Institute for Economic Studies, Keio University

Keio-IES Discussion Paper Series

Marshall meets Bartik: Revisiting the mysteries of the trade

Yasusada Murata, Ryo Nakajima

14 April, 2025
DP2025-005
https://ies.keio.ac.jp/en/publications/25234/

Institute for Economic Studies, Keio University
2-15-45 Mita, Minato-ku, Tokyo 108-8345, Japan
ies-office@adst.keio.ac.jp
14 April, 2025

Marshall meets Bartik: Revisiting the mysteries of the trade

Yasusada Murata, Ryo Nakajima

Keio-IES DP2025-005
14 April, 2025
JEL Classification: R12, O31, J61, C26
Keywords: patent productivity, inventor migration, knowledge spillovers, knowledge sharing, Bartik instruments, mysteries of the trade, idea-generating process

Abstract

We identify a causal effect of top inventor inflows on the patent productivity of local inventors by combining the idea-generating process described by Marshall (1890) with the Bartik (1991) instruments involving the state taxes and commuting zone characteristics of the United States. We find that local productivity gains go beyond organizational boundaries and co-inventor relationships, which implies the partially nonexcludable good nature of knowledge in a spatial economy and pertains to the mysteries of the trade in the air. Our counterfactual experiment suggests that the spatial distribution of inventive activity is substantially distorted by the presence of heterogeneity in state taxes.

Yasusada Murata
College of Economics, Nihon University
1-3-2 Kanda-Misakicho, Chiyoda-ku, Tokyo 101-8360, JAPAN
murata.yasusada@nihon-u.ac.jp

Ryo Nakajima
Department of Economics, Keio University
2-15-45 Mita, Minato Tokyo 108-8345, JAPAN
nakajima@econ.keio.ac.jp

Marshall meets Bartik:

Revisiting the mysteries of the trade*

Yasusada Murata† Ryo Nakajima‡

April 14, 2025

Abstract

We identify a causal effect of top inventor inflows on the patent productivity of local inventors by combining the idea-generating process described by Marshall (1890) with the Bartik (1991) instruments involving the state taxes and commuting zone characteristics of the United States. We find that local productivity gains go beyond organizational boundaries and co-inventor relationships, which implies the partially nonexcludable good nature of knowledge in a spatial economy and pertains to the mysteries of the trade in the air. Our counterfactual experiment suggests that the spatial distribution of inventive activity is substantially distorted by the presence of heterogeneity in state taxes.

Keywords: patent productivity; inventor migration; knowledge spillovers; knowledge sharing; Bartik instruments; mysteries of the trade; idea-generating process

JEL codes: R12; O31; J61; C26

*The earlier version of this paper has been circulated as Murata and Nakajima (2023) and presented at the 17th North American Meeting of the Urban Economics Association, 37th Annual Meeting of the Applied Regional Science Conference, 2024 Spring Meeting of the Japanese Economic Association, and 13th European Meeting of the Urban Economics Association. We thank the discussants and participants at these conferences and other workshops for valuable comments and suggestions.

†College of Economics, Nihon University. E-mail: murata.yasusada@nihon-u.ac.jp

‡Department of Economics, Keio University. E-mail: nakajima@econ.keio.ac.jp

1 Introduction

Knowledge creation has been central to various fields of economics such as trade, growth, and geography. However, little is known about the idea-generating process between individuals, despite Marshall's (1890) simple explanation as follows:

> *if one man starts a new idea, it is taken up by others and combined with suggestions of their own; and thus it becomes the source of further new ideas.*

While intuitive, verifying this statement has been challenging. The main difficulty lies in the possible endogeneity—those who generate new ideas tend to cluster together.

We address this problem by identifying a causal effect of a top inventor inflow on the patent productivity of local inventors at the commuting-zone level in the United States. In doing so, we use inventor-level data from the PatentsView database, which is an open data platform supported by the United States Patent and Trademark Office (USPTO). Since top inventor inflows are likely endogenous, we predict those flows by constructing Bartik (1991) instruments: the predicted probability that a top inventor migrates from origin to destination constitutes a share, and the number of top inventors in the origin corresponds to a shift.

