, wages, rents, water shortage, amenities, as well as the other portion explained by unobservables, η_{jt}.

$$\hat{\delta}_j = \beta^I (w_j - \gamma r_j) + \beta^S S_j + \beta^A \mathbf{A}_j + \eta_j, \tag{12}$$

For our water shortage variable, S_j, and amenity variables, $\mathbf{A}_j$, based on prior work that examines the long-run impact of natural disasters and extreme weather on migration decisions (Fan and Bakkensen 2022; Wrenn 2023), we calculate them using a rolling 10-year lookback for each location.

Moving to housing supply equation, we adopt a specification similar to that used in Diamond (2016):

$$r_j = (\kappa + \kappa^{Geo} \exp(x_j^{Geo}) + \kappa^{Reg} \exp(x_j^{Reg}) + \kappa^S S_j) \ln(HD_j) + \varphi_j, \tag{13}$$

In our final estimation, we use a simultaneous GMM estimator to estimate equations 12 (labor) and 13 (housing). Our model assumes that household sorting is an equilibrium outcome, with each household selecting their optimal location based on factors such as wages, rents, moving costs, water shortage, and local amenities. The coefficients for wages and rents, β^{I}, for water shortage, β^{S}, and for amenities, β^{A}, capture the marginal utility derived from each. Furthermore, κ reflects the base inverse supply elasticity, and parameters κ^{Geo}, κ^{Reg}, and κ^{S} respectively measure how $\exp(x_j^{Geo})$, $\exp(x_j^{Reg})$, and S_j influence this inverse elasticity, capturing the responsiveness of housing supply to changes in these variables.

4.2. Instrumentation

Due to potential time-varying correlations persist among η_j, wages, w_j, and rents, r_j, within the labor equation, as well as the likely correlation between housing demand, $\ln(HD_j)$, and time-varying construction costs and other dynamic unobservables, φ_j, within the housing equation, addressing endogeneity is crucial. To tackle this, we use a Bartik-style shift-share instrument, following the approaches proposed by Diamond (2016).

The Bartik instrument is designed to capture shifts in local labor demand based on two key factors: national industry growth rates and a region's specific industrial composition. When national industry trends are on an upward trajectory, areas with a pre-existing emphasis on those industries tend to experience a rise in labor demand. This increase in labor demand boosts wages, attracting more workers to those areas or convincing existing residents to stay. Both scenarios contribute to a surge in housing demand and, subsequently, rents. A significant advantage of the Bartik instrument is its exogeneity. The growth rates of industries at the national level are influenced by broader economic dynamics, not local conditions. By integrating these national growth patterns with a region's industrial profile, we can make predictions about local wages, housing demand, and rents that are unaffected by potential local confounding factors (Bartik 1991).

We define our Bartik instrument as follows:

$$\Delta Z_{jt'}^{B} = \sum_{d=1}^{D} (w_{d,-j,t} - w_{d,-j,2000})(\frac{N_{d,j,2000}}{N_{j,2000}}), \tag{14}$$

where d defines the industry. $w_{d,-j,t}$ and $w_{d,-j,2000}$ are the average log wage of workers in industry d in period t and the year 2000, respectively, excluding county j. $N_{d,j,2000}$ is the total number of workers in industry d, in county j, in the year 2000, and $N_{j,2000}$ is the total workforce in the same location and year.

To increase our identification power, we also calculate the Bartik instrument based on the skill level of workers, specifically, those with and without a four-year degree. Recent study indicates that high-skill workers exhibit a stronger response to productivity shocks (Diamond 2016; Notowidigdo 2020). Moreover, we introduce interactions between our Bartik instrument, $\Delta Z_{jt'}^{B}$, and geographic constraints, x_j^{Geo}, as well as land-use regulations, x_j^{Reg}. Both of these variables evaluate the degree of restrictiveness in the housing supply. Specifically, increased geographic restrictions and/or more stringent land-use regulations result in lower housing supply elasticities. Therefore, the intuition behind these instruments is straightforward: shocks in labor demand spur growth in labor supply, which, in turn, increases the demand for housing. Consequently, this surge in housing demand causes rents to rise more significantly in areas characterized by lower housing supply elasticities. For exogeneity, prior study has validated that interactions between the Bartik instrument and these housing supply restrictiveness variables remain uncorrelated with the error terms in equations 12 and 13 (Diamond 2016).

We use our final set of instruments:

$$\Delta Z_{jt'} = \begin{Bmatrix} \Delta Z_{jt'}^{B,High}, \Delta Z_{jt'}^{B,Low}, \\ \Delta Z_{jt'}^{B,High} x_j^{Geo}, \Delta Z_{jt'}^{B,Low} x_j^{Geo}, \\ \Delta Z_{jt'}^{B,High} x_j^{Reg}, \Delta Z_{jt'}^{B,Low} x_j^{Reg} \end{Bmatrix}, \tag{15}$$

together with a GMM procedure. We simultaneously estimate equations 12 and 13, where $E(\eta_j \, \Delta Z_{jt'})$ and $E(\varphi_j \, \Delta Z_{jt'})$ denote the respective population moment conditions.

