A Test of Market Definition in Logit Demand Models

C. Luke Watson

FDIC

April 2026

Abstract

Market definition is fundamental to demand estimation, yet in practice it is often chosen by researcher judgment. I recast market definition in aggregate logit models as a fixed-effects problem: the market-level term enters as a nuisance component that can be absorbed by sufficiently fine submarket fixed effects. As a result, demand parameters can be estimated consistently without specifying the true market in advance, provided the researcher can identify a grouping nested within the true market. I then adapt the test of Papke and Wooldridge (2023) to compare this consistent-but-inefficient estimator with structural estimates based on explicit market definitions. Monte Carlo simulations show good size and power: the test rejects market definitions that are too coarse to satisfy the required moment conditions and does not reject correctly specified ones. An application to airline demand favors airport-pair markets over city-pair markets. The approach brings formal evidence to a choice that is usually left to researcher judgment.

[*] Views opinions expressed reflect those of the authors and do not necessarily reflect those of the United States. I thank Milena Almagro and Oren Ziv for helpful conversations. All mistakes are my own.

1 Introduction

Market definition is a first-order issue in structural demand analysis because it affects elasticities, welfare calculations, and merger simulation. Yet in applied work it is often contestable and is frequently chosen on the basis of researcher judgment. This paper shows how to use the method of Papke and Wooldridge (2023) to test market definitions in aggregate logit models of the type studied by Berry (1994).

Papke and Wooldridge (2023) (hereafter PW) develop a cluster-robust Hausman test for the appropriate level of hierarchical fixed effects in linear panel models. I show how to apply their test in aggregate logit demand models to evaluate competing market definitions under the maintained assumptions of the logit framework.

The paper makes two main points. First, aggregate logit demand can be estimated consistently without specifying the true market definition, provided the researcher can identify a submarket grouping nested within the true market and the data contain sufficient within-submarket variation. In that case, submarket fixed effects absorb the market-level component that would otherwise require an explicit market definition. Second, this observation yields a practical test: one can compare the resulting consistent-but-inefficient estimator to structural estimates based on candidate market definitions, some of which may be misspecified. Under the maintained assumptions, the PW test should fail to reject the true market definition and reject definitions that are too coarse.

Three caveats are important. First, if the parameter of interest is only the price coefficient and the instrument is uncorrelated with prospective market trends, then the test may have limited ability to distinguish among market definitions. Second, the test is not about whether a given fixed-effect structure is structurally required in the outcome equation; it is about whether that structure is sufficient to make the IV moment condition hold. Third, if the instrument is invalid under the true market definition, the test may favor a finer but false definition simply because finer fixed effects restore the exclusion restriction. Accordingly, the 2SLS–PW test should be interpreted as a necessary but not sufficient diagnostic for market definition in aggregate logit demand models.

I apply the test to airline demand using the framework of Berry and Jia (2010) and

Aryal et al. (2024). The application compares three nested market definitions: airport pairs (for example, DCA–ORD), origin-airport by destination-city pairs (DCA–Chicago), and city pairs (Washington–Chicago). The results consistently reject city-pair markets relative to airport-pair markets, while the evidence distinguishing airport-pair from airport-to-city markets is mixed. In the appendix, I present a banking application comparing tracts, counties, MSAs, and commuting zones using a bank-cost-shock instrument. There the test does not reject equality across market definitions, which may reflect either broad banking markets or the non-geographic correlation structure of the instrument.

Other potential applications include grocery demand (Albuquerque and Bronnenberg, 2009; Dubois et al., 2022), housing markets (Mast, 2023; Watson and Ziv, 2025), and health insurance markets (Abraham et al., 2017; Saltzman, 2019). In all three settings, the finest observable unit is never itself treated as a market but is strictly nested within all candidate market definitions, making them natural motivating examples for the testing procedure developed in this paper.

The rest of the paper proceeds as follows. Section 2 presents the PW test for OLS and 2SLS. Section 3 introduces the aggregate logit setup and shows how the PW test applies in that setting. Section 4 presents the Monte Carlo design and results. Section 5 reports the airline application. Section 6 concludes.

2 Papke and Wooldridge (2023)

This section presents the method of Papke and Wooldridge (2023) for OLS and then extends it to 2SLS.[1]

2.1 Papke–Wooldridge (2023) Single-Coefficient Test

Papke and Wooldridge (2023, Section 3.2) propose a simple and robust procedure for testing whether the level of fixed effects (FEs) used in a linear regression is sufficient to guarantee unbiased estimation of a parameter of interest. The goal is to determine whether a coarser FE specification (e.g., group FEs) adequately controls for unobserved heterogeneity, or whether a finer specification (e.g., unit FEs rather than group) is required.

[1] PW note that their approach extends to fixed-effects IV settings, but they do not provide the derivation.

The test relies on the asymptotic linear representation of the estimator for a single coefficient β in the model:

$$y_{it} = \beta x_{it} + a_i + \lambda_t + u_{it}, \tag{1}$$

where a_i and λ_t are unobserved effects that may or may not be fully removed by the chosen FE structure.

Let $\tilde{x}_{it}$ and $\tilde{u}_{it}$ denote residuals after partialling out a given set of FEs (e.g., time and unit effects). Under the null hypothesis that the chosen FEs are sufficient, the moment condition $\mathsf{E}[\tilde{x}_{it}\tilde{u}_{it}] = 0$ holds.

We can write the OLS estimation error as:

$$\hat{\beta} - \beta = \left(\mathsf{E}[\tilde{x}_{it}^2]\right)^{-1} \mathsf{E}[\tilde{x}_{it}\tilde{u}_{it}], \tag{2}$$

and its influence function can be estimated by

$$q_{it} = \frac{\tilde{x}_{it}\hat{u}_{it}}{\overline{\tilde{x}^2}}, \tag{3}$$

where $\hat{u}_{it} = \tilde{y}_{it} - \hat{\beta}\tilde{x}_{it}$.[2]

To compare two FE structures – for example, *unit* and *group* FEs – PW propose computing q for each case, differencing them, and testing whether the mean difference is zero:

$$\text{Test statistic:} \qquad \mathtt{t} = \frac{\hat{\beta}_1 - \hat{\beta}_2}{\widehat{se}(q_1 - q_2)}. \tag{4}$$

The standard error $\widehat{se}(q_1 - q_2)$ is obtained from a regression of $q_1 - q_2$ on a constant with cluster-robust variance estimation. The resulting $\mathtt{t}$ statistic has an asymptotic standard normal distribution under the null that both FE choices yield consistent estimates.

Intuitively, the test asks: *does the coarser FE specification leave residual correlation between the regressor and the error term?* If so, the finer FE structure is required for consistency.

[2]Concretely, q_{it} is the product of the residualized regressor and the regression residual, scaled by the second moment of the residualized regressor. More intuitively, it measures how much observation (i, t) locally pushes $\hat{\beta}$, after partialling out the fixed effects.

2.2 PW for Two-Stage Least Squares (2SLS)

While PW hint at the application to 2SLS, they do not show it. I therefore derive the corresponding result for 2SLS. When the regressor x_{it} is endogenous, the same logic extends to an instrumental-variables setting. Suppose we observe instruments Z_{it} and estimate β by 2SLS after residualizing y, x, and Z with respect to a given FE structure. Let the residualized variables be $(\tilde{y}_{it}, \tilde{x}_{it}, \tilde{Z}_{it})$. The population moment condition is $\mathsf{E}[\tilde{Z}_{it}\tilde{u}_{it}] = 0$.

For 2SLS, the fitted regressor $\hat{x}_{it} = P_Z \tilde{x}_{it}$ is the projection of $\tilde{x}_{it}$ onto the (residualized) instrument space. The analogue to equation 2 is then:

$$\hat{\beta} - \beta = \left(\mathsf{E}[\tilde{x}_{it}\hat{x}_{it}]\right)^{-1} \mathsf{E}[\hat{x}_{it}\tilde{u}_{it}], \tag{5}$$

and the per-observation 2SLS influence function becomes:

$$q_{it} = \frac{\hat{x}_{it}\hat{u}_{it}}{\overline{\hat{x}\tilde{x}}}. \tag{6}$$

Note that two instances of $\tilde{x}$ from equation 3 are replaced by the fitted regressor, $\hat{x}$, to correspond to the 2SLS estimator.