To understand the driving forces behind knowledge creation among individuals, we first classify local inventors into internal and external inventors. Local inventors are considered internal if they share the same organization as the migrating top inventors and/or if they are co-inventors of the migrating top inventors. All other local inventors are external because they are not directly linked to the migrating top inventors.

We then examine two types of effects—the productivity gains of all local inventors and those of external inventors. Our baseline results suggest that the former and latter gains from an additional top inventor inflow are 6% and 4%, respectively. The former are interpreted as local aggregate gains from both knowledge sharing among internal inventors and knowledge spillovers to external inventors. The latter focus on the gains that go beyond organizational boundaries and co-inventor relationships and pertain to the most frequently quoted passage from Marshall (1890): "*The mysteries of the trade become no mysteries; but are as it were in the air.*" We thus disentangle productivity gains due to external knowledge spillovers ("knowledge in the air") from those due to internal knowledge sharing ("knowledge in the lab").

Our identification strategy consists of main three steps. We first estimate the impact of spatial and temporal variation in top earners' income tax rates on the migration probability of top inventors for any pair of origin and destination commuting zones while controlling for origin-destination characteristics. We then aggregate, for each destination commuting zone, the predicted bilateral probabilities across origin commuting zones to construct a Bartik instrument for top inventor inflows. We finally employ an instrumental variable (IV) approach, where we use the Bartik instrument in the first-stage regression and estimate a structural equation, with the outcome being local patent productivity. The identifying assumption is that local patent productivity in a destination commuting zone does not directly depend on top earners' income tax rates in other commuting zones located in *different states*.[1]

Our novelty lies in the construction of the Bartik instrument: The predicted migration probability is derived from a location choice model of top inventors who face spatial and temporal differences in individual income tax rates. Thus, our framework can be used to examine to what extent those tax differences distort the spatial distribution of inventive activity. To illustrate this, we run a counterfactual experiment by setting individual income taxes to their average and find that the existence of tax differences affects local patent productivity up to -64.8% to 72.3%, with considerable spatial heterogeneity. We further decompose those gains and losses into two types—direct gains from tax changes and indirect gains via top inventor migration induced by tax changes. We find that the former share is 0.275, while the latter share is 0.725.

The contribution of our paper is threefold. First, we shed new light on the idea-generating process described in Marshall (1890) using Bartik (1991) instruments. Our framework thus differs from natural experimental approaches to knowledge production in historical contexts (e.g., Borjas and Doran, 2012; Moser et al., 2014) or exploitation of the sudden death of inventors (e.g., Azoulay et al., 2010; Azoulay et al., 2019). We leverage the variation in tax rates across space and time to demonstrate that the tax-induced migration of top inventors leads to

[1]For example, this assumption implies that local patent productivity in destination commuting zone 37500 (Santa Clara–Monterey–Santa Cruz, CA) does not directly depend on top 5% or 1% earners' income tax rates in origin commuting zones 19600 (Bergen–Essex–Middlesex, NJ), 24300 (Cook–DuPage–Lake, IL), and so forth. In line with this assumption, we show in Section 4.5 that the main source of identifying variation comes from interstate top inventor migrations. We elaborate on this assumption in Section 4.3 and Appendix E.1.

local productivity gains in their destination, thereby contributing to the agglomeration and innovation literature (e.g., Carlino and Kerr, 2015; Kerr and Robert-Nicould, 2020).

Second, we disentangle the productivity gains due to external knowledge spillovers from those due to internal knowledge sharing through organizations or co-inventor relationships, which allows us to revisit the mysteries of the trade in the air. The theoretical foundation for separating the nonexcludable part from the excludable part of the gains dates back at least to Griliches (1979) and Romer (1990), whereas the empirical literature typically estimates the productivity gains of migrants themselves or those from internal knowledge sharing (e.g., Moretti, 2021; Prato, 2025). Thus, the productivity gains attributed to external knowledge spillovers among individuals have remained unexplored in a spatial framework using modern causal inference methods. Since this partially nonexcludable good nature of knowledge leads to market failures and constitutes a rationale for spatial agglomeration of inventive activity, our analysis contributes to the innovation policy literature (e.g., Chatterji et al. 2014; Aghion and Jaravel, 2015).

