US Centre Summer Research Grant

Recipient name: Ali Glisson

Project title: Pulling apart and sticking together: A matched-pair analysis of regional development in the United States, 1970-2019

Summary of project:

This study uses a matched-pairs design to investigate the long-term economic separation of initially similar, above-average US regions from 1970 to 2019. This ap- proach creates plausible counterfactuals to distinguish the initial endowments that set regions on different paths from the ongoing dynamics that compounded the gap. The findings reveal a predictable outcome but an unpredictable timing: a pair's ultimate separation was strongly predicted by 1970 endowments—particularly higher university graduate shares and population density—while no single initial advantage determined the specific pace of this process. The widening income gap was driven by enduring advantages in innovation (patenting) and human capital. These findings suggest that policy interventions to alter development pathways must account for both the initial conditions that set long-term potential and the dynamic, innovation- led processes that realize it over time.

1 Introduction

This study is motivated by a striking puzzle: many US metropolitan regions that were compa- rable economic peers in 1970 have followed vastly different development paths. For instance, Baltimore and Miami, Denver and Sacramento, and even San Jose and San Diego had similar economic profiles in 1970, yet instead of following the parallel trajectories predicted by neoclassical theory, they drifted apart. This drift occurred during a transformation of the American economy, as the long era of regional convergence gave way to widening spatial inequality, driven by globalization and skills-biased technological change (Blanchard et al., 1992; Fritz & Manduca, 2021; Ganong & Shoag, 2017; Kemeny & Storper, 2020).

The American urban system is complex with a range of nuanced growth pathways between polarized outcomes (Kemeny & Storper, 2023). This study focuses on a cohort of regions that occupy a theoretically interesting terrain, existing between the dominant mega-regions and small, rural areas. This cohort is both consequential to the national economy with their sizable economies, with sufficient variation in outcomes (Beeson et al., 2001; Black & Henderson, 2003).

The literature suggests several key asymmetries that could explain these different development paths, the effects of which may manifest in distinct ways within this cohort. A large body of evidence emphasizes human capital accumulation and cycles of sorting and agglomeration as a primary driver (Berry & Glaeser, 2005; Gennaioli et al., 2013; Giannone, 2021; Moretti, 2004). This fuels a region's adaptive capacity, where a skilled and innovative workforce proves critical for navigating structural shocks and improving regional resilience (Gagliardi et al., 2023; Kemeny et al., 2022; Turner & Weil, 2025). This capacity, in turn, influences a region's industrial composition, particularly its resilience to deindustrialization (Autor et al., 2013a; Autor et al., 2013b) and its ability to specialize in higher-value sectors (Autor & Dorn, 2013). Other contextual factors also play a crucial role, including a region's external connections through global investment links (Bathelt & Buchholz, 2019; Buchholz et al., 2020) and demographic trends that shape labor market outcomes and inequality (Austin et al., 2018; Hanson & Moretti, 2025).

To isolate the effects of these plausible asymmetries, this study employs a quasi-experimental matched-pairs design. By comparing the trajectories of economic peers, the analysis can distinguish the initial endowments that set regions on divergent paths from the ongoing dynamics that compounded the gap over five decades. This paper centers its analysis on income per capita, the surface-level expression of a region's evolving economy: its specialization in high- or low-wage industries, its rate of innovation, and the sophistication of its labor market. (Kemeny & Storper, 2012).

Thus, the separation of these pairs of matched regions, if, when, and why they do so, has significant implications for both theories of regional development and policymaking. Of course, the economic trajectories of these regions are further intertwined with social and political consequences (Connor et al., 2023; Rodr´ıguez-Pose, 2018; Rodr´ıguez-Pose et al., 2021).

This report begins with a discussion of the matching strategy. Then, the empirical analysis proceeds in four parts. First, Kaplan-Meier survival curves establish the general timeline of separation. Second, a logit model tests whether the ultimate outcome of separation is predictable from initial asymmetries. Third, accelerated failure time models examine if these same asymmetries influenced the speed of separation. Finally, fixed-effects models identify the economic

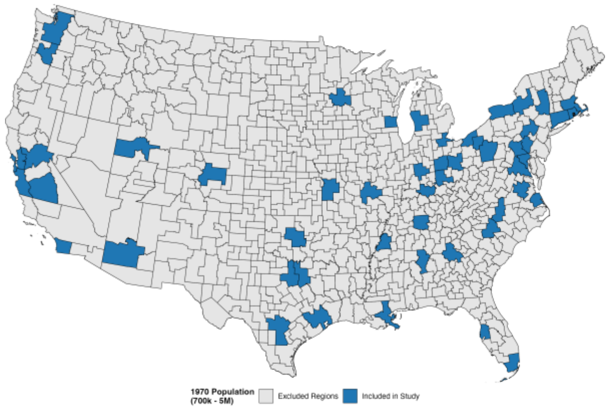


Figure 1: Regions included in analysis

factors that continuously widened the income gap between pairs over the fifty-year period.

2 Matching method and description of the pairs

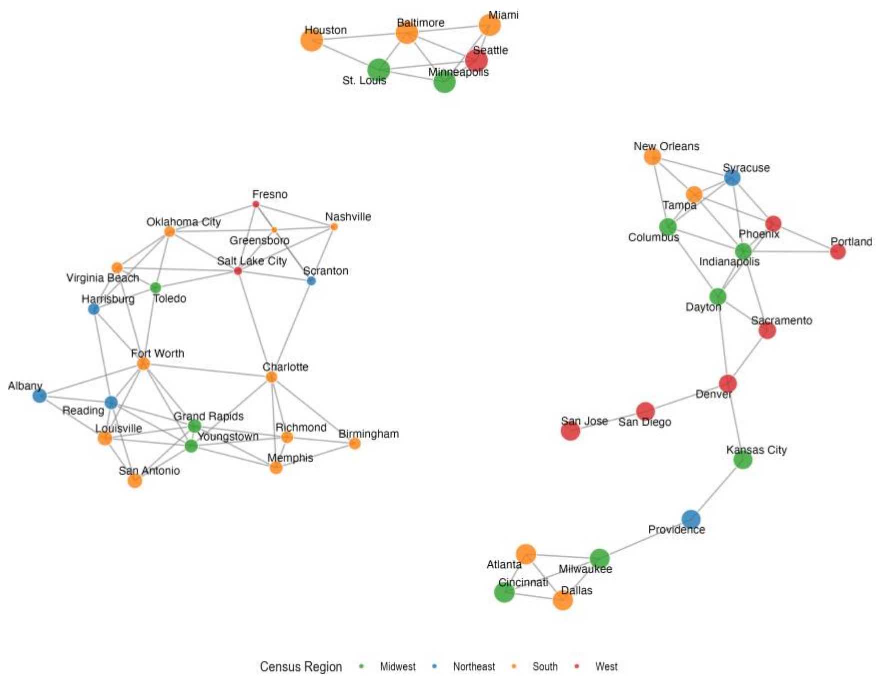
To analyze the dynamics of regional separation, this study uses 1990 commuting zones (CZs) as the unit of analysis, following Autor and Dorn (2013). The sample is strategically focused on a cohort of 51 established, above-average regions with a 1970 population between 700,000 and 5,000,000. [1]Figure 1 maps the 51 regions within this group that comprise this study group.

Each pair is required to have 1970 population difference of less than 10% and an income per capita difference of less than 10%. This is a reasonable, albeit arbitrary, threshold that implies a degree of similarity. All data is accessed via the Bureau of Economic Analysis, United States Patent and Trademark Office, and Census Bureau. Annual population data is accessed via Surveillance, Epidemiology, and End Results (SEER) Program Populations, National Cancer Institute, based on Census Bureau statistics.