5. Results

In this section, we begin by presenting the results from the first stage of our spatial equilibrium model, highlighting the significant preference heterogeneity among households with respect to relocation. We then delve into our findings from the second stage, where we focus on analyzing the labor and housing equations.

5.1. First Stage

Results from the first stage are presented in Table 3. A key advantage of using a structural model is its ability to account for moving costs. The negative coefficient on *MovingCost* suggests that households receive significant negative utility in relocating away from their birth state, a finding that is consistent with previous studies (Bayer et al. 2009; Klaiber 2014; Diamond 2016; Fan and Bakkensen 2022; Wrenn 2023).

Our findings also reveal significant heterogeneity in household preferences regarding relocation. Specifically, households with kids, and those led by single or female heads exhibit a lower propensity to move from their home state compared to their counterparts. In contrast, household heads who are college-educated, over the age of 30, white, and renters, show a higher likelihood of relocating. Most importantly, this estimation stage retrieves a full set of baseline utilities for each model, $\hat{\delta}_j$. Any variables not captured in the first stage are included in these mean indirect utility estimates (TBC).

5.2. Second Stage

Results from our second-stage estimation are presented in Table 4. Column 1 is a base model that excludes the water shortage variable from both the labor and housing equations. In line with prior studies, our results confirm that an increase in income enhances household utility. Additionally, an increase in undevelopable land along with stricter land-use regulations both contribute to higher rents (Saiz 2010; Diamond 2016). In column 2, we introduce the water shortage variable into both equations. Our assumption is that water shortage directly affects household utility, for instance, by restricting lawn watering frequencies. Moreover, water shortage indirectly shapes housing supply through factors such as increased wildfire risks (Littell et al. 2016), land degradation (Vicente-Serrano et al. 2015), and delayed or halted new construction (Shen et al. 2023).

Our results show that water shortage impact household location choices by reducing utility and raising rents, with households moving to locations where higher incomes offset these risks. Additionally, we observe a significant decrease in household utility with higher *ExtPrecip* values. This suggests that extreme precipitation, which may lead to increased flooding or other adverse conditions, detrimentally affects the quality of life and living expenses, thereby reducing overall utility. Furthermore, our housing supply equation suggests that water shortage tends to drive up rents, and results in a more inelastic housing supply, after controlling for geographic constraints and land-use regulations.

Using our marginal utility estimates from the second stage, we calculate that households are willing to pay $0.18 to avoid an additional gallon of unmet water demand per year. Regarding extreme weather events, only days with extreme precipitation are significant, with households willing to pay $XXX for each additional day with more than one inch of rain. Additionally, we estimate housing supply elasticity for each location, utilizing coefficients derived from our housing supply equation. Across counties in the West, the population-weighted average elasticity of supply is estimated at 0.98, with an unweighted average of 1.01. Counties like Whatcom, Imperial, Davis, Solano, and Marin feature the most inelastic markets, whereas Deschutes, Clark (Nevada), Washington, Yakima, and Clark (Oregon) counties have more elastic supplies. These elasticity estimates suggest that increasing water scarcity could significantly influence future housing market stability and demographic trends (TBC).

6. Conclusion

As the threat of climate change escalates, the Western U.S. is anticipated to face more frequent and intense severe droughts. Understanding how households might respond to these changes becomes crucial. Our study delves into the complex relationship between household migration and housing market responses, framed within a two-stage spatial equilibrium model, placing particular emphasis on the effects of severe droughts.

In the first stage of our analysis, we identify a marked negative utility associated with moving away from one's birth state, alongside varying preferences regarding water shortage. Specifically, households with kids and those headed by single individuals or women are less likely to relocate from their home state compared to others. Conversely, household heads who are college-educated, over the age of 30, white, and renters demonstrate a greater propensity to move.

In the second stage, we offer key insights into the factors that influence household decision-making processes, including wages, rents, drought impacts, and amenities. Our findings reveal that households prefer locations with higher income. Moreover, water shortage has a direct influence on households' utilities and also shapes their location preferences through housing costs. These insights are important for both research and policy formulation, especially as policymakers confront the immediate challenges posed by our shifting climate.

References

Allen Klaiber, H., and D.J. Phaneuf. 2010. "Valuing open space in a residential sorting model of the Twin Cities." *Journal of Environmental Economics and Management* 60(2):57–77.

Bakkensen, L.A., and L. Ma. 2020. "Sorting over flood risk and implications for policy reform." *Journal of Environmental Economics and Management* 104.