Finally, we derive the Bartik instruments from a location choice model involving policy variables à la Moretti and Wilson (2017). By construction, our model-based Bartik instruments can be used in any setting where origin-destination flows are affected by changes in location-specific policies. This paper applies these instruments to tax-induced domestic migration and conducts a counterfactual experiment to illustrate a way of bridging the gap between the tax and innovation literature (e.g., Stantcheva, 2021; Akcigit et al., 2022; Akcigit and Stantcheva, 2022) and the tax and migration literature (e.g., Kleven et al., 2020). This application provides new insights into these two strands of literature since it allows us to assess the relative importance of direct productivity gains from tax changes and indirect productivity gains through the tax-induced migration of top inventors.

The remainder of the paper is organized as follows. In Section 2, we explain the data and show descriptive statistics. In Section 3, we analyze how tax differences affect the migration of top inventors. Section 4 constructs the Bartik instruments and presents our main results on local patent productivity gains by employing the instrumental variable approach. In Section 5, we check the robustness of the main results. Section 6 discusses the underlying mechanisms

through which local productivity gains materialize. We conduct the counterfactual experiment in Section 7 and conclude the paper in Section 8.

2 Data and descriptive statistics

Our main dataset is the PatentsView database, which is an open data platform supported by the USPTO and provides various administrative data on issued patents and patent applications. The data are based on the disambiguation process and contain, for each issued patent, patent inventors, assignees, residential addresses of patent inventors, and patent citations. Our sample period is from 1977 to 2009, during which there were $3,015,305$ patent applications by 1,282,708 unique inventors (see Appendix A for a more detailed description of the data sources and construction, as well as the disambiguation of inventors and assignees).[2]

Since our objective is to estimate the impact of top inventor migration on the productivity of local inventors in the destination, we need to define the productivity of an inventor and determine (i) who qualifies as a top inventor, (ii) under what condition we detect the migration of a top inventor, and (iii) who in the destination potentially gains from top inventor inflows.

To this end, we first define the productivity of an inventor as the number of patents applied for by that inventor.[3] We then identify, for each year, the top 5% of inventors based on productivity over the last ten years and refer to them as *top inventors* for short. It follows that the status of a top inventor varies from year to year. During our sample period, there are $263,259$ top inventor $\times$ year observations, and the number of unique top inventors is $60,294$. Thus, on average, the total duration of being a top inventor is 4.366 years.

We detect the migration of a top inventor if the commuting zone of residence of that top inventor in the patent application data differs between two consecutive years.[4] Since the status

[2]The sample period and data construction are dictated by data availability and consistent with those in Moretti and Wilson (2017).

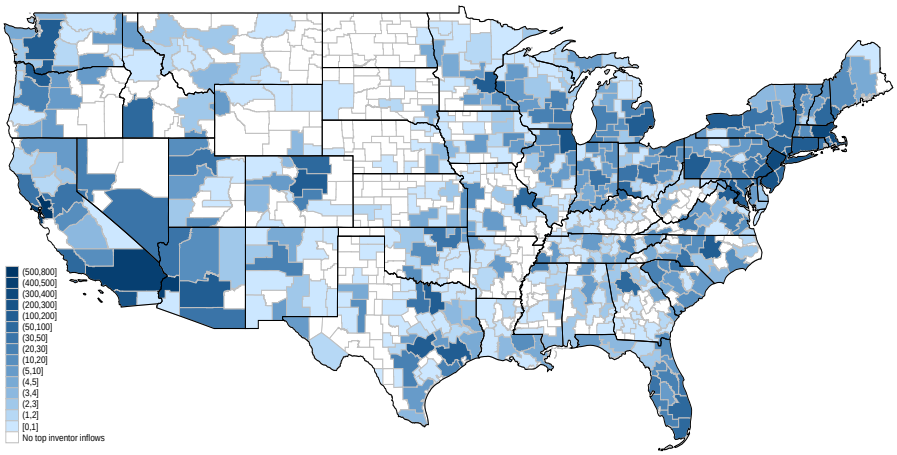
[3]If there are multiple inventors for a patent, we allocate an equal fraction of that patent to each of its inventors; i.e., if there are three inventors for a patent, one-third of that patent is allocated to each inventor.