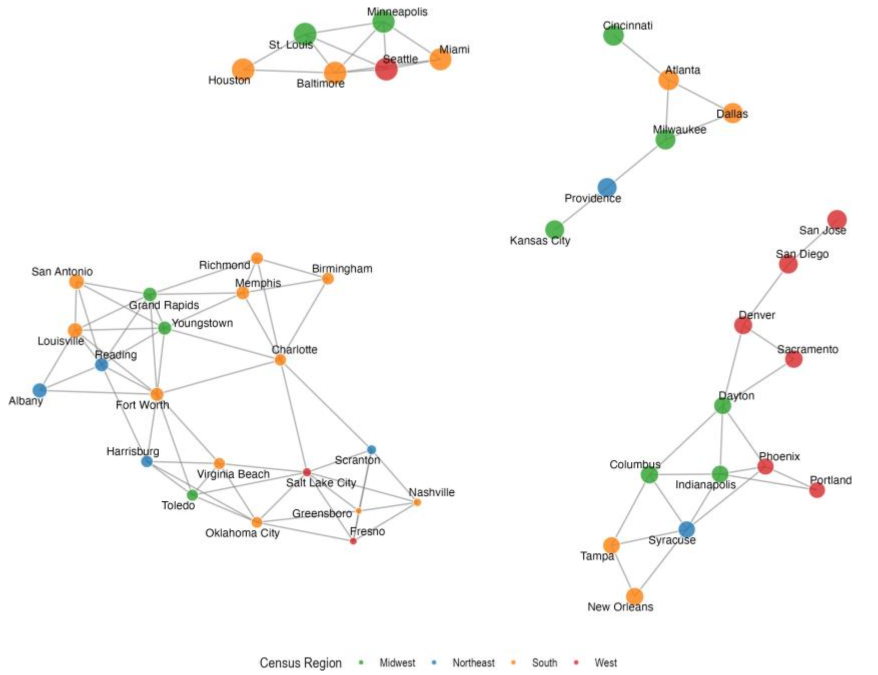
This procedure resulted in a primary set (labeled as the "primary set (10% rule)")of 99 unique pairs and a hybrid set of 91 pairs for robustness checks (read more in Appendix A). Figure 2 visualizes these matched pairs as networks.

Additional diagnostics are performed to provide confidence in the matching approach and for subsequent analysis. As detailed in the appendices, balance checks confirm that the matched pairs began in 1970 with nearly identical characteristics (Appendix B). Furthermore, a series of robustness and placebo tests demonstrates that the final set of pairs is highly stable and reflects genuine economic similarities (Appendix C).

[1]All data is accessed via the Bureau of Economic Analysis, United States Patent and Trademark Office, and Census Bureau. Annual population data is accessed via Surveillance, Epidemiology, and End Results (SEER) Program Populations, National Cancer Institute, based on Census Bureau statistics.



(a) Primary



(b) Hybrid

Figure 2: Network of matched regional pairs. Nodes are sized by 1970 population (log scale) and colored by Census Region. An edge connects two regions that form a pair, illustrating the geographic and economic landscape of the the pairs.

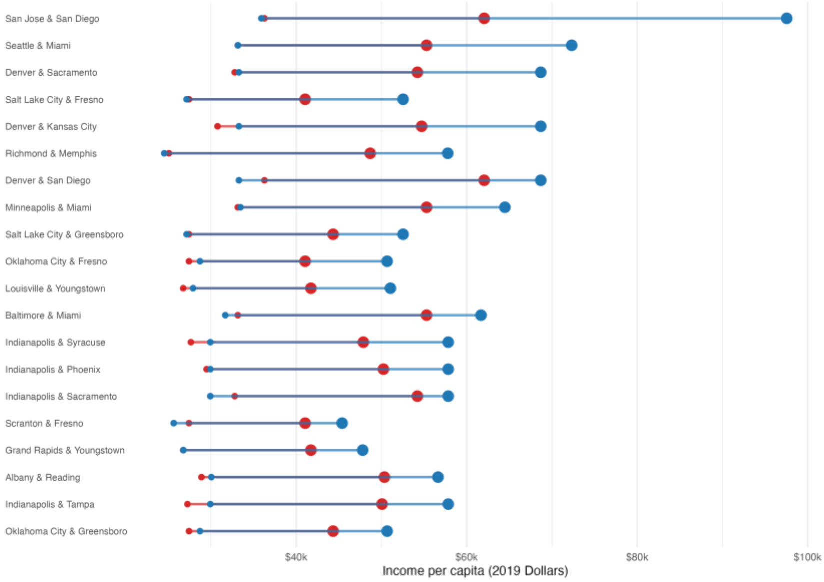


Figure 3: Real Income Trajectories of the 20 Most Separated Pairs, 1970–2019. Each line visualizes the growing gap between initially similar regions. For each pair, blue represents the winner (the region with higher cumulative real income growth over the fifty years), and red represents the loser.

3 Probability and timing of regional pair separation

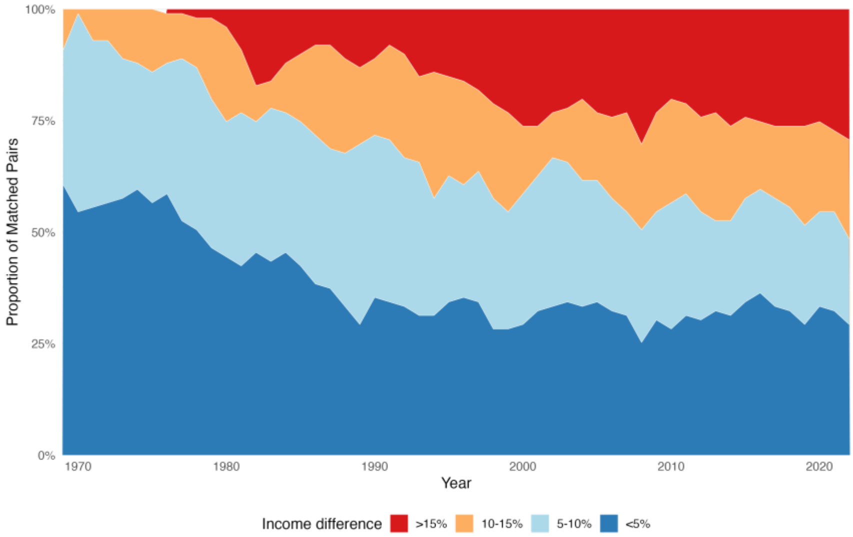
Having established a robust set of comparable regional pairs, the analysis turns to investigate the dynamics of each pair's separation over the subsequent five decades. The descriptive survival analysis is estimated in Appendix D.

Figure 4 illustrates the continuous process of divergence, showing how the initial band of similarity from 1970 eroded over subsequent decades as a growing proportion of pairs drifted into higher-difference categories. Complementing this, the Kaplan-Meier curves (Figure 5) capture the timing of these discrete separation events.

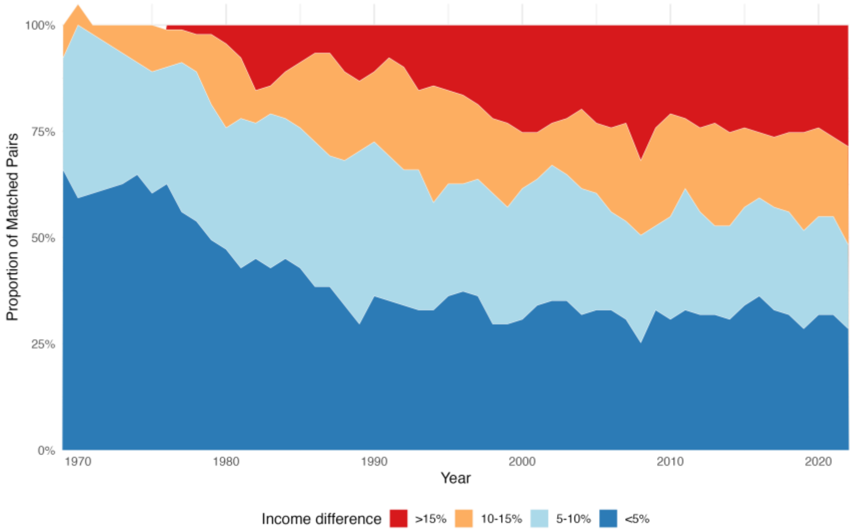
The curve shows a steady process of separation: roughly four-fifths of pairs eventually separate, with income per capita differences exceeding 10%, while about 20% of regions remain together during all five decades. The median time to separation is approximately 20 years, occurring around 1990.

While the Kaplan-Meier curves establish when separation occurred, they do not explain why. To test whether this outcome was predictable, the analysis employs a logit model. First, pairs are oriented using a post-hoc designation based on which region had higher long-run income growth. The model then tests if the initial 1970 advantages of the region that ultimately outpaced its peer—for example, a higher university graduate share or greater population density—increased the likelihood of eventual separation. The full model specification is in Appendix E, and the results are presented in Table 1.

The results show that divergence was not random but predicted by a specific set of 1970 endowments. A winner's initial advantage was most clearly associated with higher population density and a more educated workforce, underscoring the role of agglomeration and human capital. The institutional environment also mattered: an advantage for winners was associated with being in non-Right-to-Work states, suggesting a distinct model of industrial relations



(a) Primary (10% Rule)



(b) Hybrid (10% Rule + Statistical Distance)

Figure 4: Evolution of income gaps between matched pairs. *Source:* Author's analysis of BEA and Census data.