Bartik, T.J. 1991. *Who benefits from state and local economic development policies?* W.E. Upjohn Institute.

Bayer, P., N. Keohane, and C. Timmins. 2009. "Migration and hedonic valuation: The case of air quality." *Journal of Environmental Economics and Management* 58(1):1–14.

Bayer, P., and C. Timmins. 2007. "Estimating equilibrium models of sorting across locations." *Economic Journal* 117(518):353–374.

Berry, S., J. Levinsohn, and A. Pakes. 1995. "Automobile Prices in Market Equilibrium." 63(4):841–890.

Berry, S.T. 1994. "Estimating discrete-choice models of product differentiation." *The RAND Journal of Economics* 25(2):242–262.

Bieri, D.S., N. V. Kuminoff, and J.C. Pope. 2023. "National expenditures on local amenities." *Journal of Environmental Economics and Management* 117.

Cremades, R., A. Sanchez-Plaza, R.J. Hewitt, H. Mitter, J.A. Baggio, M. Olazabal, A. Broekman, B. Kropf, and N.C. Tudose. 2021. "Guiding cities under increased droughts: The limits to sustainable urban futures." *Ecological Economics* 189.

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

Diamond, R. 2016. "The determinants and welfare implications of US workers' diverging location choices by skill: 1980-2000." *American Economic Review* 106(3):479–524.

Fan, Q., and L.A. Bakkensen. 2022. "Household sorting as adaptation to hurricane risk in the United States." *Land Economics* 98(2):219–238.

Fan, Q., H.A. Klaiber, and K. Fisher-Vanden. 2016. "Does Extreme Weather Drive Interregional Brain Drain in the U.S.? Evidence from a Sorting Model." *Land Economics*. Available at: https://usa.ipums.org/usa/.

Gyourko, J., J.S. Hartley, and J. Krimmel. 2021. "The local residential land use regulatory environment across U.S. housing markets: Evidence from a new Wharton index." *Journal of Urban Economics* 124.

Gyourko, J., and A. Saiz. 2006. "Construction costs and the supply of housing structure." *Journal of Regional Science* 46(4):661–680.

Himmelberg, C., C. Mayer, and T. Sinai. 2005. “Assessing high house prices: Bubbles, fundamentals and misperceptions.” *Journal of Economic Perspectives* 19(4):67–92.

King, K.E., E.R. Cook, K.J. Anchukaitis, B.I. Cook, J.E. Smerdon, R. Seager, G.L. Harley, and B. Spei. 2024. “Increasing prevalence of hot drought across western North America since the 16th century.” Available at: https://www.science.org.

Klaiber, H.A. 2014. “Migration and household adaptation to climate: A review of empirical research.” *Energy Economics* 46:539–547.

Littell, J.S., D.L. Peterson, K.L. Riley, Y. Liu, and C.H. Luce. 2016. “A review of the relationships between drought and forest fire in the United States.” *Global change biology* 22(7):2353–2369.

Maley, A.J., and D.N. Hawkins. 2018. “The Southern Military Tradition: Sociodemographic Factors, Cultural Legacy, and U.S. Army Enlistments.” *Armed Forces and Society* 44(2):195–218.

Maxwell, J.T., and P.T. Soulé. 2009. “United States drought of 2007: Historical perspectives.” *Climate Research* 38(2):95–104.

Notowidigdo, M.J. 2020. “The incidence of local labor demand shocks.” *Journal of Labor Economics* 38(3):687–725.

Pan, S. 2023. “Health, air pollution, and location choice.” *Journal of Environmental Economics and Management* 119.

Prism climate group. 2020. “Prism climate group.” Available at: https://prism.oregonstate.edu/.

Roback, J. 1982. “Wages, Rents, and the Quality of Life.” *Source: Journal of Political Economy* 90(6):1257–1278. Available at: https://about.jstor.org/terms.

Saiz, A. 2010. “The geographic determinants of housing supply.” *The Quarterly Journal of Economics* 125(3):1253–1296. Available at: https://academic.oup.com/qje/article/125/3/1253/1903664.

Shen, M., K. Fisher-Vanden, and D.H. Wrenn. 2023. “Impacts of Water-Related Building Moratoria on California’s Housing Crisis.”

Tiebout, C.M. 1956. “A Pure Theory of Local Expenditures.” Available at: https://www.jstor.org/stable/1826343.

Timmins, C. 2007. “If you cannot take the heat, get out of the cerrado... Recovering the equilibrium amenity cost of nonmarginal climate change in Brazil.” *Journal of Regional Science* 47(1):1–25.

Vicente-Serrano, S.M., D. Cabello, M. Tomás-Burguera, N. Martín-Hernández, S. Beguería, C. Azorin-Molina, and A. El Kenawy. 2015. “Drought variability and land degradation in semiarid regions: Assessment using remote sensing data and drought indices (1982-2011).” *Remote Sensing* 7(4):4391–4423.