[4]We assume that the migration occurs at the end of the first year and use the definition of commuting zones as of 1990. When an inventor applied for more than one patent in a year, the most frequently observed commuting zone is regarded as that inventor's place of residence in that year. In case of a tie, we use the commuting zone observed for the first time in that year. We exclude commuting zones in Alaska and Hawaii. This detection requires the observations of the same inventor for two consecutive years, which may lead to the underestimation of top inventor migrations and measurement error. Thus, the same caveats as those in Moretti and Wilson (2017, pp.1864–1865) apply to our data construction.

of a top inventor varies from year to year, we consider the migration of inventors who qualify as top inventors in the first year.[5] In our sample, the total number of top inventor migrations is 9,178, and the number of unique top inventors who migrated at least once is 5,725. Thus, on average, each top inventor moved 1.603 times, conditional on moving at least once.

Since we analyze the impact of top inventor migration on the productivity of local inventors in the destination, we aggregate migration flows at the destination level. Figure 1 depicts the geographic distribution of all 9,178 top inventor inflows by commuting zone, and Table 1 summarizes top 10 commuting zones by top inventor inflows.[6]

Figure 1: Geographic distribution of top inventor inflows.



Notes: Inflows are defined as the number of top inventors who migrated into each commuting zone from 1977 to 2009.

Table 1: Top 10 commuting zones by top inventor inflows.

rank	cz number	counties	state	inflows
1	37500	Santa Clara–Monterey–Santa Cruz	CA	724
2	37800	Alameda–Contra Costa–San Francisco	CA	557
3	38300	Los Angeles–Orange–San Bernardino	CA	408
4	19600	Bergen–Essex–Middlesex	NJ	372
5	20500	Middlesex–Worcester–Essex	MA	335
6	38000	San Diego	CA	266
7	19400	Kings–Queens–New York	NY	240
8	19700	Philadelphia–Montgomery–Delaware	PA	220
9	24300	Cook-DuPage–Lake	IL	219
10	20901	Hartford–Fairfield–New Haven	CT	195

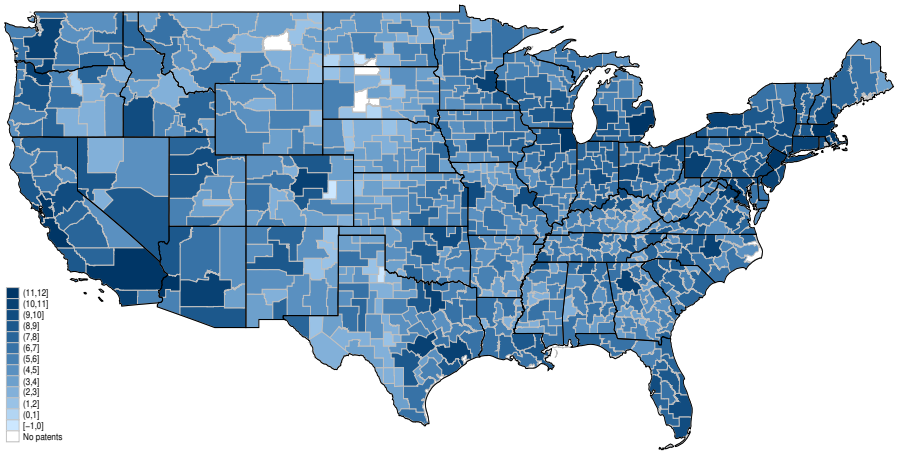
Notes: Inflows are defined as the number of top inventors who migrated into each commuting zone from 1977 to 2009.

[5] As a robustness check, we consider the migration of inventors who qualify as top inventors in both years in Appendix C.1.