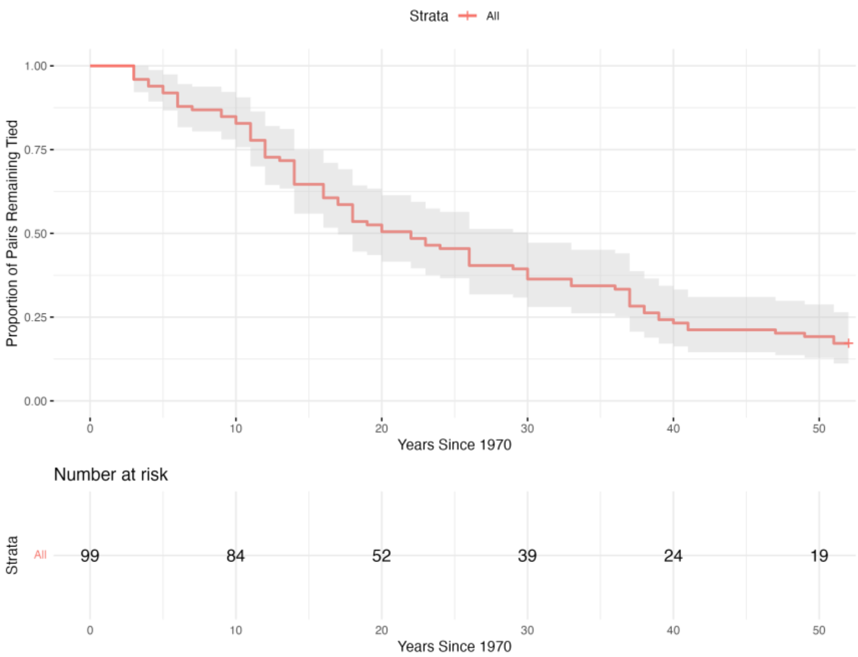
Table 1: Logit Estimates of Initial Endowments on the Probability of Pair Separation

Covariate (advantage)	**Estimate**	**Std. Error**	**p-value**
Intercept	-1.068	0.343	0.002**
Education (UniGradRatio)	1.115	0.433	0.010*
Manufacturing employment	-0.016	0.422	0.970
Poverty rate (lower is advant.)	-0.772	0.502	0.124
Patents per capita	-0.015	0.311	0.962
Nonwhite share	0.339	0.337	0.314
Right-to-work status	-1.223	0.418	0.003**
Population density	1.129	0.380	0.003**

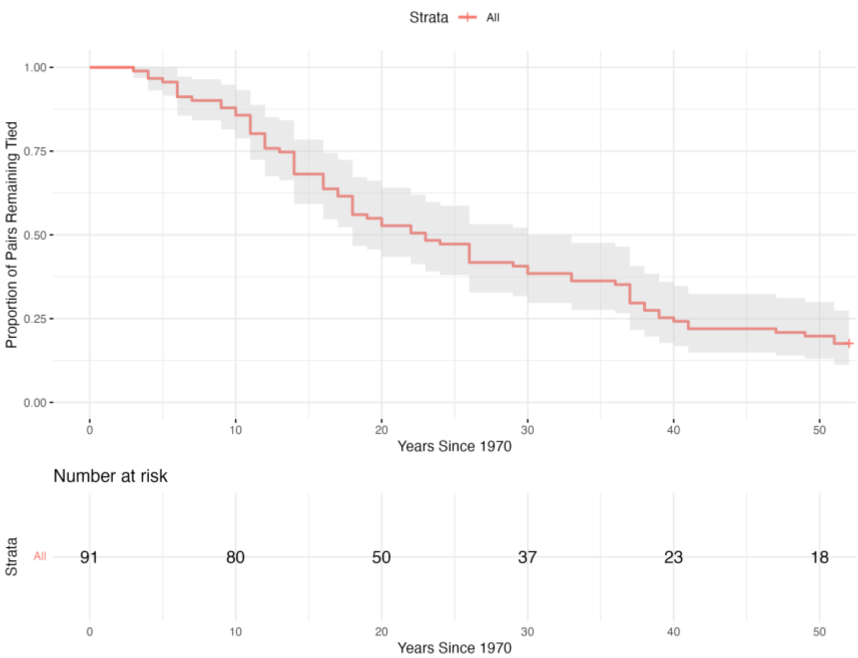
Notes: Estimates are log-odds coefficients from a logit model.
Observations = 99; Log-Likelihood = -37.9; Adj. Pseudo R^2 = 0.294; BIC = 112.6.
Significance: * $p < 0.05$, ** $p < 0.01$.

contributed to their capacity for growth (Austin & Lilley, 2021; VanHeuvelen, 2023).

Despite pairs starting with a moderate difference in manufacturing share (see Appendix B),



(a) Primary (10% Rule)

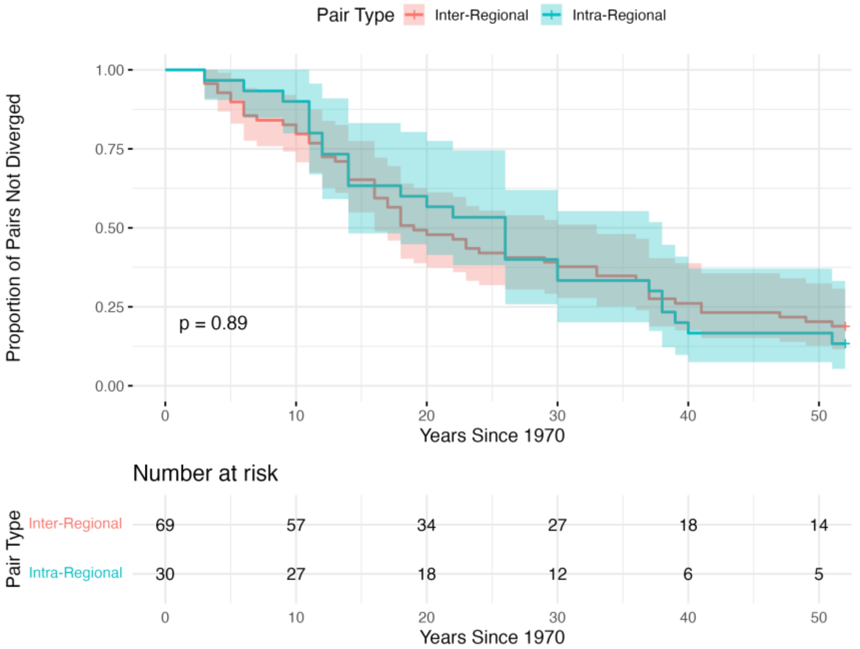


(b) Hybrid (10% Rule + Statistical Distance)

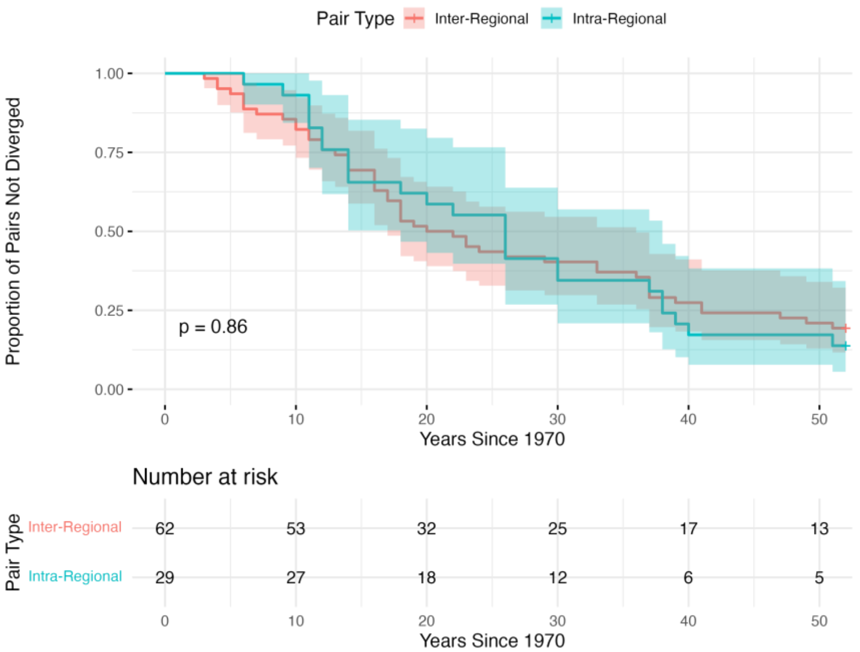
Figure 5: Overall Kaplan–Meier survival curves showing the probability that a matched pair has not yet separated (i.e., income per capita gap remains within 10%) since 1970. Results are shown separately for the primary (10% rule) and hybrid (10% rule + statistical distance) matched sets.

this initial discrepancy had no bearing on whether a pair would eventually separate ($p = 0.970$).