Walsh, R. 2007. “Endogenous open space amenities in a locational equilibrium.” *Journal of Urban Economics* 61(2):319–344.

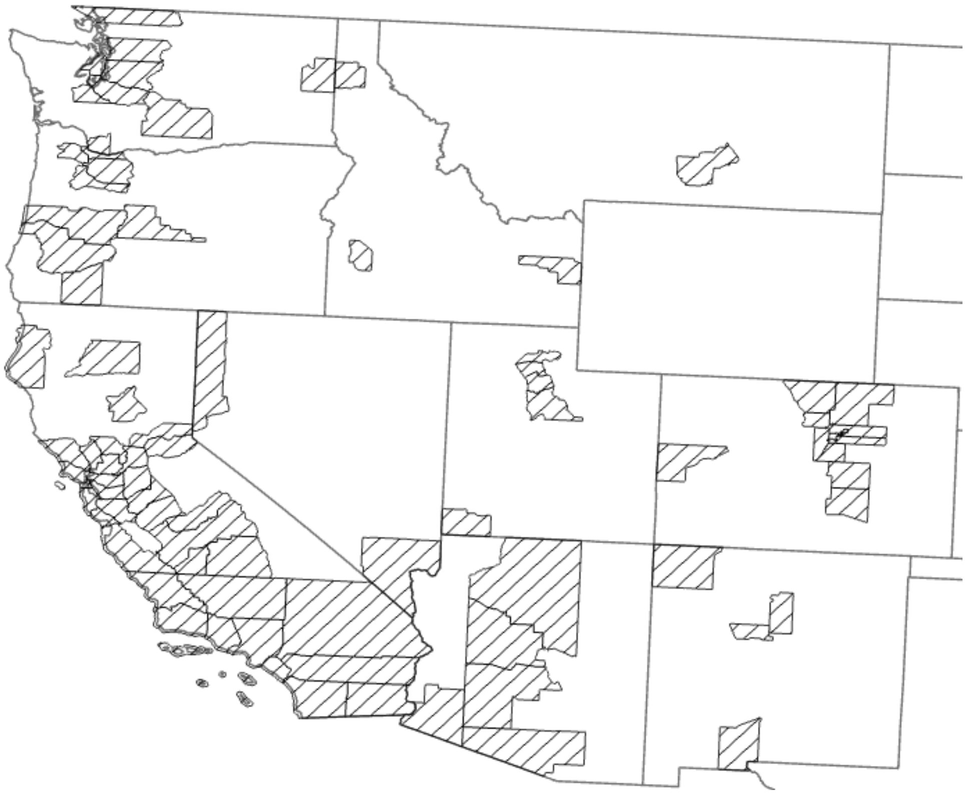
Wrenn, D.H. 2023. “The effect of natural disasters and extreme weather on household location choice and economic welfare *.” Available at: https://ssrn.com/abstract=4078002.

Yuan, X., Y. Wang, P. Ji, P. Wu, J. Sheffield, and J.A. Otkin. 2023. “A global transition to flash droughts under climate change.” *Science* 380(6641):187–191. Available at: https://www.science.org.

Zhou, Q., G. Leng, and J. Peng. 2018. “Recent changes in the occurrences and damages of floods and droughts in the United States.” *Water* 10(9):1109.