[6] We aggregate migration flows at the commuting zone level because it captures stronger commuting ties and thus more inventor interactions within labor market areas and because knowledge spillovers tend to be localized at short distances (see, e.g., Murata et al. 2014). We check the robustness of the result regarding geographic space in Section 5.2.

When assessing the impact of top inventor migration into a commuting zone, we focus on the *local inventors* who lived in that commuting zone at that time while excluding the top inventors who had already moved in that commuting zone. In our sample, the number of those local inventors is $1,274,192$, and they applied for $2,027,777$ patents from 1977 to 2009. Figure 2 illustrates the geographic distribution of local patent productivity (in logs), and Table 2 summarizes the top 10 commuting zones by local patent productivity.

Figure 2: Geographic distribution of local patent productivity (in logs).



Notes: The productivity in each commuting zone is defined as the total number of patents applied for by local inventors between 1977 and 2009. If there are multiple inventors for a patent, then we allocate an equal fraction of that patent to each of its inventors.

Table 2: Top 10 commuting zones by local patent productivity.

rank	cz number	counties	state	productivity
1	37500	Santa Clara-Monterey-Santa Cruz	CA	143,069.321
2	38300	Los Angeles-Orange-San Bernardino	CA	116,303.629
3	37800	Alameda-Contra Costa-San Francisco	CA	81,828.603
4	20500	Middlesex-Worcester-Essex	MA	80,486.767
5	19600	Bergen-Essex-Middlesex	NJ	77,093.518
6	24300	Cook-DuPage-Lake	IL	75,095.386
7	11600	Wayne-Oakland-Macomb	MI	63,437.324
8	19400	Kings-Queens-New York	NY	54,570.527
9	21501	Hennepin-Ramsey-Dakota	MN	50,587.528
10	39400	King-Pierce-Snohomish	WA	49,760.915

Notes: The productivity in each commuting zone is defined as the total number of patents applied for by local inventors between 1977 and 2009. If there are multiple inventors for a patent, then we allocate an equal fraction of that patent to each of its inventors.

As seen from Figures 1 and 2, their spatial patterns are quite similar. The correlation between the top inventor inflows and local patent productivity (the log of local patent productivity) is 0.97 (0.52) and their rank correlation is 0.85. However, since correlation does not necessarily imply causation, we take an instrumental variable approach to examine the

causal effect of the top inventor inflows on the productivity of local inventors. To this end, we use the variation in the individual income average tax rates for top earners (ATRs) by state and year to construct predicted flows of top inventors by commuting zone and year.[7] We also consider corporate income tax rates (CITRs), investment tax credits (ITCs), and R&D tax credits (RTCs) that can affect top inventor migration.[8] Table 3 presents summary statistics for the top inventor inflows and the local patent productivity at the commuting zone $\times$ year level. We show in Appendix A the summary statistics for other commuting zone-level variables, as well as state taxes and tax credits, that we use in the subsequent analysis.

Table 3: Summary statistics (main variables).

	total	mean	sd	min	max
Local patent productivity (overall)	2,027,776.570	85.821	386.089	0.000	10,205.625
Local patent productivity (internal)	1,061,199.866	44.913	252.589	0.000	8,232.745
Local patent productivity (external)	966,576.704	40.908	152.466	0.000	3,108.842
Top inventor inflows (overall)	9,178.000	0.388	2.139	0.000	81.000
Top inventor inflows (intrastate)	2,271.000	0.096	0.929	0.000	37.000
Top inventor inflows (interstate)	6,907.000	0.292	1.453	0.000	48.000
Number of observations					23,628
Number of commuting zones					716
Number of years					33

Notes: Summary statistics are based on the data described in Section 2 for the years 1977 to 2009. The local patent productivity can be decomposed into two: One is by the internal inventors who share the same assignee as the migrating top inventors and/or who are co-inventors of the migrating top inventors; and the other is by the external inventors. The top inventor inflows can be decomposed into intrastate and interstate migration. Of the 722 commuting zones, four have no patents and two have only one patent during the sample period. We thus use 716 commuting zones in our regression analysis with fixed effects.

