Instrument choice and geographic clustering in the Galician literacy panel

Lluis Eriksson

November 2026

Abstract

An audit of a public replication package for a study of emigration and literacy in Galicia finds that the estimated coefficient and the diagnostics printed under one column use different sets of instruments. We reconstruct the historical panel, compare instrument sets on identical observations, and test the hypothesis of zero effect with two geographic definitions of a cluster. Across 47 judicial districts, a restricted wild bootstrap gives $p = 0.1246$ with the instrument used for the coefficient and $p = 0.0282$ with both instruments used for the diagnostics. The positive point estimates survive deletion of each municipality in turn. The result identifies which inference can accompany the archived estimate.

1 Which specification does the table report?

Fernández Sánchez studies the educational consequences of Galician emigration [1]. Column 4 of Table 2 in the deposited package [2] relates standardized literacy to the lagged standardized share of absent men. Line 56 of `Code/table_2.do` fits the model using a destination-demand instrument. Line 89 also includes its interaction with a historical migration-network measure. Lines 234–235 place the resulting diagnostics in several columns. The archive prints a coefficient of 0.247, a first-stage statistic of 20.30 and an overidentification p-value of 0.48 for column 4. The author's July 2024 working paper (p. 24) displays the same three entries. The version of record's typeset table was unavailable for comparison.

With one excluded instrument and one endogenous variable, the fitted equation has zero overidentifying restrictions. The printed 0.48 can therefore describe the two-instrument calculation only. This mismatch alone says nothing about the size or significance of the literacy coefficient. We evaluate that question using the actual observations.

2 Fixed-sample comparison

We standardize on the source panel and then keep the complete observations used by the specification: 321 municipalities in 1900, 1910, 1920 and 1930 (1,284 observations). We absorb municipality and year effects, lagged population and lagged literacy. The single-instrument estimate reproduces the deposited coefficient to its printed precision. Adding only the interaction gives a larger estimate. Both remain positive when any one municipality is omitted (Table 1).

Table 1: Instrument sets evaluated on identical observations

	Demand	Demand and interaction
Standardized two-stage coefficient	0.24727	0.29200
Partial first-stage R^2	0.02470	0.03194
Leave-one-municipality-out minimum	0.20434	0.26089
Leave-one-municipality-out maximum	0.29844	0.33588

3 Inference with geographic dependence

For the null of zero structural effect, the Anderson–Rubin construction [3] tests whether the excluded instruments have zero coefficients in the outcome regression conditional on the same controls. We use a joint Wald statistic with uncorrected cluster-robust covariance (CR0). For each of two geographic groupings we compare its asymptotic chi-square reference distribution with a restricted wild cluster bootstrap. We fit the null, multiply its residuals by one independent random sign per cluster, add the fitted null outcome, refit the alternative and recompute the studentized statistic. The bootstrap p-value is $(1 + h)/10{,}000$, where h is the number of exceedances in 9,999 draws. The code checks its cluster-level calculations against full observation-level fits on the observed data and the first five draws of each specification.

The replication panel's district identifier assigns each municipality to exactly one of 47 judicial districts across all four censuses. Districts contain 3–12 sampled municipalities. Assigning a common sign at district level allows dependence across municipalities within each district.

Table 2: Tests of zero effect. Intervals quantify Monte Carlo sampling error only.

Cluster	Instruments	Asymptotic p	Bootstrap p	95% MC interval
Municipality	Demand	0.0677	0.0720	[0.0669, 0.0771]
Municipality	Both	0.0145	0.0167	[0.0142, 0.0193]
Judicial district	Demand	0.0892	0.1246	[0.1181, 0.1311]
Judicial district	Both	0.0096	0.0282	[0.0250, 0.0315]

At a 5% threshold, only the specification with the added interaction rejects zero under either grouping. Neither grouping establishes exogeneity or accounts for every possible common shock at migration destinations. These calculations concern the four-census literacy equation. Their scope does not extend to long-run outcomes or to other columns of the study.

Data and reproduction. Obtain the versioned data at [2]. The accompanying scripts `reproduce.py` and `reproduce-districts.py` verify the data archive's SHA-256, reconstruct the fixed sample and generate machine-readable results. The bootstrap seed is 20261120.

Model assessments and review context. AI assistance was used in preparing this audit. Any subsequent assessment belongs to the exact PDF identified in its separate review record and should be read with its stated authorship context.

References. [1] M. Fernández Sánchez, *Mass Emigration and Human Capital over a Century: Evidence from the Galician Diaspora*, JEEA, doi:10.1093/jeea/jvag074; author's working-paper page.

[2] M. Fernández Sánchez, versioned replication deposit, doi:10.5281/zenodo.20114638.

[3] T. W. Anderson and H. Rubin (1949), *Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations*, doi:10.1214/aoms/1177730090.