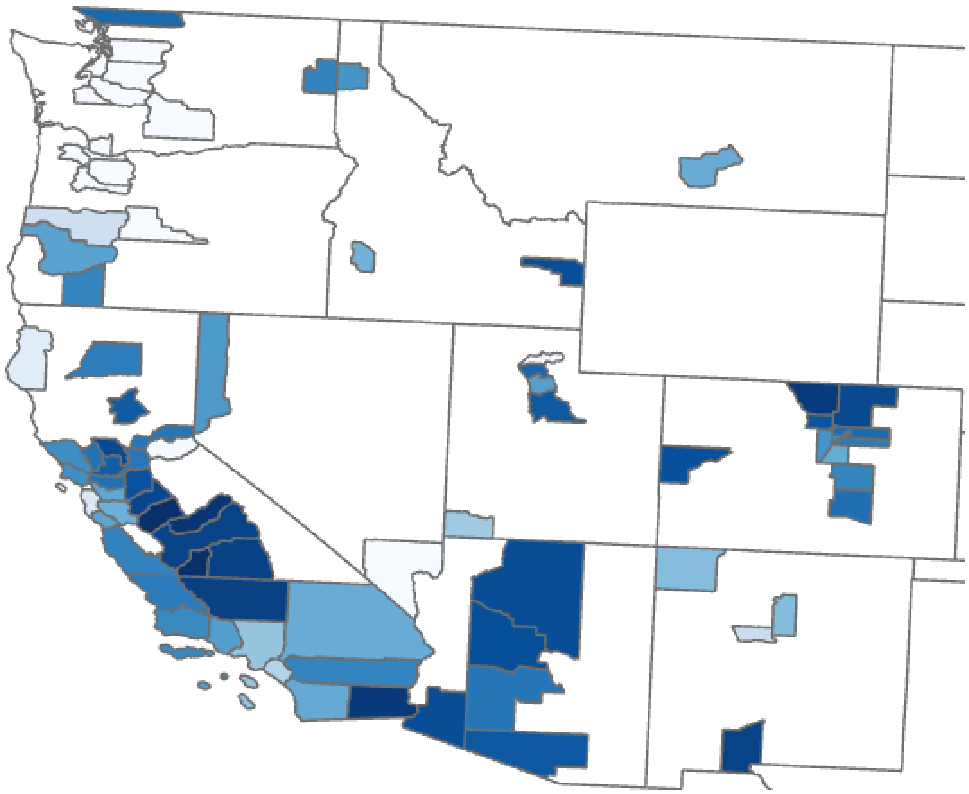
Figures

Figure 1: 83 Counties in the Western U.S. Identified in 2020 Census Sample



Sources: 1% microdata sample from the 2020 5-year (2016-2020) American Community Survey.

Figure 2: Unmet Water Demand in the West



Notes: This figure plots the log of unmet water demand (m^3/yr) for all counties in our choice set. Darker areas represent more severe water shortage.

Tables

Table 1: Summary Statistics for Data Used in the First Stage of the Model

Variables	Description	Mean	Std. Dev.
MovingCost	Whether county j is out of the head-of-household i's birth state (Yes = 1; No = 0)	0.44	0.50
Kid	Whether the head-of-household i has any kids under 19 years old (Yes = 1; No = 0)	0.37	0.48
ColGrad	Whether the head-of-household i is a college graduate (Yes = 1; No = 0)	0.52	0.50
Age30-39	Whether the head-of-household i is between 30 and 39 years old (Yes = 1; No = 0)	0.94	0.24
Age40-49	Whether the head-of-household i is between 40 and 49 years old (Yes = 1; No = 0)	0.94	0.24
Age50-64	Whether the head-of-household i is between 50 and 64 years old (Yes = 1; No = 0)	0.94	0.24
White	Whether the head-of-household i is White (Yes = 1; No = 0)	0.85	0.35
Single	Whether the head-of-household i is single (Yes = 1; No = 0)	0.46	0.50
Female	Whether the head-of-household i is female (Yes = 1; No = 0)	0.48	0.50
Renter	Whether the head-of-household i is a renter (Yes = 1; No = 0)	0.38	0.49

Notes: The demographic data are sourced from the census 2020 5-year American Community Survey (ACS) sample.

Table 2: Summary Statistics for Data Used in the Second Stage of the Model

Variables	Description	Mean	Std. Dev.
Shortage	Unmet water demand (m^3/yr) in county j	7.83	4.89
HotDay	Number of days with a maximum temperature above 90 °F in county j	31.09	32.98
ColdDay	Number of days with a minimum temperature below 32 °F in county j	73.08	63.98
ExtPrecip	Number of days with total precipitation above one inch in county j	7.00	5.85
ViolentCrime	Count of violent crime per 100K people in county j	173.60	71.23
GeoConstraint	Share of undevelopable area in county j	0.21	0.18
WRLURI	Wharton Residential Land Use Regulation Index in county j	0.47	0.83

Notes: Water shortage data are from the University of New Hampshire's water balance model. Weather data are from the PRISM Climate Group (PRISM Climate Group, 2020). Crime data are from FBI Uniform Crime Statistics 1960-2017, County and City Data Book, U.S. Census Bureau. All shortage, weather, and crime variables represent average annual, over a decade-long retrospective window, specifically, 2010-2019 for the water shortage data, 2011-2020 for weather data, and 2008-2017 for crime data. Geographic data are from the United States Geological Survey, and land-use regulations data are from the survey conducted by Gyourko et al. in 2018 (Gyourko et al. 2021).

Table 3: Results from the First Stage of the Model

	Estimates	Std. Err.
MovingCost	-3.3257***	0.0231
MovingCost X Kid	-0.1232***	0.0128
MovingCost X ColGrad	0.2644***	0.0110
MovingCost X Age30to39	0.0669***	0.0178
MovingCost X Age40to49	0.2413***	0.0185
MovingCost X Age50to64	0.3255***	0.0178
MovingCost X White	0.2758***	0.0145
MovingCost X Single	-0.1165***	0.0122
MovingCost X Female	-0.0667***	0.0108
MovingCost X Renter	0.3327***	0.0126

Notes: Results are shown for the first-stage multinomial logit models, with the dependent variable being the log of indirect utility in equation 10. The microdata used are from the 2020 ACS 5-year survey. The first column presents the estimates, and robust standard errors are presented in the second column. County fixed effects are not shown.
* Significant at 10%; ** significant at 5%; *** significant at 1%.