Computing q under two FE structures and testing the difference, as in the OLS case, yields a valid Hausman-type comparison for IV models. The test statistic is

$$t = \frac{\hat{\beta}_{\text{coarse}} - \hat{\beta}_{\text{fine}}}{\widehat{se}(q_{\text{coarse}} - q_{\text{fine}})}. \tag{7}$$

The asymptotic logic mirrors the OLS version: if the coarser FE specification is sufficient for the IV orthogonality condition $\mathsf{E}[Z'u] = 0$, both estimators are consistent and the difference should vanish asymptotically.

Interpretation of PW 2SLS Test Conceptually, this test is not about the *structural* necessity of the fixed effects in the outcome equation, but about whether the chosen FE structure is sufficient to make the moment condition $\mathsf{E}[Z'u] = 0$ hold. That is, the relevant question is not whether a_i or λ_t enter the true model for y or x, but whether residual correlation between the instrument and the error term remains after the chosen residualization.

Equivalently, the test can be interpreted as asking whether the *reduced form* requires the finer FE structure to achieve instrument validity. Even if the structural relationship between y and x does not depend on any fixed effects, the moment condition may still fail at a coarser FE level if Z has unobserved components correlated with the disturbance u.

For example, in a hierarchical panel with states and counties, one might generate instruments of the form

$$Z_{it} = z_{c(i)} + \nu_{it}, \tag{8}$$

where $z_{c(i)}$ is a county-level component correlated with the county effect in u_{it}. With state FEs, the county-level component of Z remains in the residualized instruments, violating $\mathsf{E}[Z'u] = 0$ and rendering the 2SLS (and reduced form) inconsistent. With county FEs, the residualization removes $z_{c(i)}$, restoring $\mathsf{E}[Z'u] = 0$ and consistency.

3 Aggregate Logit Demand Model Estimation and Testing

This section introduces the aggregate logit model, shows how it can be estimated without specifying the true market, and explains how to apply the PW test.

3.1 Model Notation

Let j index products, c index local geographies (e.g., counties), s index mutually exclusive groups of local geographies (e.g., states), and let t index time. Let m denote the true market definition, which the econometrician does not observe.

Assume the conditions under which aggregate demand at the market level can be written as:

$$s_{jmt} = \frac{\mathrm{e}^{\beta X_{jmt} + \alpha p_{jmt} + \delta_{jmt}}}{1 + \sum_{k \in J} \mathrm{e}^{\beta X_{kmt} + \alpha p_{kmt} + \delta_{kmt}}}, \tag{9}$$

where s are market shares, X are observed product characteristics, p are observed prices, and δ are unobserved product characteristics (e.g., 'product quality'). Using the arguments from Berry (1994), we can invert the share equation to arrive at the following linear equation:

$$\ln[s_{jmt}] - \ln[s_{0mt}] = \beta X_{jmt} + \alpha p_{jmt} + \delta_{jmt}, \tag{10}$$

where s_0 is the outside good share. We assume that $\mathsf{E}[X'\delta] = 0$ but $\mathsf{E}[p \cdot \delta \mid X] \neq 0$, and so we look for instruments that satisfy exclusion and relevancy for price.

The standard approach is to estimate equation 10 by 2SLS, which may or may not include fixed effects at the market-time level. This requires choosing a market definition and calculating product and outside good shares.

Implications of Market Definition. Logit demand elasticities depend on market shares; for example, the own-price elasticity is $\varepsilon_j = \alpha p_j(1 - s_j)$. Consumer welfare calculations also depend on market shares. Merger analysis depends on the market definition when solving for post-merger prices and quantities.

3.2 Markets, Submarkets, and Identification

A key insight to the derivation below is that market share is a ratio of quantity sold and total market quantity:

$$s_{jmt} = \frac{q_{jmt}}{Q_{mt}} \implies \ln[s_{jmt}] = \ln[q_{jmt}] - \ln[Q_{mt}]. \tag{11}$$

This implies we can rewrite equation 10 as:

$$\begin{aligned} \ln[q_{jmt}] &= (\ln[Q_{mt}] + \ln[s_{0mt}]) + \beta X_{jmt} + \alpha p_{jmt} + \delta_{jmt} & (12) \\ &= \lambda_{mt} + \beta X_{jmt} + \alpha p_{jmt} + \delta_{jmt} & (13) \\ &= \beta X_{jmt} + \alpha p_{jmt} + (\delta_{jmt} + \lambda_{mt}). & (14) \end{aligned}$$

In equation 12 I have grouped the market quantities $\{Q_{mt}, s_{0mt}\}$, and in equation 13 I have relabeled them as a single nuisance variable λ_{mt}. The last equation 14 groups the unobservable product quality and the market shifter together to emphasize that these are simply variables we need to ensure are uncorrelated with our instrument. This framing also highlights that the aggregate logit model is ultimately a regression of log quantity on prices, and the α parameter is a semi-elasticity.

In aggregate logit markets, it is common to define markets by geography-time; i.e., a county-year will be a market for the structural model.

Submarkets Suppose that potential market definitions can be ordered into hierarchical groupings: $g_1 \subset g_2 \subset \cdots \subset g_G$. For example, states, counties, and Census tracts. If one is willing to assume that the true market is larger than some grouping, $g \subset m$, then one can call g a submarket.

Identification Equation 14 implies that the true market definition need not be known to estimate α (or β). Let $\xi = \delta + \lambda$. Consistent estimation requires only that $\mathsf{E}[Z'\xi \mid X, D_g] = 0$ for a set of controls X and submarket fixed effects D_g. Thus, if a known submarket grouping exists and the data contain sufficient within-submarket variation, the utility parameters can be estimated consistently without specifying the true market boundary. For example, if the true spatial market is believed to be either the state or the MSA, county-by-time fixed effects can still deliver consistent estimates of the aggregate logit parameters.[3] Again, the correctly specified logit demand model would be the most efficient estimation of the parameters; however, it is not necessary for consistent estimation.

4 Monte Carlo Design

This section describes a Monte Carlo experiment designed to evaluate the performance of the proposed market-definition test in a controlled environment. Each Monte Carlo draw simulates data from a random coefficients logit demand model with a known hierarchical market structure and a deliberately contaminated price instrument. The experiment assesses whether the PW test correctly favors sufficiently fine fixed-effect structures and rejects specifications that are too coarse.

4.1 Hierarchical Market Structure

The simulated economy consists of a three-level geographic hierarchy. There are $S = 50$ states, each containing $M_s = 15$ markets, for a total of $M = 750$ markets. Each market contains $G_m = 5$ tracts, yielding $G = 3{,}750$ tracts in total. The nesting structure is

$$\text{Tract } g \subset \text{Market } m \subset \text{State } s.$$

[3] Watson and Ziv (2025) apply this logic in their pass-through test in urban housing markets using Census tract by year FEs despite housing markets being larger geographically but generally unknown.

The true demand market is the market (middle) level, while tracts serve as submarkets in the sense of Section 3.2.

Each market contains approximately 25 products on average, drawn from a normal distribution and bounded between 20 and 35 products. Products are assigned to multi-product firms, with each market containing approximately 15 firms and each firm operating between 1 and 3 products. Products are randomly assigned to tracts within their market. The market structure is held fixed across Monte Carlo replications, while product-level variables are redrawn in each iteration.

4.2 Preferences and Demand

Consumer preferences follow a random coefficients logit model. For consumer i evaluating product j in market m, indirect utility is:

$$u_{ijm} = \beta_0 + \nu_{i0} + \alpha p_{jm} + (\beta_1 + \nu_{i1})x_{1,jm} + \beta_2 x_{2,jm} + \xi_{jm} + \varepsilon_{ijm}, \tag{15}$$

where p_{jm} is price, $x_{1,jm}$ and $x_{2,jm}$ are observed product characteristics drawn independently from $U(0,2)$, ξ_{jm} is an unobserved demand shock, and ε_{ijm} is a Type I extreme value error. The random coefficients (ν_{i0}, ν_{i1}) are drawn independently from $N(0,1)$, introducing consumer heterogeneity in preferences for the outside option and for characteristic x_1.