We further classify local inventors into *internal* and *external* inventors. Local inventors are internal if they share the same assignee as the migrating top inventors and/or if they are co-inventors of the migrating top inventors. All the other local inventors are external because they are not directly linked to the migrating top inventors. In our sample, 42.20% of local inventors are internal, whereas the remaining 57.80% are external.[9]

[7]We assume that top inventors are taxpayers at the ninety-fifth (ninety-ninth) percentile of the U.S. income distribution as a baseline (as a robustness check). In Appendix C.1, we further check the robustness of our results using statutory marginal tax rates (MTRs). As discussed in Moretti and Wilson (2017), ATRs and MTRs are highly correlated, and indeed, we obtain similar results regardless of the choice of tax rates.

[8]Data on state taxes and tax credits for the years 1977 to 2009 are provided by Moretti and Wilson (2017).

[9]The number of internal inventors is 537, 717, and that of external inventors is 736, 475. Internal inventors are classified into three groups: 89, 637 inventors share the same assignee as the migrating top inventors and are co-inventors of the migrating top inventors, 419, 378 inventors share the same assignee as the migrating top inventors but are not co-inventors of the migrating top inventors, and 28, 702 inventors do not share the same assignee as the migrating top inventors but are co-inventors of the migrating top inventors.

The knowledge of the migrating top inventors can be shared with internal inventors within the same organization and/or through co-inventors relationships ("knowledge in the lab") or can spill over to external inventors within the same commuting zone ("knowledge in the air"). We call the former *internal knowledge sharing* and the latter *external knowledge spillovers*.

3 Tax differences and the migration of top inventors

In this section, we estimate the impact of tax differences across states on the migration of top inventors from origin commuting zone o to destination commuting zone d, which reproduces the results in Moretti and Wilson (2017). In the next section, we use the predicted migration of top inventors to develop a new method of constructing a Bartik instrument.

Let $\sigma(o)$ and $\sigma(d)$ denote the states to which origin and destination commuting zones belong, respectively. In the beginning of period t, top inventors in o, whose number is denoted by I_{ot}, observe individual income tax rates in origin and all possible destination commuting zones, $\tau_{\sigma(o)t}$ and $\{\tau_{\sigma(d)t}\}_{d\neq o}$. By the end of period t, they decide whether to migrate to d or to stay in o. The number and share of top inventors who migrate from o to d in period t is defined as M_{odt} and $P_{odt} = M_{odt}/I_{ot}$, respectively. Similarly, the number and share of top inventors who stay in o in period t is defined as M_{oot} and $P_{oot} = M_{oot}/I_{ot}$.

3.1 Inventors

In each period, top inventors choose the location that gives them the highest utility. The utility of top inventor i, who lived in commuting zone o in the previous period and moves to commuting zone d in the current period t, is given by $U_{iodt} = \alpha \ln(1-\tau_{\sigma(d)t}) + \alpha \ln w_{dt} + Z_d - C_{od} + \varepsilon_{iodt}$, where $\tau_{\sigma(d)t}$ and w_{dt} are the individual income tax rate and wage in d, respectively; α is the coefficient on the log of after-tax income; Z_d captures consumption amenities and the cost of living in d; C_{od} is the cost of migration measured in utility; and ε_{iodt} represents time-varying idiosyncratic preferences for locations. The utility of top inventor i who stays in o is given by $U_{ioot} = \alpha \ln(1-\tau_{\sigma(o)t}) + \alpha \ln w_{ot} + Z_o - C_{oo} + \varepsilon_{ioot}$, where we assume that $C_{oo} = 0$. Taking the difference between U_{iodt} and U_{ioot} yields the utility change for top inventor i, conditional on moving from o to d. Assume that ε_{iodt} is independent and identically Gumbel distributed. Let P_{odt}/P_{oot} denote the share of top inventors who move from o to d relative to

the share of top inventors who stay in o. The log odds ratio for top inventors is then given by

$$\ln(P_{odt}/P_{oot}) = \alpha[\ln(1-\tau_{\sigma(d)t}) - \ln(1-\tau_{\sigma(o)t})] + \alpha[\ln w_{dt} - \ln w_{ot}] + [Z_d - Z_o] - C_{od}. \quad (1)$$