Table 4: Results from the Second Stage of the Model

	(1)	(2)
	A. Utility	
Income	0.224***	0.337***
	(0.058)	(0.061)
Shortage		-0.152***
		(0.043)
HotDay	-0.017***	-0.005
	(0.004)	(0.005)
ColdDay	0.001	0.001
	(0.001)	(0.001)
ExtPrecip	-0.094***	-0.065*
	(0.032)	(0.035)
ViolentCrime	-0.002	-0.002
	(0.002)	(0.002)
	B. Rents	
HD	0.908***	0.857***
	(0.043)	(0.050)
GeoConstraint X HD	0.025	0.031
	(0.034)	(0.033)
WRLURI X HD	0.011***	0.007***
	(0.002)	(0.002)
Shortage X HD		0.007***
		(0.002)

Notes: This table shows results from GMM estimation of the second stage of our spatial equilibrium model. Data include observations from 83 counties. Income enters the model in the form of a budget constraint, $w_j - \gamma \Delta r_j$, where w_j is the log of real monthly household wages, r_j is the natural log of real monthly rent, and γ is the expenditure share on local goods, set at a value of 0.2. We instrument for *Income* and $\ln(HD_j)$ using a shift-share method similar to Diamond (2016). Standard errors are clustered at the county level and shown in parentheses.
* Significant at 10%; ** significant at 5%; *** significant at 1%.

Appendix—Extra Table and Derivation

Derivation of Indirect Utility Function (equation 2) from the Utility Function (equation 1)

The original utility function and budget constraint are presented as equation A1, which corresponds to equation 1 in the main body of the paper.

$$\begin{aligned} &\max_{C,H_j} \quad \ln(C^{1-\gamma})+\ln(H_j^{\gamma})+S_j+\mathbf{A}_j+M_{ij} \\ &\text{s.t.} \quad PC+R_jH_j \le W_j, \end{aligned} \tag{A1}$$

where C is the consumption of the national good; H_j is the consumption of the local good; γ is the relative taste parameter for the local good; P is the price of the national good; R_j is the price of the local good, and W_j is the total wages of a household in county j.

To maximize the utility function, we set up the Lagrange function:

$$L(C,H_j,\lambda)=\ln(C^{1-\gamma})+\ln(H_j^{\gamma})+S_j+\mathbf{A}_j+M_{ij}-\lambda(PC+R_jH_j-W_j). \tag{A2}$$

Taking partial derivatives of L with respect to C and H_j gives us the following first-order conditions.

For C:

$$\frac{\partial L}{\partial C}=\frac{1-\gamma}{C}-\lambda P=0, \tag{A3}$$

which yields:

$$\lambda=\frac{1-\gamma}{CP}. \tag{A4}$$

For H_j:

$$\frac{\partial L}{\partial H_j} = \frac{\gamma}{H_j} - \lambda R_j = 0, \tag{A5}$$

which yields:

$$\lambda = \frac{\gamma}{H_j R_j}. \tag{A6}$$

Now, we equate the two expressions for λ to first solve for C:

$$\frac{1-\gamma}{CP} = \frac{\gamma}{H_j R_j}, \tag{A7}$$

which yields:

$$C = \frac{(1-\gamma) H_j R_j}{\gamma P}. \tag{A8}$$

We substitute C from equation A8 into the budget constraint A1 to solve for H_j:

$$
\begin{aligned}
&P\left(\frac{(1-\gamma)H_j R_j}{\gamma P}\right)+R_j H_j = W_j,\\
&\left(\frac{(1-\gamma)H_j R_j}{\gamma}\right)+R_j H_j = W_j,\\
&H_j R_j\left(\frac{(1-\gamma)}{\gamma}+1\right) = W_j,\\
&H_j R_j\left(\frac{1-\gamma+\gamma}{\gamma}\right) = W_j,\\
&H_j = \frac{\gamma W_j}{R_j}.
\end{aligned}
\tag{A9}
$$

Now that we have H_j, we can substitute this back into equation A8 to solve for C:

$$
\begin{aligned}
&C = \frac{(1-\gamma)\left(\frac{\gamma W_j}{R_j}\right)R_j}{\gamma P},\\
&C = \frac{(1-\gamma)W_j}{P}.
\end{aligned}
\tag{A10}
$$

So, we have derived expressions for C and H_j in terms of the wages W_j and the prices P and R_j. Since the utility function is of the Cobb-Douglas form and the budget constraint is linear, we can also derive expressions for C and H_j without using the Lagrange method. This means that a household will allocate the share γ of their budget to the local good H_j, and the share $1-\gamma$ of their budget to the national good C, which is consistent with our results.