The logit model reveals that a pair's likelihood of separating was predictable. This raises a second question: did these same initial advantages also determine the pace at which that separation unfolded? For instance, since an initial advantage in university graduates made separation more likely, it stands to reason that a larger educational advantage might also lead to quicker separation. An Accelerated Failure Time (AFT) model tests whether the magnitude of



(a) Primary (10% Rule)



(b) Hybrid (10% Rule + Statistical Distance)

Figure 6: Overall Kaplan–Meier survival curves showing the probability that a matched pair has not yet separated (i.e., income per capita gap remains within 10%) since 1970. Results are shown separately for the primary (10% rule) and hybrid (10% rule + statistical distance) matched sets.

these same 1970 advantages systematically accelerated or delayed the timing of that separation (see Appendix G).

None of the initial 1970 advantages are statistically significant predictors of the timing of separation. Some point estimates are directionally suggestive (for example, a higher non-white share (TR = 0.83) hints at faster separation), but the 95% confidence interval for every variable comfortably includes 1.0. This outcome reinforces the conclusion from stratified survival analysis

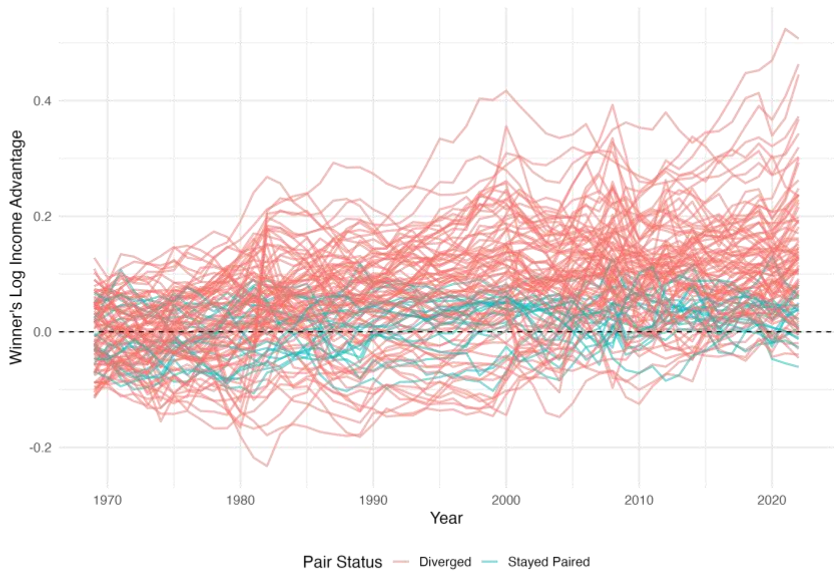


Figure 7: Evolution of the log income gap between matched "winner" and "loser" regions, 1970–2019. Each line represents one matched pair, oriented such that the region with higher long-term growth is designated the winner. The vertical axis shows the winner's log income advantage, i.e., the natural log of the ratio of winner to loser per-capita income. A value of zero indicates income parity; positive values reflect the growing proportional advantage of winners relative to their matched peers.

and log-rank tests in Appendix F.

In summary, while the conditions in 1970 determined if a pair would likely separate, no single endowment significantly predicted when or how quickly that separation would occur.

Table 2: Lognormal Accelerated Failure Time (AFT) Models: Per-1 SD Time Ratios (TR) with 95% CIs

Covariate (advantage)	Primary TR	SE	95% CI	Hybrid TR
Education (UniGradRatio)	0.94	0.16	[0.68, 1.28]	0.86
Manufacturing employment	1.02	0.22	[0.66, 1.57]	0.97
Finance employment	1.03	0.17	[0.74, 1.44]	1.02
Poverty rate (lower advant.)	1.22	0.20	[0.83, 1.79]	1.27
Patents per capita	1.05	0.11	[0.84, 1.31]	1.08
Employment-to-population	1.00	0.30	[0.56, 1.78]	1.54
Prime-age employment (EmpWA)	0.99	0.26	[0.59, 1.66]	0.69
Nonwhite share	0.83	0.17	[0.60, 1.14]	0.79

Notes: Entries are per-1 SD time ratios (TR = e^{β}). TR<1 = faster separation; TR>1 = slower separation. SE = standard error of β, transformed via delta method. 95% CI = $e^{\beta \pm 1.96 \cdot SE}$.
Models: Lognormal AFT; samples are persistent separation (K=3) with 10% threshold.

4 Comparing economic profiles of regional pairs in 2019

The preceding analysis establishes that a pair's separation was a predictable outcome, albeit with unpredictable timing. But, what were the tangible consequences (besides one region out- pacing its peer in income per capita growth)? This section examines the end-states in 2019, comparing the economic profiles of winner, loser and stay paired groups of pairs. Table 3) compares the summary statistics for each group at the beginning and end of the study period.

Table 3: Summary Statistics by Separation, 1970 vs 2019

	1970			2019		
	Winners	Losers	Stayed Paired	Winners	Losers	Stayed Paired
Panel A: Means						
Income per Capita	4,134.91	4,124.47	4,114.43	56,231.83	51,948.07	53,301.09
Population	1,171,152	1,176,126	1,121,392	2,491,997	2,299,565	2,228,663
Population Density	100.53	108.01	107.33	188.94	193.47	192.52
Employment-Pop Ratio	0.67	0.66	0.66	0.80	0.77	0.79
Manufacturing Emp.	0.34	0.35	0.34	0.09	0.10	0.10
University Grads	0.11	0.10	0.10	0.34	0.31	0.33
Patents per 10k	3.48	2.81	3.49	12.48	7.20	9.60
Non-White Share	0.12	0.12	0.12	0.24	0.22	0.23
Panel B: Counts of CZs by RTW Status						
Non-RTW States	19	20	12	14	16	7
RTW States	14	13	8	19	17	13
Panel C: Counts of CZs by Census Region						
Midwest	5	10	5	5	10	5
Northeast	5	4	3	5	4	3
South	16	14	10	16	14	10
West	7	5	2	7	5	2

While the groups began in 1970 with similar economic profiles, a clear hierarchy had formed by 2019. The most dramatic difference appeared in patents per capita, reflecting evidence that innovation clusters exhibit much stronger spatial concentration than manufacturing activities and serve as the primary mechanism for sustained regional advantage (Carlino & Kerr, 2015).

This success was also reflected in higher shares of university graduates and stronger employment-to-population ratios, cementing the profile of a winner as a prosperous, highly skilled, and innovative urban economy. Ultimately, the separation that began from a state of a high degree of similarity in 1970 culminated in a landscape of difference fifty years later.

5 Mechanisms driving separation

Table 4: Drivers of Regional Income Gaps: Alternative Lag Specifications

Variable	1-Year Lag	3-Year Lag	5-Year Lag
Manufacturing Emp. Gap	−0.082* (0.036)	−0.043 (0.036)	−0.035 (0.046)
University Grad. Gap	0.722*** (0.179)	0.740*** (0.179)	0.756*** (0.190)
Patents Gap	0.0035*** (0.0007)	0.0034*** (0.0009)	0.0034** (0.0011)
Nonwhite Pop. Gap	−0.279† (0.155)	−0.430** (0.156)	−0.533** (0.165)
Employment–Pop. Gap	0.819*** (0.061)	0.586*** (0.054)	0.364*** (0.057)
Population Density Gap	−0.00011 (0.00009)	−0.00008 (0.00007)	−0.00009 (0.00007)
RTW Status Gap	0.0065 (0.0072)	0.0101 (0.0079)	0.0135 (0.0091)
Observations	4,851	4,851	4,851
R^2	0.783	0.734	0.697
Within R^2	0.416	0.282	0.184

Notes. Each column reports coefficients from a fixed-effects OLS model with pair and year effects. Standard errors (in parentheses) clustered by pair. Significance codes: $***p < 0.001$, $**p < 0.01$, $*p < 0.05$, $^{\dagger}p < 0.10$.

The final empirical question is how this income gap emerged and continued to widen year after year. This section moves beyond static initial conditions and outcomes to analyze the process of separation utilizing a fixed-effects model leveraging fifty years of panel data. The full model specification is in Appendix I.