The true parameter values are $\alpha = -1.5$, $\beta_0 = 2$, $\beta_1 = 2$, and $\beta_2 = 2$. The unobserved demand shock is drawn as $\xi_{jm} \sim N(0, \sigma_\xi^2)$ with $\sigma_\xi = 2.5$, independently across products. This relatively large variance in ξ generates substantial share heterogeneity within markets, which in turn produces meaningful variation in equilibrium markups and price endogeneity through the Nash–Bertrand pricing channel.

Given preferences, product-level market shares and equilibrium prices are computed using the random coefficients logit framework implemented in `pyblp` (Conlon and Gortmaker, 2020). The resulting structural estimating equation takes the standard Berry inversion form:

$$\log s_{jm} - \log s_{0m} = \beta_0 + \alpha p_{jm} + \beta_1 x_{1,jm} + \beta_2 x_{2,jm} + \xi_{jm}. \tag{16}$$

4.3 Costs, Instruments, and Endogenous Prices

Marginal costs follow a linear specification:

$$c_{jm} = \gamma_0 + \gamma_1 x_{1,jm} + \gamma_2 x_{2,jm} + \gamma_3 z_{jm} + \omega_{jm}, \tag{17}$$

where z_{jm} is a cost shifter, $\omega_{jm} \sim N(0, \sigma_\omega^2)$ is an idiosyncratic cost shock with $\sigma_\omega = 0.4$, and $(\gamma_0, \gamma_1, \gamma_2, \gamma_3) = (0.5, 0.5, 0.5, 0.75)$.

The cost shifter z_{jm} serves as the candidate instrument and is constructed to contain both product-level and market-level components:

$$z_{jm} = \tilde{z}_{jm} + \zeta_m, \tag{18}$$

where $\tilde{z}_{jm} \sim N(0, \sigma_{\tilde{z}}^2)$, where $\sigma_{\tilde{z}} = 0.4$, is a product-specific cost shock and ζ_m is a market-level cost component.

The key feature of the design is that the market-level cost shock is correlated with a market-level demand shifter λ_m:

$$\lambda_m \sim N(0, \sigma_\lambda^2), \tag{19}$$

$$\zeta_m = \rho \lambda_m + \sqrt{1 - \rho^2} \cdot \eta_m, \quad \eta_m \sim N(0, \sigma_\zeta^2), \tag{20}$$

with $\sigma_\lambda = 0.4$, $\sigma_\zeta = 0.3$, and $\rho = 0.5$. This correlation implies that the instrument z_{jm} is correlated with market-level demand conditions, violating the exclusion restriction $\mathsf{E}[z_{jm} \cdot u_{jm}] = 0$ unless market-level fixed effects are included to absorb the common demand component.

Given marginal costs and demand, equilibrium prices are computed by solving firms' Nash–Bertrand pricing first-order conditions under the random coefficients logit demand system, accounting for multi-product firm ownership within each market.

4.4 Estimation Strategies

Each simulated dataset is used to estimate the price coefficient α under competing specifications, organized into two main approaches: the direct fixed effects approach and the structural demand approach. Table 1 summarizes all specifications.

Direct 2SLS Approach (Models 0–2). The direct approach regresses log market shares on prices, product characteristics, and fixed effects at varying levels of geographic aggregation, using the cost shifter Z as an instrument for price:

$$\log s_{jg} = \alpha p_{jg} + \beta_1 x_{1,jg} + \beta_2 x_{2,jg} + \lambda_g + \xi_{jg}, \tag{21}$$

where λ_g represents fixed effects at the tract, county, or state level. This approach does not require specifying a market definition for computing shares; instead, it relies on fixed effects to absorb market-level demand shifters.

Model 0 uses tract fixed effects and serves as the benchmark for the Papke–Wooldridge test. Because tracts are strictly nested within the true county markets, tract fixed effects absorb the market-level demand component and yield consistent estimates under the maintained assumptions. Model 1 uses county fixed effects (the true market level), and Model 2 uses state fixed effects (coarser than the true market). The comparison of Models 0, 1, and 2 forms the basis for the PW test of fixed effects adequacy.

Structural Demand Approach (Models 3–10). The structural approach estimates the aggregate logit demand model under explicit market definitions. Inverting the market share equation yields the estimating equation:

$$\log s_{jm} - \log s_{0m} = \beta_0 + \alpha p_{jm} + \beta_1 x_{1,jm} + \beta_2 x_{2,jm} + \xi_{jm}, \tag{22}$$

where s_{jm} and s_{0m} are the product and outside good shares computed under the assumed market definition. Unlike the direct approach, this specification requires the researcher to define market boundaries in order to compute shares.

Models 3–8 use the correct county market definition but vary the instrument and fixed effects structure. Model 3 uses the cost shifter Z, which is valid under the correct market definition. Models 4, 5, and 6 use BLP instruments computed at the county, tract, and state levels, respectively.[4] Comparing Models 4, 5, and 6 reveals whether the geographic level of BLP instrument construction matters when the market definition is correct.

Models 7 and 8 add county fixed effects to the correctly specified structural model. Because the market definition is already correct, these fixed effects are not necessary for consistency but may affect efficiency. These specifications test whether adding fixed effects to a correctly specified structural model is harmless, as suggested by the identification argument in Section 3.2.

Models 9 and 10 use incorrect market definitions with the cost instrument. Model 9 defines markets at the tract level (finer than truth), while Model 10 defines markets at the state level (coarser than truth). Both specifications compute shares at the wrong geographic level, inducing measurement error in the dependent variable and potentially biasing the price coefficient. These specifications illustrate the consequences of market misspecification in structural demand estimation.

4.5 Papke–Wooldridge Tests

Each simulated dataset is used to conduct the PW test comparing the benchmark tract fixed effects estimator to estimators using coarser fixed effects.

Benchmark Estimator. The benchmark estimator uses tract fixed effects and the cost shifter Z as an instrument:

$$\hat{\alpha}^{(0)} \equiv \text{2SLS with tract fixed effects.}$$

Fixed-Effect Comparisons. The PW test is applied to compare the benchmark to estimators using market and state fixed effects:

$$\hat{\alpha}^{(0)} \text{ vs } \hat{\alpha}^{(1)}, \qquad \hat{\alpha}^{(0)} \text{ vs } \hat{\alpha}^{(2)}.$$

[4] BLP instruments include the count of own-firm and rival-firm products and the sum of own-firm and rival-firm characteristics x_1 and x_2, computed at the indicated geographic level.

Table 1 – Monte Carlo Estimation Specifications

Model	Description	Market Def.	Instrument	Fixed Effects
		Direct-2SLS		
0	Direct-Tract (Baseline)	–	Z	Tract
1	Direct-County	–	Z	County
2	Direct-State	–	Z	State
		Correct market definition (county)		
3	Market-C + Cost IV	County	Z	None
4	Market-C + BLP-C	County	BLP (county)	None
5	Market-C + BLP-T	County	BLP (tract)	None
6	Market-C + BLP-S	County	BLP (state)	None
7	Market-C + Cost IV + FE-C	County	Z	County
8	Market-C + BLP-C + FE-C	County	BLP (county)	County
		Incorrect market definition		
9	Market-T + Cost IV	Tract	Z	None
10	Market-S + Cost IV	State	Z	None

Note: All specifications estimate demand by 2SLS with clustering at the state level. "Market Def." indicates the geographic level used to compute product shares s_{jm} and outside good shares s_{0m}; "–" indicates no market definition is required. The true market is the county. Z denotes the cost shifter instrument. BLP instruments include the count of own-firm and rival-firm products and the sum of own-firm and rival-firm characteristics x_1 and x_2, computed at the indicated geographic level.