Next, we substitute C and H_j into the utility function A1 to get the indirect utility:

$$
V_{ij} = \ln\left(\frac{(1-\gamma)W_j}{P}^{1-\gamma}\right)+\ln\left(\frac{\gamma W_j}{R_j}^{\gamma}\right)+S_j+\mathbf{A}_j+M_{ij},
\tag{A11}
$$

expanding the logarithms by the power rule of logarithms $\ln(a^b)=b\ln(a)$, we get:

$$V_{ij}=(1-\gamma)\ln\left(\frac{(1-\gamma)W_j}{P}\right)+\gamma\ln\left(\frac{\gamma W_j}{R_j}\right)+S_j+\mathbf{A}_j+M_{ij}. \tag{A12}$$

Notice that the expenditure shares $(1-\gamma)$ and γ in the indirect utility function will not change the maximization problem since they are constants. Therefore, they can be omitted from the logarithmic terms, yielding:

$$V_{ij}=(1-\gamma)\ln\left(\frac{W_j}{P}\right)+\gamma\ln\left(\frac{W_j}{R_j}\right)+S_j+\mathbf{A}_j+M_{ij}. \tag{A13}$$

Finally, we distribute $(1-\gamma)$ and γ:

$$V_{ij}=\ln\left(\frac{W_j}{P}\right)-\gamma\ln\left(\frac{W_j}{P}\right)+\gamma\ln\left(\frac{W_j}{R_j}\right)+S_j+\mathbf{A}_j+M_{ij}, \tag{A14}$$

which simplifies to the indirect utility function presented as equation 2 in the main body of our paper:

$$V_{ij}=\ln\left(\frac{W_j}{P}\right)-\gamma\ln\left(\frac{R_j}{P}\right)+S_j+\mathbf{A}_j+M_{ij}. \tag{A15}$$

Wages and Rents

Our models capture the spatial equilibrium outcome, represented by the shares of households in each county. These outcomes are produced as households choose their optimal location based on the wages, rents, and local amenities available in each location. Therefore, it is essential to estimate wages and rents for every location.

Wages

To estimate wages in each location, we employ a wage model that considers several key variables as follows:

$$\begin{aligned}\ln\left(\frac{W_{ij}}{P}\right) = w_{ij} = \beta_j + \beta_j^{White} White_i + \beta_j^{Male} Male_i + \beta_j^{Age} Age_i + \beta_j^{Age^2} Age_i^2 + \beta_j^{Head} Head_i \\ + \beta_j^{HSG} HSG_i + \beta_j^{SC} SC_i + \beta_j^{CG} CG_i + \beta_j^{CF} CF_i + \beta_j^{CF^2} CF_i^{\,2} + \varepsilon_{ij}^{w},\end{aligned} \tag{A16}$$

where the dependent variable is the log of the monthly wages of an individual residing in county j, adjusted for inflation using CPI-U price index, P. Each regression includes a set of individual characteristics that are most influential in determining wages, such as race (whether the person is white), gender (whether the person is a male), age and its square (a proxy for experience), head-of-household status, and educational attainment. The education levels are categorized as less than high school $(LHSG)$, high school graduate (HSG), some college (SC), and college graduate (CG). The last two variables in the model, CF_i and CF_i^2 are control functions that capture all unobserved factors affecting an individual’s wage for a given location.

The intuition is that people do not randomly choose where to live. Their decisions are often influenced by a variety of observed and unobserved factors. For instance, consider individual A with a master's degree. When seeking employment, this individual is more likely to relocate to a region rich in high-tech sector opportunities and the prospect of higher wages. In contrast, individual B, who has a high school degree, might opt for a region with more manufacturing or blue-collar job opportunities. Their relocation decisions, clearly, are not random. They are rooted in their educational backgrounds and the potential for higher wages in locations that match their skill sets. Also, there might be unobserved state-specific factors influencing wages. By including a control function that reflects both education background and migration path, we can control for the potential non-random sorting of individuals.

To calculate our control functions, we first categorize individuals based on their education levels. We then determine migration shares for each educational group based on established relationships between birth state and current state of residency (Wrenn 2023; Pan 2023). As a result, each individual in our dataset is allocated a specific share, conditional on their educational background and migration path, described as:

$$CF = Share[State_a, State_b \mid LHSG, HSG, SC, CG], \tag{A17}$$

which is the share of people with a given level of education born in state a and currently living in state b. These shares are the CF_i control functions in equation A17, aiding in controlling for potential endogenous sorting between states based on education levels.

With the control functions in place, equation A17 is estimated separately for each location. Following each estimation, we remove the control function terms and predict wages for each individual using only demographic variables. We then aggregate the wages of individuals across households to get total household income, and compute an average wage for each county. These averages serve as our county estimates of household wages, $\hat{W}_j$.