Across all lag specifications, enduring and growing advantages in education and patenting activity consistently predict a wider income gap (Table 4). The effect of employment-to-population ratios is strong in the short run but attenuates over time, whereas the influence of non-white population share grows stronger with longer lags. Conversely, the manufacturing gap shows a statistically significant negative effect in the short term (-0.082 at a 1-year lag), confirming that a heavy specialization in manufacturing acted as a drag on relative income growth, even as the effect fades in later years.

These patterns suggest that some drivers of separation (human capital, innovation) are persistent, while others (labor market participation, industrial mix) matter primarily in the near term.

The robustness of these findings was confirmed through a jackknife analysis and the use of two-way clustered standard errors, which demonstrate that the core results are not driven by influential outliers or sensitive to alternative error structures (see Appendices J and K).

Finally, the analysis tested whether these separation mechanisms are context-dependent. The results show that the core drivers of advantages in human capital and innovation predict a widening income gap, with similar strength for both intra-regional (within the same Census region) and inter-regional (across different Census regions) pairs. A formal interaction model confirms this null finding, suggesting the forces driving separation operate are not fundamentally altered by local proximity or shared regional institutions (Appendix L).

6 Implications

This paper set out to understand why initially economically similar, established but not dominant, US regions followed dramatically different development paths from 1970 onwards. By employing a matched-pairs design, the analysis reveals a nuanced story of path dependency, dynamic separation, and the tangible consequences in the modern economy.

A region's destiny to diverge from its peer was possible to predict decades in advance from its 1970 endowments in human capital and density, yet the timing of this separation was not. This divergence was not a single event but a continuous process, fueled by the persistent accumulation of knowledge-based assets that interact with the existing path. The fixed-effects models show that enduring advantages in innovation and education actively widen the income gap over time, a mechanism that appears to operate irrespective of geography or historical industrial legacy.

These findings have several implications for both economic theory and public policy. First, they suggest that models of divergence must account for both initial conditions that set long-term potential and the dynamic processes that realize it over time.

For policymakers, the predictability of the outcome from 1970 conditions suggests that policy interventions to alter a region's trajectory must be early and sustained over the long-term. The findings strongly support a focus on building human capital and innovation ecosystems. The fact that manufacturing decline was a universal feature suggests that industrial policy should focus on transition and adaptation, not preservation.

References

Austin, B., Glaeser, E., & Summers, L. (2018). *Jobs for the Heartland: Place-Based Policies in 21st Century America* (tech. rep. No. w24548). National Bureau of Economic Research. Cambridge, MA. https://doi.org/10.3386/w24548

Austin, B., & Lilley, M. (2021). The long-run effects of right to work laws. *Mathew Lilley Ph. D Candidate in Business Economics*, *11*, 13.

Autor, D., & Dorn, D. (2013). The Growth of Low-Skill Service Jobs and the Polarization of the US Labor Market. *American Economic Review*, *103* (5), 1553–1597. https://doi.org/10.1257/aer.103.5.1553

Autor, D., Dorn, D., & Hanson, G. (2013a). The China Syndrome: Local Labor Market Effects of Import Competition in the United States. *American Economic Review*, *103* (6), 2121–2168. https://doi.org/10.1257/aer.103.6.2121

Autor, D. H., Dorn, D., & Hanson, G. H. (2013b). The Geography of Trade and Technology Shocks in the United States. *American Economic Review*, *103* (3), 220–225. https://doi.org/10.1257/aer.103.3.220

Bathelt, H., & Buchholz, M. (2019). Outward Foreign Direct Investments as a Catalyst of Urban-Regional Income Development? Evidence from the United States. *Economic Geography*, *95* (5), 442–466. https://doi.org/10.1080/00130095.2019.1665465

Beeson, P. E., DeJong, D. N., & Troesken, W. (2001). Population growth in U.S. counties, 1840–1990. *Regional Science and Urban Economics*, *31* (6), 669–699. https://doi.org/10.1016/S0166-0462(01)00065-5

Berry, C., & Glaeser, E. (2005). The divergence of human capital levels across cities*. *Papers in Regional Science*, *84* (3), 407–444. https://doi.org/10.1111/j.1435-5957.2005.00047.x

Black, D., & Henderson, V. (2003). Urban evolution in the USA. *Journal of Economic Geography*, *3* (4), 343–372. https://doi.org/10.1093/jeg/lbg017

Blanchard, O., Katz, L., Hall, R., & Eichengreen, B. (1992). Regional Evolutions. *Brookings Papers on Economic Activity*, *1992* (1), 1. https://doi.org/10.2307/2534556

Buchholz, M., Bathelt, H., & Cantwell, J. A. (2020). Income divergence and global connectivity of U.S. urban regions. *Journal of International Business Policy*, *3* (3), 229–248. https://doi.org/10.1057/s42214-020-00057-7

Carlino, G., & Kerr, W. R. (2015). Agglomeration and Innovation. In *Handbook of Regional and Urban Economics* (pp. 349–404, Vol. 5). Elsevier. https://doi.org/10.1016/B978-0-444-59517-1.00006-4

Connor, D. S., Berg, A. K., Kemeny, T., & Kedron, P. J. (2023). Who gets left behind by left behind places? *Cambridge Journal of Regions, Economy and Society*, rsad031. https://doi.org/10.1093/cjres/rsad031

Fritz, B., & Manduca, R. (2021). The economic complexity of US metropolitan areas. *Regional Studies*, *55* (7), 1299–1310. https://doi.org/10.1080/00343404.2021.1884215

Gagliardi, L., Moretti, E., & Serafinelli, M. (2023, December). *The World's Rust Belts: The Heterogeneous Effects of Deindustrialization on 1,993 Cities in Six Countries* (tech. rep. No. w31948). National Bureau of Economic Research. Cambridge, MA. https://doi.org/10.3386/w31948

Ganong, P., & Shoag, D. (2017). Why has regional income convergence in the U.S. declined? *Journal of Urban Economics*, *102*, 76–90. https://doi.org/10.1016/j.jue.2017.07.002

Gennaioli, N., La Porta, R., Lopez-de-Silanes, F., & Shleifer, A. (2013). Human Capital and Regional Development*. *The Quarterly Journal of Economics*, *128* (1), 105–164. https://doi.org/10.1093/qje/qjs050

Giannone, E. (2021). Skill-biased Technological Change and Regional Convergence. *Unpublished*. https://crei.cat/wp-content/uploads/2022/01/JMP March2021.pdf

Hanson, G., & Moretti, E. (2025, March). *Where Have All the Good Jobs Gone? Changes in the Geography of Work in the US, 1980-2021* (tech. rep. No. w33631). National Bureau of Economic Research. Cambridge, MA. https://doi.org/10.3386/w33631

Kemeny, T., & Storper, M. (2012). The Sources of Urban Development: Wages, Housing and Amenity Gaps across American Cities. *Journal of Regional Science*, *52* (1), 85–108.

Kemeny, T., & Storper, M. (2020). Superstar Cities and Left-Behind Places: Disruptive In- novation, Labor Demand, and Interregional Inequality [Publisher: Unpublished]. *LSE International Inequality Institute*, (Working Paper 41). https://doi.org/10.13140/RG.2. 2.19192.19202

Kemeny, T., Petralia, S., & Storper, M. (2022). Disruptive innovation and spatial inequality. *Regional Studies*, 1–18. https://doi.org/10.1080/00343404.2022.2076824

Kemeny, T., & Storper, M. (2023). The Changing Shape of Spatial Income Disparities in the United States. *Economic Geography*, 1–30. https://doi.org/10.1080/00130095.2023. 2244111

Moretti, E. (2004). Chapter 51 Human capital externalities in cities. In *Handbook of Regional and Urban Economics* (pp. 2243–2291, Vol. 4). Elsevier. https://doi.org/10.1016/S1574-0080(04)80008-7

Rodr´ıguez-Pose, A. (2018). The revenge of the places that don't matter (and what to do about it). *Cambridge Journal of Regions, Economy and Society*, *11* (1), 189–209. https://doi.org/10.1093/cjres/rsx024

Rodr´ıguez-Pose, A., Lee, N., & Lipp, C. (2021). Golfing with Trump. Social capital, decline, inequality, and the rise of populism in the US. *Cambridge Journal of Regions, Economy and Society*, *14* (3), 457–481. https://doi.org/10.1093/cjres/rsab026

Stuart, E. A. (2010). Matching Methods for Causal Inference: A Review and a Look Forward. *Statistical Science*, *25* (1). https://doi.org/10.1214/09-STS313

Turner, M., & Weil, D. (2025, January). *Are Big Cities Important for Economic Growth?* (Tech. rep. No. w33334). National Bureau of Economic Research. Cambridge, MA. https://doi.org/10.3386/w33334

VanHeuvelen, T. (2023). The Right to Work and American Inequality. *American Sociological Review*, *88* (5), 810–843. https://doi.org/10.1177/00031224231197630

A Hybrid matching for robustness

I employ a secondary approach to ensure the selection of the most genuinely comparable pairs, incorporating a measure of statistical distance to avoid selecting statistically poor pairs that happen to fall within the threshold. The hybrid matching procedure first calculates a weighted Mahalanobis distance for all possible pairs, ranking them based on a measure of similarity. The "hybrid set" consists of pairs that are in the top 33% of statistical similarity (i.e., have the lowest Mahalanobis distance) and also meet the 10% hard threshold rule.