Under the null that a candidate fixed-effect structure is sufficiently fine, both estimators are consistent and should not differ asymptotically. Rejection indicates that the coarser specification fails to purge the correlation between the instrument and the error term.

Standard errors for the PW test statistic are computed using differences in influence functions with clustering at the state level, following Section 2.2.

4.6 Monte Carlo Implementation

For each of the $R = 500$ Monte Carlo replications, I draw market-level shocks (λ_m, ζ_m) and product-level variables $(x_1, x_2, \tilde{z}, \xi, \omega)$ for all products. Given these primitives, I compute equilibrium prices and shares using `pyblp`, construct the BLP instruments at the tract, market, and state levels, and estimate all model specifications. I then apply the PW test by comparing Models 1 and 2 to the benchmark Model 0. Across replications, I record the distribution of the estimated price coefficients, their bias relative to the true parameter $\alpha = -1.5$, and the rejection frequencies of the PW test at the 5% significance level.

4.7 Monte Carlo Results

Table 2 reports summary statistics for the estimated price coefficient $\hat{\alpha}$ across 500 Monte Carlo replications.

Table 2 – Monte Carlo Results: Price Coefficient Estimates

Model	Description	Mean $\hat{\alpha}$	Std. Dev.	Bias
	Direct 2SLS Approach			
0	Direct + Tract FE (benchmark)	−1.486	0.067	0.014
1	Direct + County FE	−1.487	0.064	0.013
2	Direct + State FE	−0.862	0.056	0.638
	Structural: Correct Market (County)			
3	Market-C + Cost IV	−1.490	0.051	0.010
4	Market-C + BLP-C	−1.283	0.558	0.217
5	Market-C + BLP-T	−1.150	1.094	0.350
6	Market-C + BLP-S	−1.742	0.524	−0.242
7	Market-C + Cost IV + FE-C	−1.487	0.064	0.013
8	Market-C + BLP-C + FE-C	−1.497	0.505	0.003
	Structural: Incorrect Market			
9	Market-T + Cost IV	−0.863	0.056	0.637
10	Market-S + Cost IV	−0.836	0.055	0.664

Note: Results based on 500 Monte Carlo replications. True parameter: $\alpha = -1.5$. Bias = $\bar{\hat{\alpha}} - \alpha$.

Direct 2SLS Models. The benchmark tract FE estimator (Model 0) and county FE estimator (Model 1) are both approximately unbiased, with mean estimates of −1.486 and −1.487. State FE (Model 2) exhibits substantial attenuation bias (mean = −0.862), confirming that coarse fixed effects fail to purge instrument contamination.[5]

Structural Models with Correct Market. Model 3 (correct market, cost IV) performs well with mean −1.490 and the smallest variance. Because this is the correct structural specification with a strong instrument, it is the most efficient estimate. Despite using the correct market definition, Models 4-6 perform poorly: the mean estimates are imprecise and in some cases materially biased, with substantially larger variance. As the derivation

[5] While unreported, this coefficient estimate is close to the OLS estimate that does not account for price endogeneity.

in Section 3.2 suggests, Model 7 yields essentially the same result as Model 1. Interestingly, adding county fixed effects to the correctly specified model with BLP IVs yields unbiased estimates, a substantial improvement relative to Model 4, although the variance remains higher than in Models 0, 1, and 3.

Structural Models with Incorrect Market. Both tract-level (Model 9) and state-level (Model 10) market definitions produce severely biased estimates around −0.85, demonstrating that share mismeasurement from incorrect market boundaries cannot be overcome by a valid instrument alone. In Appendix A, we show that adding either market or submarket FEs to a misspecified problem can lead to unbiased estimates.

PW Test Performance. The PW test rejects at 4.0% when comparing Models 0 and 1 (close to the 5% nominal size) and at 100% when comparing Models 0 and 2 (correct power against the invalid specification).

Table 3 – PW test Rejection Rates

Comparison	Alternative Specification	Rejection Rate
Model 0 vs. Model 1	Direct-County + Z	0.04
Model 0 vs. Model 2	Direct-State + Z	1.00

Note: Rejection rates at the 5% significance level across 500 Monte Carlo replications. The benchmark (Model 0) uses tract fixed effects with the cost instrument Z. Standard errors are computed using clustered influence function differences at the state level.

5 Application: Air Travel Demand

I apply the test to U.S. domestic airline demand using the replication data from Aryal et al. (2024), originally drawn from the DOT DB1B Airline Origin and Destination Survey. This setting is well-suited to the exercise because the candidate market definitions are naturally hierarchical: airport pairs are nested within origin-airport by destination-city pairs, which are nested within city pairs. The application asks whether the PW test can distinguish among these market definitions in a canonical aggregate-demand setting.

5.1 Data and Market Hierarchy

The data come from the second quarter of each year from 1998 to 2016. The DB1B is a 10% sample of airline tickets, and I follow the standard practice of scaling passenger counts by ten to recover estimated totals. Market size is defined as

$$M_m = 0.9 \times \sqrt{\text{pop}_o \times \text{pop}_d},$$

where origin and destination populations are 2010 Census CBSA populations. The outside share is computed as $s_{0m} = 1 - \sum_j s_{jm}$ at each candidate market definition. After standard sample restrictions, the common estimation sample contains 863,182 carrier-market-year observations across 18,982 airport pairs and 16 carriers.

I consider three nested market definitions. The finest market is a directional airport pair, such as DCA–ORD. The intermediate market is origin airport by destination city, so that DCA–Chicago pools ORD and MDW at the destination. The coarsest market is the city pair, such as Washington–Chicago. Airport-to-MSA assignments are based on a static IATA-to-CBSA crosswalk. By construction, each finer definition is strictly nested within the next coarser definition, which is the structure required for the PW test.

5.2 Empirical Setup

I estimate the price-income coefficient under two approaches. The first is the direct estimator implied by Section 3.2,

$$\ln[q_{jmt}] = \alpha(p_{jmt}/y_{ot}) + \beta X_{jmt} + \mu_j + \lambda_{mt} + u_{jmt}, \tag{23}$$

where the market definition enters through interacted market-by-year fixed effects. The second is the standard structural logit estimator,

$$\ln[s_{jmt}] - \ln[s_{0mt}] = \alpha(p_{jmt}/y_{ot}) + \beta X_{jmt} + \mu_j + \lambda_m + \lambda_t + \xi_{jmt}, \tag{24}$$

where shares are recomputed under each candidate market definition and fixed effects enter separately.

In both specifications, the endogenous regressor is the carrier's average fare normalized by origin-market income, p_{jmt}/y_{ot}. The included controls are the carrier's nonstop passenger share and mean routing circuity. I report results for three excluded-instrument sets: own cost shifters, BLP-style rival characteristics, and the combination of the two. Construction details are reported in Appendix B. Table 8 shows that first-stage strength is comfortably above conventional thresholds in all specifications.

5.3 Results

Table 4 reports the second-stage estimates. Three patterns stand out. First, the own cost-shifter IVs yield stable and economically sensible negative coefficients across market definitions in both the direct and structural specifications. Second, in the direct specification, the BLP-rival and combined instrument sets become more negative as the market definition becomes coarser. Third, in the structural logit, the BLP-rival and combined sets yield positive coefficients despite strong first stages, suggesting that without granular interacted fixed effects these instruments retain demand-side variation that violates the exclusion restriction.

Table 5 reports the PW tests for the direct IV estimator. The main finding is straightforward: the test rejects the city-pair market in nearly all comparisons and, under the BLP-rival and combined instrument sets, also rejects the intermediate origin-airport by destination-city definition relative to the airport pair. Only one comparison fails to reject at the 5% level: the own-cost-shifter specification comparing the fine and intermediate definitions ($p = 0.34$). Overall, the evidence favors the airport pair as the most defensible market definition in this application, while allowing that the intermediate definition may also be adequate under the pure cost-shifter IV set.