Given that we have 83 separate regressions, it is challenging to present all the results altogether. Instead, we gather estimates and their respective standard errors from each regression and take a yearly average. These averages are summarized in Table A1. Overall, our results are consistent with prior research: Whites typically earn more, men tend to have higher wages, income rises with age, non-head-of-household generally earn less, and higher education links to higher wages.

Rents

Similar to wages, we restrict our sample to focus only on homeowners. Each household is associated with an estimated house value, HP_{hj}, where h denotes the household, j represents the county. To compute an equivalent monthly rent value for each homeowner, we leverage the concept of user costs. This idea treats homeowners as if they were renting from themselves. Specifically, it converts actual house values by multiplying them with a coefficient. This coefficient translates the values to reflect the amount a household would pay if they were renting the same property, adjusting for any benefits and costs of homeownership relative to renting (Bieri, Kuminoff and Pope 2023; Himmelberg, Mayer and Sinai 2005). We source user costs data from Wrenn (2023). Since Wrenn's data are presented at the community zone level, we average them out to the state level, assuming all counties within a state share a uniform state-level user cost, UC_j. Moreover, we utilize the user costs data from 2010 as a proxy for our 2020 sample. We then multiply each household's house value by the user costs to derive the rent values, $R_{hj} = UC_j * HP_{hj}$.

We generate county rent estimates, $\hat{R}_j$, using individual house values, R_{hj}. Building on prior research, we implement rent regressions that hedonically adjust rents. This allows us to net out the pricing effects of housing characteristics, such as the structure of the house and its age (Fan and Bakkensen 2022; Wrenn 2023). We estimate the model as follows:

$$\begin{aligned} \ln\left(\frac{R_{hj}}{P}\right) = r_{hj} = \beta^{Acre} Acre_{hj} + \beta^{Unit} Unit_{hj} + \beta^{Room} Room_{hj} \\ + \beta^{Bedroom} Bedroom_{hj} + \beta^{Age} Age_{hj} + \rho_j + \varepsilon_{hj}, \end{aligned} \tag{A18}$$

where the dependent variable is the natural log of the monthly rents of a household residing in county j, adjusted for inflation using CPI-U price index, P. Each model includes a full set of controls for housing characteristics, such as acres, numbers of units in the building, numbers of rooms and bedrooms, and the age of the house. We also introduce a county fixed effect, ρ_j, which quantifies the rental price for each county. This rent index serves as our estimates for county rents in the second stage of our model. Results from our hedonic rent regressions are presented in Table A2.

Table A1: Results from Dahl Wages Regressions

	$\bar{\beta}$	$\bar{S.E.}$
White	0.092	0.053
Male	0.476	0.032
Age	0.174	0.010
AgeSqrd	-0.002	0.001
Head	0.177	0.035
HSG	0.336	0.234
SG	0.473	0.241
CG	0.875	0.293
N	83	
R^2	0.302	

Notes: This table displays average coefficient values and standard errors from the wage regression in equation A16. The regression runs separately for 83 counties. The columns show the mean coefficients and standard errors taken across all regressions. Average R^2 values are shown at the bottom of the table.

Table A2: Results from Hedonic Rents Regressions

largelot	0.482***
	(0.024)
unit2	-0.735***
	(0.135)
unit3	1.472***
	(0.047)
unit4	1.392***
	(0.049)
unit5	1.575***
	(0.068)
unit6	1.338***
	(0.055)
unit7	1.311***
	(0.058)
unit8	1.351***
	(0.071)
unit9	1.457***
	(0.052)
unit10	1.560***
	(0.061)
room2	-0.005
	(0.102)
room3	0.115
	(0.103)
room4	0.154
	(0.104)
room5	0.195*
	(0.106)
room6	0.280**
	(0.107)
room7	0.374***
	(0.106)
room8	0.451***
	(0.105)
room9	0.636***
	(0.104)
bedroom2	-0.188**
	(0.082)
bedroom3	-0.005
	(0.086)
bedroom4	0.132
	(0.086)
bedroom5	0.261***
	(0.088)

bedroom6	0.384***
	(0.089)
builtyr2	-0.134***
	(0.012)
builtyr3	-0.254***
	(0.014)
builtyr4	-0.296***
	(0.016)
builtyr5	-0.399***
	(0.023)
builtyr6	-0.458***
	(0.021)
builtyr7	-0.389***
	(0.027)
N	336,882
R^2	0.492
Mean Rent ($)	1,509

Notes: This table displays results from hedonic models estimated using 2020 census microdata on housing rents. Model controls for acres, building size in units, numbers of rooms and bedrooms, and age of the structure. The models also include a full set of county fixed effects (not shown), which are used to generate the quality-adjusted rent values used in the model. Mean Rent is the unlogged, average monthly rent, in $2000.