B Covariate Balance of Matched Pairs

To validate the quality of the matched pairs, a balance table (Table 5) is presented, which reports the difference in means for 1970 characteristics. The primary diagnostic is the Abso- lute Standardized Difference (ASD)—a scale-free measure of between-group difference that is independent of sample size (Stuart, 2010). An ASD below 0.25 is typically taken to indicate negligible imbalance, providing confidence in the matching procedure (Stuart, 2010).

The Paired vs. Unpaired comparison reveals a key feature of the data: the seven regions that remain unpaired are systematically larger, wealthier, and more innovative (e.g., ASD for population is 2.367). These regions lack a suitable comparison within the study's criteria.

Second, the Within-Pair comparison confirms that the matching created pairs with nearly identical foundational characteristics. For the explicit matching variables of population and income, the ASD values are exceptionally low (0.035 and 0.079, respectively). Furthermore, the pairs are well-balanced across most demographic and economic dimensions, including educational attainment, age structure, and poverty. However, the matching process did not fully resolve pre-existing differences in industrial structure; a moderate imbalance persists for Manufacturing Employment (ASD = 0.424) and Finance Employment (ASD = 0.388).

Third, the nearly identical balance statistics for the primary and hybrid sets demonstrate the robustness of the matching procedure itself. Rather than one method being superior, this consistency shows that the core set of comparable pairs and their characteristics are not sensitive to the specific matching algorithm used (i.e., a simple caliper vs. a statistically-refined hybrid approach). This gives us high confidence that the selected group of 99 primary pairs is a stable and representative sample for the main analysis that follows.

Table 5: Covariate Balance Diagnostics Across Matching Specifications

	Within-Pair Balance						Paired vs. Unpaired Balance		
	Primary (10% Rule)			Hybrid (10% Rule + Stat. Dist.)			Primary (10% Rule)		
Variable	Mean A	Mean B	ASD	Mean A	Mean B	ASD	Mean Paired	Mean Unpaired	ASD
Population	1158258	1186774	0.035	1158258	1186774	0.035	1194148	3149084	2.367
Income per capita	4127	4169	0.079	4127	4169	0.079	4162	4931	1.462
University Graduates (%)	0.106	0.110	0.150	0.106	0.110	0.150	0.108	0.138	0.991
Manufacturing Emp. (%)	0.372	0.323	0.424	0.372	0.323	0.424	0.345	0.364	0.168
Finance Emp. (%)	0.061	0.068	0.388	0.061	0.068	0.388	0.064	0.069	0.311
Over 65 Pop. (%)	0.091	0.093	0.091	0.091	0.093	0.091	0.094	0.094	0.025
Poverty Rate (%)	0.116	0.123	0.180	0.116	0.123	0.180	0.120	0.091	0.781
Population Density	106	96	0.112	106	96	0.112	107	265	1.648
Total Patent Count	490	403	0.106	490	403	0.106	432	2360	2.344
Patents per capita	0.000	0.000	0.212	0.000	0.000	0.212	0.000	0.001	1.291
Employment–Population (%)	0.661	0.670	0.175	0.661	0.670	0.175	0.662	0.664	0.040
Prime Working Age Emp. (%)	0.761	0.773	0.253	0.761	0.773	0.253	0.765	0.766	0.026
Nonwhite Pop. (%)	0.116	0.124	0.091	0.116	0.124	0.091	0.120	0.115	0.059
Regional Price Parity (2019)	101	102	0.190	101	102	0.190	101.547	105.208	0.588

Notes. ASD stands for Absolute Standardized Difference. For the Within-Pair panels, Mean A and Mean B represent the averages for the two arbitrarily assigned sides of each matched pair. For the Paired vs. Unpaired panel, "Mean Paired" is the average for all unique regions included in at least one pair, while "Mean Unpaired" is the average for regions in the initial sample that could not be matched. An ASD value below 0.25 is generally considered to indicate good covariate balance.

C Robustness of the matching strategy

The validity of this paper's findings depends on the quality and stability of the matched pairs. To ensure the results are not an artifact of specific methodological choices, four tests confirm that the identified pairs are robust and reflect genuine economic similarities, strengthening the confidence in the analysis.

First, I tested whether our choice of a weighted Mahalanobis distance metric was driving the results by re-running the match using standardized Euclidean and standardized Manhattan distances. The choice of metric had a negligible impact. Both alternative metrics produced matched sets with a Jaccard similarity of 0.92 compared to our primary set. This high degree of overlap shows that the pairs reflect true similarities in the data, not the specific formula used to measure them.

Second, the primary model weights population 2.5 times more than income (2.5:1) to prioritize similarity in scale. I tested this against alternative schemes, from equal weighting (1:1) to a heavy population emphasis (4:1). The matching was highly stable across reasonable weighting schemes (Jaccard similarity > 0.92). Only the most extreme 4:1 weighting showed a meaningful deviation. This confirms our 2.5:1 weight is a balanced, non-arbitrary choice that falls within a stable parameter range.

Third, the 10% caliper for initial differences in population and income ensures intuitive similarity. I also considered a stricter (5%) and a looser (15%) caliper. The number of pairs scaled predictably with the caliper's strictness, from 27 at 5% to 132 at 15%. The choice of 10% yields 99 pairs, representing a conservative balance between match quality and sufficient sample size for robust analysis.

To ensure our algorithm captures a real economic relationship, I also conducted a placebo test, where 1970 income per capita data was randomly shuffled across regions—breaking the link between population and income. I then reran the match, and as one would expect, the match quality collapsed. The Jaccard similarity between the original pairs and the placebo pairs fell to just 0.24. This confirms the procedure identifies genuine, non-random economic relationships, not spurious correlations driven by population.

D Kaplan-Meier estimator specification

To assess how long initially matched pairs remain on similar growth paths, I estimate Kaplan–Meier survival functions. This is descriptive evidence only.

The Kaplan-Meier estimator, denoted as $\hat{S}(t)$, is a non-parametric statistic used to estimate the survival function from time-to-event data, which takes the following form:

$$\hat{S}(t) = \prod_{i:\, t_i \leq t} \left(1 - \frac{d_i}{n_i}\right) \tag{1}$$

where $\hat{S}(t)$ is the estimated probability that a matched pair has remained paired (i.e., has not separated) by year t; t_i denotes the years (since 1970) in which at least one separation event occurred; d_i is the number of pairs that diverged in year t_i; and n_i is the number of pairs still paired ("at risk") just before year t_i.

E Logit model specification

The analysis next examines whether the separation of a pair was a predictable, path-dependent process determined by initial conditions. To do this, a logit model is employed to test if the subtle 1970 advantages of the region that would ultimately become the long-term “winner”predict the probability of a persistent economic separation. A positive finding would provide evidence that these divergent trajectories were not random but were shaped by asymmetries present at the outset. The logit takes the following form:

$$\Pr(Y_i = 1 \mid Z_i) = \Lambda\left(\alpha + Z_i^{\top}\beta\right), \qquad \Lambda(u) = \frac{1}{1 + e^{-u}},$$

where Y_i is an indicator equal to one if pair i eventually separates, and Z_i is a vector of standardized 1970 advantages, including university graduate ratio, manufacturing employment ratio, poverty ratio, patents per capita, non-white population ratio, RTW status, and population density. Coefficients β_j capture the effect of a one–standard deviation increase in the baseline advantage on the log-odds of separation. Odds ratios are given by $\exp(\beta_j)$.