Substantively, the results suggest that travelers do not treat airports serving the same metropolitan area as perfect substitutes within a route. Aggregating to the city-pair level pools heterogeneous choice situations and leaves residual market-level variation that the coarser fixed effects do not absorb. In that sense, the application illustrates the main value of the test: it converts a judgment call about market boundaries into an empirical restriction that can be assessed formally.

Table 4 – Price/Income Coefficient ($\hat{\alpha}$)

Market definition	Airport Pair (1)	Orig.Airport–Dest.City (2)	City Pair (3)
Panel A: Direct, Interacted market-by-year FE			
Own cost shifters	−3.19	−3.20	−3.23
	(0.14)	(0.16)	(0.16)
BLP rival chars	−1.63	−2.55	−2.63
	(0.09)	(0.15)	(0.14)
Cost + BLP	−1.44	−2.21	−2.33
	(0.07)	(0.09)	(0.09)
Market × year FE	Yes	Yes	Yes
Carrier FE	Yes	Yes	Yes
N	863,182	863,182	863,182
Panel B: Structural Logit, Separate market + year FEs			
Own cost shifters	−2.59	−2.58	−2.63
	(0.13)	(0.13)	(0.14)
BLP rival chars	1.02	1.03	0.92
	(0.03)	(0.03)	(0.03)
Cost + BLP	0.57	0.60	0.51
	(0.02)	(0.02)	(0.02)
Market FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Carrier FE	Yes	Yes	Yes
N	863,209	794,987	730,308

Note: This table reports the price/income coefficient $\hat{\alpha}$ from direct and structural logit specifications across three nested market definitions. Panel A reports the direct approach, in which the market definition governs the granularity of the interacted market-by-year fixed effects. Panel B reports the structural logit approach, in which the market definition governs both the level at which shares are computed and the market fixed effects. All specifications include carrier fixed effects and two carrier-route-year controls. Cluster-robust standard errors are clustered at the market level.

Table 5 – Papke–Wooldridge Market Definition Tests

IV Set	Comparison	$\hat{\alpha}_{\text{diff}}$	SE	p	Rej.
Own cost shifters	Fine v. Interm.	−0.01	0.015	0.343	
	Fine v. Coarse	−0.05	0.023	0.036	⋆
	Interm. v. Coarse	−0.03	0.015	0.023	⋆
BLP rival chars	Fine v. Interm.	−0.91	0.048	0.000	⋆
	Fine v. Coarse	−1.00	0.063	0.000	⋆
	Interm. v. Coarse	−0.09	0.036	0.017	⋆
Cost + BLP	Fine v. Interm.	−0.77	0.037	0.000	⋆
	Fine v. Coarse	−0.90	0.049	0.000	⋆
	Interm. v. Coarse	−0.13	0.028	0.000	⋆

Note: This table reports Papke–Wooldridge tests comparing $\hat{\alpha}$ across pairs of nested market definitions: airport pair (fine), origin-airport × destination-city (intermediate), and city pair (coarse). Standard errors are cluster-robust at the coarser market level. ⋆ denotes rejection at the 5% level.

6 Conclusion

This paper shows how the fixed-effects test of Papke and Wooldridge (2023) can be used to study market definition in aggregate logit demand models. By rewriting the aggregate logit model as a regression of log quantities on prices with a market-level nuisance component, I show that consistent estimation of demand parameters does not require the researcher to know the true market definition in advance, provided sufficiently fine submarket fixed effects are available and the relevant IV moment conditions hold.

This observation makes it possible to compare a consistent-but-inefficient estimator based on fine fixed effects with structural estimators that rely on explicit market definitions. In this setting, the PW test asks whether a proposed market definition is fine enough to render the instrument orthogonal to the structural error. Definitions that are too coarse are rejected; correctly specified definitions should not be.

The Monte Carlo results and the airline application show how the procedure can be used in practice. At the same time, the test has clear limits: it can rule out overly coarse market definitions, but it is only a necessary and not a sufficient condition for correct specification. Even so, it gives applied researchers a disciplined and transparent way to evaluate market-definition choices that are otherwise often made informally.

References

Abraham, J., C. Drake, D. W. Sacks, and K. I. Simon (2017): "Demand for health insurance marketplace plans was highly elastic in 2014–2015," *NBER Working Paper*.

Albuquerque, P. and B. J. Bronnenberg (2009): "Estimating Demand in Consumer Packaged Goods Markets Using Scanner Data," *Journal of Marketing Research*, 46, 356–372.

Aryal, G., D. J. Campbell, F. Ciliberto, and E. A. Khmelnitskaya (2024): "Common Subcontracting and Airline Prices," *Review of Economics and Statistics*, forthcoming.

Begenau, J. and E. Stafford (2023): "Uniform rate setting and the deposit channel," Tech. rep., Manuscript.

Berry, S., M. Carnall, and P. T. Spiller (2006): "Airline Hubs: Costs, Markups and the Implications of Customer Heterogeneity," in *Advances in Airline Economics, Vol. 1: Competition Policy and Antitrust*, ed. by D. Lee, Elsevier, 183–214.

Berry, S. and P. Jia (2010): "Tracing the Woes: An Empirical Analysis of the Airline Industry," *American Economic Journal: Microeconomics*, 2, 1–43.

Berry, S. T. (1994): "Estimating Discrete-Choice Models of Product Differentiation," *RAND Journal of Economics*, 25, 242–262.

Conlon, C. and J. Gortmaker (2020): "Best practices for differentiated products demand estimation with pyblp," *The RAND Journal of Economics*, 51, 1108–1161.

Dubois, P., R. Griffith, and M. O'Connell (2022): "The Use of Scanner Data for Economics Research," *Annual Review of Economics*, 14, 723–745.

Granja, J. and N. Paixao (2023): "Bank consolidation and uniform pricing," Tech. rep., Manuscript.

Mast, E. (2023): "Local Effects of New Housing in Low-Income Areas," *Journal of Urban Economics*, 133, 103383.

Papke, L. E. and J. M. Wooldridge (2023): "A Simple, Robust Test for Choosing the Level of Fixed Effects in Linear Panel Data Models," *Empirical Economics*, 64, 2683–2701.

Saltzman, E. (2019): "Demand for Health Insurance: Evidence from the California and Washington ACA Exchanges," *Journal of Health Economics*, 66, 102–120.

Watson, C. L. and O. Ziv (2025): "A Test for Pricing Power in Urban Housing Markets," *Review of Economics and Statistics*, Accepted.

A Additional Monte Carlo Results

This appendix reports results for additional structural demand specifications that examine whether fixed effects can mitigate bias from market misspecification. Table 6 summarizes these specifications.

Table 6 – Additional Structural Model Specifications

Model	Description	Market Def.	Instrument	Fixed Effects
Correct market with finer fixed effects				
A1	Market-C + Cost IV + FE-T	County	Z	Tract
A2	Market-C + BLP-C + FE-T	County	BLP (county)	Tract
Market too fine (tract) with fixed effects				
A3	Market-T + Cost IV + FE-T	Tract	Z	Tract
A4	Market-T + BLP-C + FE-T	Tract	BLP (county)	Tract
Market too coarse (state) with fixed effects				
A5	Market-S + Cost IV + FE-C	State	Z	County
A6	Market-S + BLP-C + FE-C	State	BLP (county)	County

Notes: All specifications estimate demand by 2SLS with clustering at the state level. "Market Def." indicates the geographic level used to compute product shares s_{jm} and outside good shares s_{0m}. The true market is the county. Z denotes the cost shifter instrument. BLP instruments include the count of own-firm and rival-firm products and the sum of own-firm and rival-firm characteristics x_1 and x_2, computed at the indicated geographic level.

Correct Market with Finer Fixed Effects (Models A1–A2). These specifications use the correct county-level market definition but include tract fixed effects, which are finer than necessary. Because tracts are nested within counties, tract fixed effects absorb county-level variation and may reduce efficiency without improving consistency. These specifications test whether "over-controlling" with unnecessarily fine fixed effects is harmful.