F Stratified survival analysis and log-rank test

Table 6: P-values from Stratified Survival Analyses

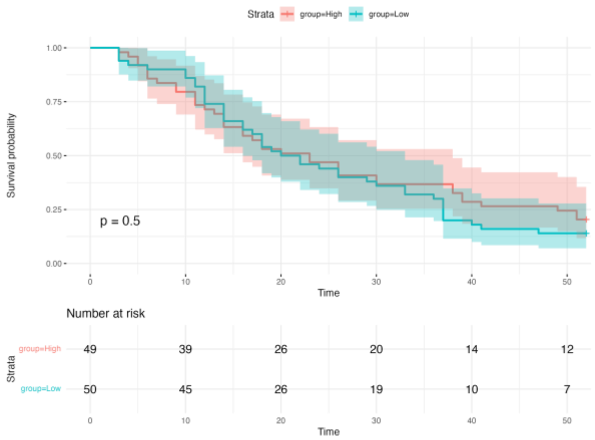
Variable	Primary (10% Rule)	Hybrid (10% + Stat. Dist.)
University Graduates (%)	0.687	0.483
Manufacturing Employment (%)	0.503	0.511
Finance Employment (%)	0.380	0.343
Poverty Rate (%)	0.470	0.369
Patents per Capita	0.870	0.979
Employment–Population (%)	0.682	0.542
Prime-Age Employment (%)	0.630	0.489
Non-White Population (%)	0.777	0.892

Note. P-values are rounded to three decimal places for readability.

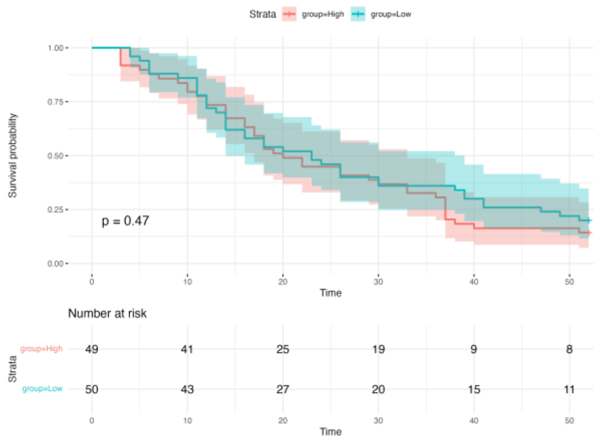
To test if the magnitude of a winner’s initial advantage affected the pace of separation, pairs were stratified into high (above median) and low advantage groups for each 1970 characteristic (relative to the median), and their survival curves were compared.

The Kaplan–Meier curves in Figures 8 and 9 illustrate how long pairs of regions stayed close to one another before one pulled away, and whether an early advantage tipped the balance. Across all the stratified plots is that the confidence intervals for both the high and low groups overlap extensively, indicating that the differences between the curves are likely not statistically meaningful.

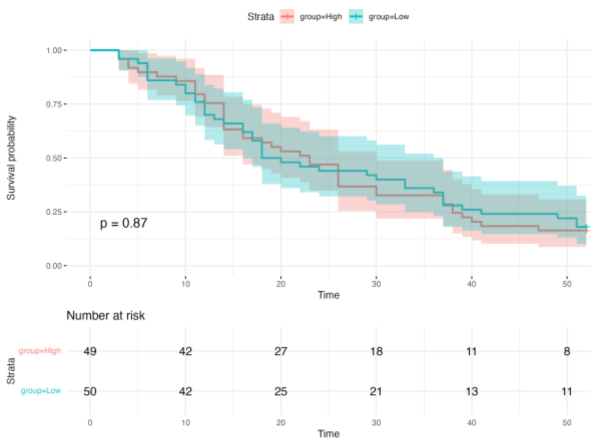
The formal log-rank tests confirm this interpretation (Table 6). The p-values for all tested variables are well above the conventional 0.05 significance threshold, indicating that there are no statistically significant differences between the survival curves for the ”High” versus ”Low” advantage groups. Any visual separation between the curves in Figures 8 and 9 is consistent with random sampling variation and cannot be attributed to a true, systematic effect of that initial endowment.



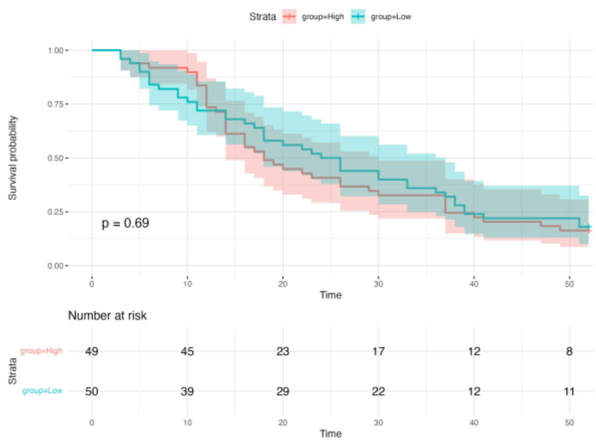
(a) Manufacturing advantage



(b) Poverty rate advantage



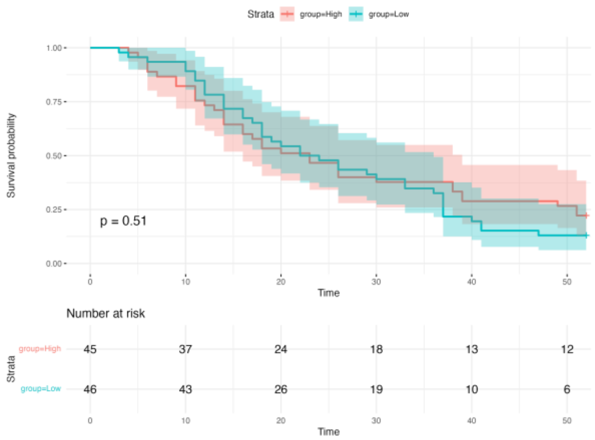
(a) Patents per capita advantage



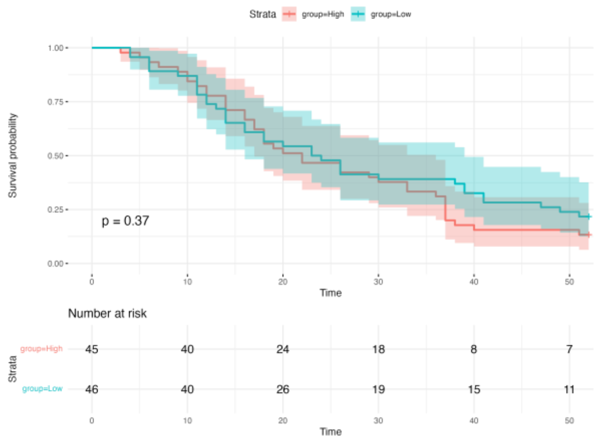
(d) University graduate ratio

Figure 8: Kaplan–Meier survival analyses for selected covariate advantages (primary set, 10% rule). Each panel shows the probability that a matched pair remains within 10% income similarity over time, comparing pairs where the eventual winner began with a clear 1970 advantage (turquoise) against pairs without such an advantage (coral).

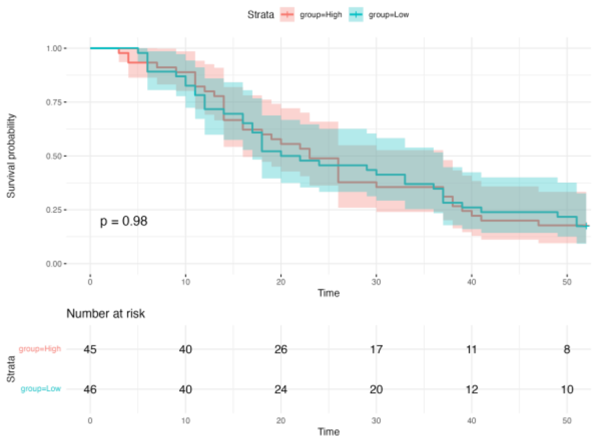
This null finding on the timing of separation contrasts the logit model's results, which showed that initial endowments could predict the ultimate probability of whether a pair would separate. Thus, while a region's destiny to separate from its pair may be influenced by its initial conditions, the specific pace at which that separation unfolds is not strongly determined by any single one of these factors.



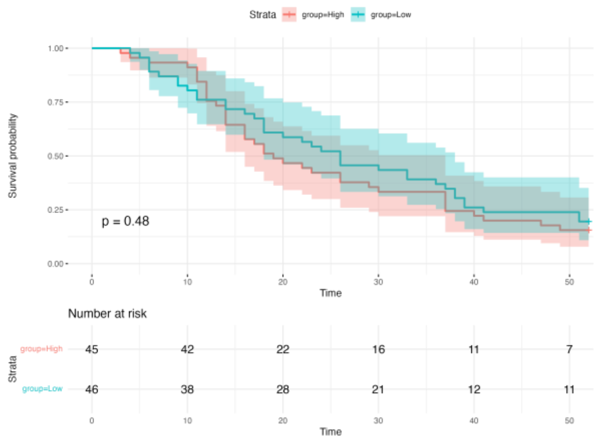
(a) Manufacturing advantage



(b) Poverty rate advantage



(c) Patents per capita advantage



(d) University graduate ratio

Figure 9: Kaplan–Meier survival analyses for selected covariate advantages (hybrid set, 10% rule). Each panel shows the probability that a matched pair remains within 10% income simi- larity over time, comparing pairs where the eventual winner began with a clear 1970 advantage (turquoise) against pairs without such an advantage (coral).

G Accelerated Failure Time (AFT) model specification

I employ an Accelerated Failure Time (AFT) model, which directly estimates how a given characteristic stretches or compresses the time until a pair diverges. The model takes the following form:

$$\log T_i = \mathbf{X}_{i,1970}^{\top}\boldsymbol{\beta} + \sigma\varepsilon_i,$$

where T_i is the time to separation for pair i and $\mathbf{X}_{i,1970}$ is a vector of the winner's standardized 1970 advantages. The results are interpreted as *Time Ratios* (TR), defined as

$$\mathrm{TR} = e^{\beta}.$$

A TR < 1 indicates that an advantage accelerates separation (makes it happen faster), while a TR > 1 indicates that it decelerates separation (makes it happen slower). A TR = 1 implies no effect.

H AFT model selection via AIC

Table 7: AIC Comparison of AFT Distributions

Distribution	Primary AIC	Hybrid AIC
Weibull	734.95	674.41
Log-logistic	733.55	670.97
Log-normal	733.46	670.67

Notes: Lower AIC indicates better model fit. Log-normal is preferred for both samples. Models estimated on persistent separation (K=3) with 10% threshold.

I test three common distributional assumptions for the AFT model, each implying a different shape for the underlying hazard function of pair separation. The Weibull distribution assumes a monotonic hazard, where the risk of separation is either strictly increasing or strictly decreasing over the entire period. In contrast, both the log-normal and log-logistic distributions allow for a non-monotonic hazard function.

To select the most appropriate model, I compared the Akaike Information Criterion (AIC) for each specification. Across both the primary and hybrid matched sets, the log-normal distri- bution provided the best fit, as shown in Table 7.

I Fixed-effects model specification

To investigate the mechanisms driving the separation in income per capita between matched regional pairs, I employ a series of two-way fixed-effects models to understand how one region pulls ahead of its specific economic peer year over year. The model specification is as follows:

$$Y_{it} = \alpha_i + \gamma_t + \sum_{k=1}^{K} \beta_k X_{k,it-L} + \epsilon_{it} \tag{2}$$

Where:

- Y_{it} is the outcome variable, the log-difference in income per capita between the two regions in pair i in year t.
- α_i represents the pair fixed effects, controlling for all time-invariant characteristics of each unique pair i.
- γ_t represents the year fixed effects, controlling for all time-specific shocks common to all pairs in year t.
- $X_{k,it-L}$ is the value of the k^{th} control variable for pair i at time $t - L$, where L is the lag length (e.g., 1, 3, or 5 years). The set of K control variables includes the gaps in manufacturing employment, university graduate ratio, patents per capita, etc.
- β_k are the coefficients of interest, measuring the effect of a one-unit change in each control variable on the income gap.
- ϵ_{it} is the idiosyncratic error term for pair i in year t. Standard errors are clustered at the pair level.

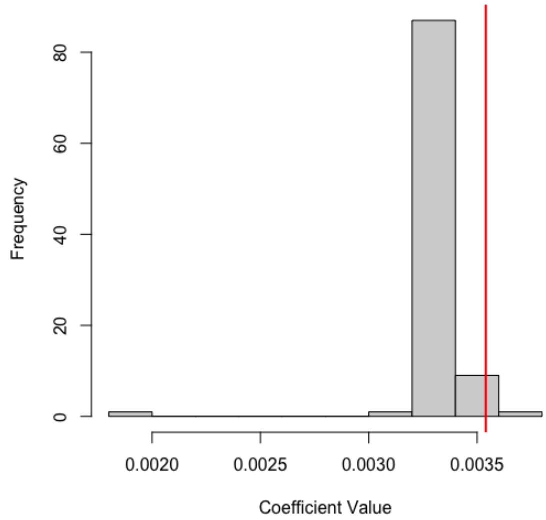


Figure 10: Jackknife histogram of patents per capita

J Robustness check: jackknife analysis

To assess whether the estimated effect of patenting is driven by any single matched pair, I conducted a jackknife test to re-estimate the model repeatedly while omitting one observation (here, one matched pair) at a time, and then examining the distribution of the resulting coefficients. The histogram below displays the leave-one-out estimates for the patents coefficient, with the red line marking the full-sample estimate. The distribution is tightly clustered around the full estimate, and the sign of the coefficient is consistently positive across all replications. While a handful of outliers exert modest influence, there is no evidence that the result depends on any one region.

This sensitivity check demonstrates that the positive effect of patenting on pair separation is robust and not an artifact of a small number of influential cases.

K Robustness check: two-way clustering of standard errors

This test checks whether the results are sensitive to the way standard errors are calculated. When switching from clustering by pair only to two-way clustering by both pair and year, the overall pattern of results is unchanged. Human capital (university graduates), innovation (patents), and broad labor force participation (employment-to-population) remain consistently strong and significant predictors of income gaps. The non-white share remains marginally significant, while population density and manufacturing employment do not show systematic effects. These findings confirm that the main conclusions are robust to alternative error structures, strengthening confidence in the results.

Table 8: Drivers of Regional Income Gaps: One-Year Lag Model with Two-Way Clustered SEs

Variable	Estimate (SE)
Manufacturing Emp. Gap	−0.082 (0.050)
University Grad. Gap	0.722** (0.217)
Patents Gap	0.0035*** (0.0007)
Nonwhite Pop. Gap	−0.279† (0.159)
Employment–Pop. Gap	0.819*** (0.072)
Population Density Gap	−0.00011 (0.00009)
RTW Status Gap	0.0065 (0.0089)
RMSE	0.043
Adj. R^2	0.783
Within R^2	0.416
Observations	4,851

Notes. Coefficients are from an OLS model with pair and year fixed effects. Standard errors (in parentheses) are two-way clustered by pair and year. Significance codes: $***p < 0.001$, $**p < 0.01$, $*p < 0.05$, $^{\dagger}p < 0.10$.

L Subgroup analysis by geographic proximity

The analysis in Table 9 addresses a crucial question: does the geographic scope of a comparison fundamentally change the drivers of separation? We explore this by splitting the sample and running a formal interaction model to see if the recipe for economic separation is consistent across different spatial contexts.

First, we examine intra-regional pairs—those located in the same state or census region. This tests whether the same rules of separation apply to close, similar neighbors. Advantages in university graduates and patenting activity remain strong and reliable predictors of a widening income gap.

Next, looking at inter-regional pairs that span different states and broader regions, the same pattern holds. Once again, education and innovation gaps are the key drivers that predict which region pulls ahead. The consistency of these findings across both local and long-distance comparisons highlights the robustness of these central mechanisms.

A formal interaction model tests whether the strength of these fundamental drivers differs between intra- and inter-regional pairs. The findings show no robust evidence that the effects of human capital or innovation vary by geographic scope, as the interaction terms are not statistically significant. This suggests that the same core mechanisms govern why peer regions grow apart, regardless of whether they are neighbors or located in entirely different parts of the country.

Table 9: Drivers of Regional Income Gaps: Intra- vs. Inter-Regional Pairs and Interac- tion Model

Variable	Intra-Regional Pairs	Inter-Regional Pairs	Interaction Model
Manufacturing Emp. Gap	0.080 (0.072)	0.012 (0.040)	0.075 (0.068)
University Grad. Gap	1.100** (0.326)	1.362*** (0.223)	1.088*** (0.314)
Patents Gap	0.0024* (0.0009)	0.0029* (0.0011)	0.0025** (0.0009)
Manufacturing × Inter	—	—	−0.059 (0.075)
University × Inter	—	—	0.281 (0.374)
Patents × Inter	—	—	0.0003 (0.0014)
Observations	1,470	3,381	4,851
RMSE	0.0523	0.0519	0.0522
Adj. R^2	0.621	0.707	0.686
Within R^2	0.151	0.156	0.154

Notes. All models estimated with OLS and two-way fixed effects (pair and year). Standard errors (in parentheses) clustered by pair. Significance codes: $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\dagger}p < 0.10$. Interaction model includes product terms with the Inter-Regional indicator; the main effect of Inter-Regional is omitted due to collinearity.