Market Too Fine with Fixed Effects (Models A3–A4). These specifications define markets at the tract level, which is finer than the true county-level market. The dependent variable $\log s_{jg} - \log s_{0g}$ uses tract-level shares, which differ from the true county-level shares. Models A3 and A4 add tract fixed effects and use either the cost instrument or BLP instruments computed at the true county level. These specifications test whether fixed effects or correctly-leveled BLP instruments can rescue a model with markets defined too narrowly.

Market Too Coarse with Fixed Effects (Models A5–A6). These specifications define markets at the state level, which is coarser than the true county-level market. The dependent variable uses state-level shares, aggregating across the true county markets. Models A5 and A6 add county fixed effects (the true market level) and use either the cost instrument or BLP instruments computed at the county level. These specifications test whether adding the correct-level fixed effects can rescue a model with markets defined too broadly.

The key question is whether the fixed effects can compensate for the share mismeasurement induced by incorrect market definitions. Because the structural inversion computes $\log s_{jm} - \log s_{0m}$ using the assumed market definition, incorrect boundaries create measurement error in the dependent variable that fixed effects cannot fully address.

Results Table 7 reports results for specifications examining whether fixed effects can mitigate market misspecification.

Table 7 – Appendix Results: Fixed Effects and Market Misspecification

Model	Description	Mean $\hat{\alpha}$	Std. Dev.	Bias
A1	Market-C + Cost IV + FE-T	−1.486	0.067	0.014
A2	Market-C + BLP-C + FE-T	−1.490	0.537	0.010
A3	Market-T + Cost IV + FE-T	−1.486	0.067	0.014
A4	Market-T + BLP-C + FE-T	−1.490	0.537	0.010
A5	Market-S + Cost IV + FE-C	−1.487	0.064	0.013
A6	Market-S + BLP-C + FE-C	−1.497	0.505	0.003

Notes: Results based on 500 replications. True $\alpha = -1.5$.

All appendix specifications yield approximately unbiased estimates. This suggests that sufficiently fine fixed effects can absorb the bias from market misspecification—at least in this simulation design where the primary source of endogeneity operates at the market level. However, the structural interpretation of parameters estimated under incorrect market definitions remains problematic for downstream analyses (e.g., elasticities, welfare) that depend on correctly measured market shares.

B Additional Details for the Airline Application

This appendix collects implementation details for the airline application in Section 5. The goal is to keep the main text focused on the market-definition test itself while documenting the construction of the estimation sample, the candidate market definitions, the estimating equations, and the instrument sets in a single place.

B.1 Data and Sample Construction

The application uses the replication data from Aryal et al. (2024), based on the DOT DB1B Airline Origin and Destination Survey. The DB1B is a 10% sample of airline tickets, and the analysis uses the second quarter of each year from 1998 to 2016. Following Berry and Jia (2010) and Berry et al. (2006), passenger counts are multiplied by ten to recover estimated totals.

The raw sample contains 1,123,368 carrier-market-year observations for 16 ticketing carriers. Market size is defined as

$$M_m = 0.9 \times \sqrt{\text{pop}_o \times \text{pop}_d},$$

where pop_o and pop_d are 2010 Census CBSA populations for the origin and destination markets. The outside share is computed as

$$s_{0m} = 1 - \sum_j s_{jm}$$

at each candidate market definition.

The estimation sample imposes the same restrictions across market definitions. I drop thin markets with fewer than 200 estimated passengers, infrequent carriers, observations with outside share outside the interval $[0.5, 0.999]$, and singleton observations. The resulting common sample used for the direct IV and PW comparisons contains 863,182 observations across 18,982 airport pairs and 16 carriers. The median outside share is 0.987, close to the value reported by Berry and Jia (2010).

B.2 Candidate Market Definitions

The application compares three candidate market definitions arranged in a strict hierarchy.

1. **Airport pair (fine):** a directional airport-to-airport market, such as DCA–ORD.
2. **Origin-airport × destination-city (intermediate):** the origin airport is held fixed while destination airports within the destination MSA are pooled, so that DCA–Chicago combines ORD and MDW.
3. **City pair (coarse):** the origin MSA and destination MSA are each pooled, as in Washington–Chicago.

Each market definition is indexed separately by year. Airport-to-MSA assignments are based on a static IATA-to-CBSA crosswalk covering the 170 airports in the estimation sample. By construction, each finer definition is strictly nested within the next coarser definition, which is the hierarchical structure required by the PW procedure.

B.3 Estimating Equations

The main text reports results from two estimators.

Direct estimator. The direct IV estimator is

$$\ln[q_{jmt}] = \alpha(p_{jmt}/y_{ot}) + \beta X_{jmt} + \mu_j + \lambda_{mt} + u_{jmt}, \tag{25}$$

where q_{jmt} is carrier passenger quantity, p_{jmt}/y_{ot} is the price-income ratio, X_{jmt} is the vector of included controls, μ_j are carrier fixed effects, and λ_{mt} are market-by-year fixed effects. Under this approach, the market definition enters only through the fixed effects.

Structural logit estimator. The structural logit estimator is

$$\ln[s_{jmt}] - \ln[s_{0mt}] = \alpha(p_{jmt}/y_{ot}) + \beta X_{jmt} + \mu_j + \lambda_m + \lambda_t + \xi_{jmt}, \tag{26}$$

where product and outside shares are recomputed at each candidate market definition. Under this approach, the market definition affects both the dependent variable and the fixed effects structure.

The endogenous regressor in both specifications is

$$\text{price\_income}_{jmt} = \frac{p_{jmt}}{y_{ot}/1000}, \tag{27}$$

where p_{jmt} is the carrier's average fare and y_{ot} is origin-market median household income. This scaling gives α the interpretation of a marginal utility of income.

B.4 Included Controls

The included controls are

$$X_{jmt} = \left(\text{NonstopPassengerShare}_{jmt},\ \text{MeanCircuity}_{jmt}\right). \tag{28}$$

Nonstop passenger share. This variable is the fraction of a carrier's passengers in a carrier-market-year cell who travel on nonstop itineraries. At the itinerary level, an itinerary is classified as nonstop when no second flight segment is observed. The carrier-market-year measure is then constructed as the passenger-weighted share of traffic that is nonstop.

Mean circuity. Mean circuity is the passenger-weighted average of itinerary circuity, where

$$\text{circuity} = \frac{\text{total itinerary distance}}{\text{nonstop market distance}}. \tag{29}$$

A value of one corresponds to a nonstop itinerary, while larger values indicate more circuitous routing.

B.5 Excluded Instrument Sets

The empirical application reports results for three excluded-instrument configurations.

B.5.1 Own Cost Shifters

The own cost-shifter instrument set is

$$Z^{\text{cost}}_{jmt} = \left(\text{HasNonstop}_{jmt}, \text{MeanOwnedRegional}_{jmt}, \text{MeanUnownedRegional}_{jmt}\right). \tag{30}$$

HasNonstop. This indicator equals one if the carrier offers at least one nonstop itinerary in the market-year.

MeanOwnedRegional and MeanUnownedRegional. These variables are passenger-weighted averages of itinerary-level regional subcontracting measures and summarize the extent to which the carrier's service on the route is operated by owned regional affiliates or by unowned regional partners.

B.5.2 BLP-Style Rival Characteristics

The BLP-style rival instrument set is

$$Z_{jmt}^{\text{BLP}} = (\text{RivalNonstopCount}_{jmt}, \text{RivalPctNonstop}_{jmt}, \text{RivalMeanCircuity}_{jmt}, \\ \text{RivalMeanOwnedReg}_{jmt}, \text{RivalMeanUnownedReg}_{jmt}, \text{NumRivals}_{mt}). \quad (31)$$

These variables are constructed as leave-own-out market aggregates. For each carrier in a market-year, the rival instruments summarize the characteristics of competing carriers in the same market-year, excluding the carrier's own observation. Rival nonstop count is the number of rival carriers that offer at least one nonstop itinerary. Rival nonstop share is the leave-own-out average nonstop passenger share among rivals. Rival circuity and rival regional usage are the corresponding leave-own-out averages of service characteristics. Number of rivals is the number of carriers present in the market-year.

B.5.3 Combined Cost + BLP Set

The third instrument set stacks the own cost shifters and BLP-style rival characteristics together.

B.6 Aggregation Across Market Definitions

For the direct IV specifications, the same control and instrument definitions are used across the three candidate market definitions, with the market definition entering only through the fixed effects.

For the structural logit specifications, the data are re-aggregated to each candidate market definition before estimation. When this occurs, carrier-level variables such as fares, nonstop passenger share, circuity, regional usage, and network size are aggregated using passenger-weighted means, while `has_nonstop` is updated as the maximum across the underlying finer routes. The rival-characteristic instruments are then recomputed at the relevant market definition using leave-own-out formulas.

B.7 First-Stage Results

Table 8 reports the first-stage Kleibergen–Paap rk Wald F statistics for each instrument set under each market definition and estimation approach.

Table 8 – First-Stage Kleibergen–Paap rk Wald F Statistics

Market definition	Airport Pair (1)	Orig.Airport–Dest.City (2)	City Pair (3)
Panel A: Direct, Interacted market-by-year FE			
Own cost shifters	207.8	155.6	159.9
BLP rival chars	129.6	57.3	65.9
Cost + BLP	137.4	90.4	107.5
Market × year FE	Yes	Yes	Yes
Carrier FE	Yes	Yes	Yes
N	863,182	863,182	863,182
Panel B: Structural Logit, Separate market + year FEs			
Own cost shifters	151.1	137.4	125.1
BLP rival chars	291.7	281.2	273.1
Cost + BLP	241.2	231.2	221.8
Market FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Carrier FE	Yes	Yes	Yes
N	863,209	794,987	730,308

Note: This table reports first-stage Kleibergen–Paap rk Wald F statistics from regressions of the price/income ratio on the excluded instruments. Panel A uses the direct specification with interacted market-by-year fixed effects. Panel B uses the structural logit specification with separate market and year fixed effects. All specifications include carrier fixed effects and the two carrier-route-year controls.

The first stages are strong in all cases. Accordingly, the unusual second-stage behavior of the BLP-rival and combined instrument sets in the structural logit does not appear to be driven by weak identification.

B.8 Additional Discussion of the Second-Stage Results

The main text emphasizes the broad patterns in Table 4. A few additional remarks are useful.

First, the own cost-shifter IVs yield stable negative coefficients across candidate market definitions in both the direct and structural logit specifications. This stability is consistent with the view that these instruments primarily capture variation that is less sensitive to the precise level of market aggregation.

Second, the BLP-rival and combined instrument sets behave differently across the two estimators. In the direct specification, the estimates become more negative as the market definition becomes coarser. This is exactly the kind of pattern the PW test is designed to detect, since coarser fixed effects may leave residual market-level demand variation in the structural error.

Third, in the structural logit specification, the BLP-rival and combined instrument sets generate positive coefficients despite strong first stages. The natural interpretation is that, in this implementation, the rival-characteristic instruments still contain demand-side variation when the specification uses only separate market and year effects rather than interacted market-by-year effects. That pattern is informative about instrument validity under alternative fixed-effect structures, even if the structural estimates themselves are not economically plausible.

B.9 Robustness: Carrying Forward Fine-Level Rival IVs

The baseline structural logit specification recomputes the leave-own-out rival instruments separately at each candidate market definition after aggregation. An alternative approach is to hold the airport-pair-level rival IVs fixed and aggregate them upward to coarser definitions using the same passenger-weighted averaging used for own characteristics. The two approaches are identical at the airport-pair level by construction.

Table 9 compares these two approaches. The coefficient estimates and first-stage statistics are very similar across them, with only minor differences at the coarser market definitions. This robustness check indicates that the structural-logit results are not sensitive to whether the rival IVs are recomputed after aggregation or instead carried forward from the fine market level.

Table 9 – Structural Logit: BLP Rival IV Comparison

	Airport Pair (1)	Orig.Airport–Dest.City (2)	City Pair (3)
BLP rival (recomputed)	1.02	1.03	0.92
	(0.03)	(0.03)	(0.03)
BLP rival (airport-pair)	1.02	1.03	0.98
	(0.03)	(0.03)	(0.04)
First-stage F (recomputed)	291.7	281.2	273.1
First-stage F (airport-pair)	291.7	251.3	232.9
Market FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Carrier FE	Yes	Yes	Yes
N	863,209	794,987	730,308

Note: "Recomputed" recomputes leave-own-out rival IVs at each candidate market level after aggregation. "Airport-pair" carries the airport-pair-level rival IVs forward to coarser definitions by passenger-weighted averaging. Both approaches recompute shares at the candidate market level.

B.10 Interpretation

The included controls are intended to capture observed service quality and routing characteristics, while the excluded instruments provide variation intended to shift the endogenous price-income ratio. Because the application reports results across several instrument sets and across both direct and structural specifications, the appendix lists all definitions explicitly even though the main text emphasizes only the patterns most relevant for the market-definition test.

C Bank Markets Application

Market definition has long been a contested issue in empirical banking research. Traditional geographic markets such as counties, metropolitan statistical areas (MSAs), or states are often adopted for institutional or regulatory convenience, yet consumer substitution patterns in deposit markets may operate over narrower or broader geographic areas. Moreover, banks typically operate multi-branch networks, and deposit pricing decisions are often set at the bank or regional level rather than at the individual branch level.

This section applies the Papke–Wooldridge (PW) test developed above to branch-level banking data. The goal is to assess whether commonly used geographic market definitions provide a sufficiently rich fixed-effect structure to satisfy the identifying assumptions of a branch-level demand model with endogenous bank-level pricing.

We will consider three sets of PW tests: tracts versus counties, tracts versus metropolitan statistical areas (MSAs), and tracts versus commuting zones (CZs). Our submarket will be census tracts. Both MSAs and CZs nest counties; however, MSAs and CZs are not coterminous.

C.1 Branch Demand Model

The underlying demand model assumes that depositors choose among bank branches. The empirical question is which geographic market best captures that choice set.

Branch-level demand is modeled using a log-quantity specification motivated by the aggregate logit framework developed in Section 3.2. Let b index branches, $j(b)$ index the parent bank of branch b, $m(b)$ index the geographic market in which branch b is located, and t index calendar years.

The structural logit function is:

$$\log(s_{bjmt}) - \log(s_{0,mt}) = \alpha\, p_{bjmt} + \beta X_{bjmt} + \varepsilon_{bjmt}, \tag{32}$$

where s_{bjmt} is the branch market share and $s_{0,mt}$ is the outside good share, p_{bjmt} is a deposit rate (paid to the depositor), X_{bjmt} is a vector of branch characteristics, and ε_{bjmt} is an unobserved branch-level demand shock. The branch market shares are measured based on the volume of deposits at a given branch.

C.2 Data

To estimate the model, I use data spanning 2015 to 2025. For branch data, I use the Summary of Deposits, which is annual regulatory data on the location and deposits of all bank branches, plus limited branch characteristics. For bank data, I use Call Reports, which is quarterly regulatory data on bank financial data. I augment the banking data with quarterly averages of the 1-Year Treasury rates, which we use to measure funding shocks.

C.3 Branch Variables

We do not observe branch level pricing, so we instead assume that deposit rates are set at the bank level. We discuss this variable more below.

The branch characteristics include branch age, an indicator for main-office branches, and an indicator for full-service branches. These variables capture persistent differences in branch size, functionality, and visibility that are plausibly relevant for deposit demand.

I use bank fixed effects to absorb time-invariant bank attributes such as brand reputation, long-run service quality, and persistent differences in product menus.

Bank-Level Price Measurement Branch-level deposit rates are not observed. We construct bank-level average deposit rates from Call Report data and assign the rate to all branches of the bank in a given year. Recent work supports that banks appear to set rates nationally or regionally rather than locally (Granja and Paixao, 2023; Begenau and Stafford, 2023).

Specifically, let q index calendar quarters. The baseline deposit rate measure is defined as an annualized flow-stock ratio over the four quarters ending in Q2 of year t:

$$p_{jt} = \frac{\sum_{q=\mathrm{Q3}(t-1)}^{\mathrm{Q2}(t)} \mathrm{DepositInterestExpense}_{jq}}{\frac{1}{4}\sum_{q=\mathrm{Q3}(t-1)}^{\mathrm{Q2}(t)} \mathrm{Deposits}_{jq}}. \tag{33}$$

This measure captures the average interest rate paid on deposits over the period most relevant for the Q2 deposit stock. Any residual within-bank heterogeneity in branch-level pricing is absorbed into the error term.

C.4 Endogeneity and Instrumental Variables

Bank-level deposit rates are potentially endogenous if unobserved changes in deposit demand are correlated with pricing decisions. To address this concern, equation (32) is estimated by two-stage least squares (2SLS) using instruments constructed from bank funding structure and national interest rate movements.

Our instrument is a funding shock exposure. This instrument type follows a shift-share design that exploits heterogeneity in banks' liability composition combined with common national rate shocks.

Let Δr_t denote the change in a national short-term interest rate between Q2 of year $t-1$ and Q2 of year t. Let exposure measures be defined using lagged balance sheet shares measured in Q2 of year $t-1$: (1) wholesale funding share, (2) brokered deposit share, and (3) large time deposit share.

The baseline instruments are

$$Z_{bt}^{(k)} = \text{Exposure}_{b,t-1}^{(k)} \times \Delta r_t, \qquad k = 1, 2, \ldots \tag{34}$$

where exposures vary across banks and Δr_t is common to all banks.

These instruments shift banks' marginal cost of funds and therefore affect deposit pricing, while their interaction structure ensures that identification comes from cross-bank differences rather than aggregate time variation.

C.5 Empirical Specification

The PW test is applied at the branch level, with census tracts as the submarket definition. We compare three candidate geographic markets against tracts: counties, commuting zones (CZs), and metropolitan statistical areas (MSAs). For the MSA comparison, area that do not belong to any MSA are assigned to their state, so that all observations are retained and the sample does not shrink relative to the county or CZ comparisons.

The estimated demand equation follows equation (32), with branch-by-year observations. The dependent variable is log branch deposits. The endogenous variable is the bank-level deposit rate constructed in equation (33). Branch controls include an indicator for full-service branches, an indicator for main-office branches, the log of branch age in months, and the log of months since the branch was most recently purchased. Bank fixed effects are included to absorb time-invariant bank-level attributes; this implies that banks operating a single branch are excluded from estimation. Area fixed effects are specified as geography-by-year interactions at the candidate market level, absorbing local time-varying economic conditions within each candidate market.

The deposit rate is instrumented using the funding-shock exposure instruments described in equation (34), constructed from lagged brokered deposit and wholesale funding shares interacted with changes in the national short-term rate.

We present results under two sample definitions. In the first version, each pairwise PW test (tracts vs. counties, tracts vs. CZs, tracts vs. MSAs) is estimated on its own unrestricted sample; that is, we do not require observations to be common across the three comparisons. In the second version, we restrict to a common sample across all three tests.

C.6 Results

Table 10 presents the 2SLS estimates under all seven specifications and Table 11 presents the corresponding PW tests. The two sample versions yield sharply different conclusions, and understanding the source of the discrepancy is central to interpreting the results.

Instrument strength is not a concern in any specification. Cluster-robust first-stage F-statistics range from 66.2 to 72.7 across all seven regressions, well above conventional thresholds, and are stable across both sample definitions.

Consider first the unrestricted sample, in which each pairwise comparison is estimated on its own sample. The PW test rejects all three candidate markets with overwhelming precision ($p < 0.001$ in each case), and the estimated price coefficient under tract-by-year fixed effects, 26.85, is noticeably smaller than the estimates under county, CZ, and MSA fixed effects, which range from 29.09 to 30.50.[6] However, this pattern is an artifact of a

[6]Since deposit rates are measured in decimals, a 100bp increase (0.01) implies roughly a 0.27-0.31 in-

compositional difference across specifications. The tract specification contains 717,475 observations, while the county, CZ, and MSA specifications each contain roughly 886,000—a gap of nearly 170,000 branches. The missing observations are branches located in census tracts that contain only a single branch. In the tract specification, these single-branch tracts are perfectly absorbed by the tract-by-year fixed effects and contribute no identifying variation to the price coefficient; in the coarser-geography specifications, they remain in the sample and help identify the estimate. The PW test therefore compares a price estimate identified from multi-branch tracts against estimates identified from a broader population of branches, and the resulting differences reflect this compositional mismatch rather than a failure of the coarser market definitions.

The common-sample results resolve the discrepancy. Restricting all four specifications to the same 717,475 observations eliminates the compositional difference, and the PW test fails to reject the null for all three candidate markets: the p-values are 1.000 for counties, 0.997 for commuting zones, and 0.997 for MSAs. The estimated price coefficient under tract-by-year fixed effects remains 26.85, while the estimates under county, CZ, and MSA fixed effects are 26.08, 29.03, and 30.13, respectively. Although these point estimates still differ, the PW test finds the differences statistically indistinguishable from zero. Conditional on the exogenous variation identified by the funding-shock instruments, tract-level deposit demand heterogeneity does not alter the estimated price coefficient relative to what is obtained under any of the three coarser geographic market definitions.

Tract fixed effects clearly capture meaningful variation in deposit levels, most of it likely reflecting persistent differences in tract-level demographics and branch density. The PW test, however, isolates whether this additional variation changes the identified relationship between deposit rates and deposit demand. It does not, and the failure to reject on the common sample indicates that county, CZ, and MSA fixed effects provide a sufficiently rich structure for the purposes of estimating the price coefficient in this demand model.

crease in log deposits.

Table 10 – Branch-Level Deposit Demand: 2SLS Estimates

		Unrestricted sample			Common sample		
	Tract	County	CZ	MSA	County	CZ	MSA
Deposit rate	26.85	29.09	30.10	30.50	26.08	29.03	30.13
	(6.41)	(5.12)	(4.93)	(4.92)	(5.73)	(5.53)	(5.51)
Other controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Bank FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year × Geo. FE	Tract	County	CZ	MSA	County	CZ	MSA
First-stage F	72.65	66.17	67.36	69.46	66.20	66.36	67.97
Observations	717,475	886,380	887,589	887,649	717,475	717,383	717,383

Note: Dependent variable is log branch deposits (in millions). Deposit rate is instrumented using lagged brokered deposit and wholesale funding shares interacted with the change in the federal funds rate. Bank-clustered standard errors in parentheses. First-stage F-statistics are cluster-robust. The tract specification is identical across sample versions.

Table 11 – Papke–Wooldridge Market Definition Tests

Comparison	$\hat{\alpha}_{\text{tract}}$	$\hat{\alpha}_{\text{coarse}}$	Difference	t-statistic	p-value
Panel A: Unrestricted sample					
Tract vs. County	26.85	29.09	−2.24	−17.60	< 0.001
Tract vs. CZ	26.85	30.10	−3.25	−17.40	< 0.001
Tract vs. MSA	26.85	30.50	−3.65	−17.40	< 0.001
Panel B: Common sample					
Tract vs. County	26.85	26.08	0.77	0.00	1.000
Tract vs. CZ	26.85	29.03	−2.18	0.00	0.997
Tract vs. MSA	26.85	30.13	−3.28	0.00	0.997

Note: The PW test assesses whether tract-by-year fixed effects provide statistically significant improvement over the coarser geography-by-year fixed effects in the instrumented demand equation. $\hat{\alpha}_{\text{tract}}$ and $\hat{\alpha}_{\text{coarse}}$ are the 2SLS estimates of the deposit rate coefficient under tract-by-year and coarser-geography-by-year fixed effects, respectively. The null hypothesis is that the two coefficients are equal. The unrestricted sample allows each comparison to use its own sample; the common sample restricts all specifications to branches in tracts with more than one branch. p-values are one-sided.