3.2 Firms

In each period, firms choose a location and hire a top inventor to maximize profit. The profit of firm j, which was located in commuting zone o in the previous period and moves to commuting zone d in the current period t, is given by $\ln \pi_{jodt} = \beta \ln(1-\tau'_{\sigma(d)t}) - \ln w_{dt} + Z'_d - C'_{od} + \varepsilon'_{jodt}$, where $\tau'_{\sigma(d)t}$ stands for state policies such as the CITR, ITC, and RTC in $\sigma(d)$; Z'_d captures production amenities in d; C'_{od} is the cost of migration for a firm; and ε'_{jodt} represents time-varying idiosyncratic firm productivity shocks. As in the case with inventors, assume that $C'_{oo} = 0$, and that ε'_{jodt} is independent and identically Gumbel distributed. Let P'_{odt}/P'_{oot} denote the share of firms that move from o to d relative to the share of firms that stay in o. The log odds ratio for firms is then given by

$$\ln(P'_{odt}/P'_{oot}) = \beta[\ln(1-\tau'_{\sigma(d)t}) - \ln(1-\tau'_{\sigma(o)t})] - [\ln w_{dt} - \ln w_{ot}] + [Z'_d - Z'_o] - C'_{od}. \quad (2)$$

3.3 Equilibrium

In equilibrium, the demand for top inventors must equal the supply of top inventors in each commuting zone in each year. To derive an equilibrium relationship between tax differences and the migration of top inventors, we first solve (2) for $\ln w_{dt} - \ln w_{ot}$. We then plug the resulting expression into (1) and set $\ln(P'_{odt}/P'_{oot}) = \ln(P_{odt}/P_{oot})$ as in Moretti and Wilson (2017), which yields the equation we estimate as follows (see Appendix B.1 for the derivation):

$$\begin{aligned} \ln(P_{odt}/P_{oot}) \;=\; & \eta[\ln(1-\tau_{\sigma(d)t}) - \ln(1-\tau_{\sigma(o)t})] + \eta'[\ln(1-\tau'_{\sigma(d)t}) - \ln(1-\tau'_{\sigma(o)t})] \\ & + \gamma_d + \gamma_o + \gamma_{od} + u_{odt}, \end{aligned} \quad (3)$$

where $\eta = \frac{\alpha}{1+\alpha}$ and $\eta' = \frac{\alpha\beta}{1+\alpha}$ are parameters governing inventor and firm mobility, respectively; $\gamma_d = \frac{1}{1+\alpha}[Z_d + \alpha Z'_d]$ and $\gamma_o = -\frac{1}{1+\alpha}[Z_o + \alpha Z'_o]$ are destination and origin fixed effects that account for consumption and production amenities; $\gamma_{od} = -\frac{1}{1+\alpha}[C_{od} + \alpha C'_{od}]$ denotes fixed effects that are specific to each pair of commuting zones to capture the cost of migration for inventors and firms; and u_{odt} is an error term. We consider different combinations of fixed effects in the next subsection.

3.4 Estimation

When estimating (3), we proxy $\tau_{\sigma(d)t}$ by the ATR for a hypothetical taxpayer at the ninety-fifth or ninety-ninth percentile of the U.S. income distribution because, as in Moretti and Wilson (2017), we do not observe top inventors' income.[10] We regard $\tau'_{\sigma(d)t}$ as consisting of the CITR, ITC, and RTC. We use different combinations of fixed effects in (3), as well as year fixed effects or region pair × year fixed effects and report robust standard errors that allow for three-way clustering by commuting zone pair, origin-state × year, and destination-state × year.

Table 4: The impact of tax differences on the migration of top inventors.

	(1)	(2)	(3)	(4)
$\Delta \ln(1 - \text{ATR})$	7.357	6.902	6.406	6.586
	(1.611)	(1.420)	(1.292)	(1.124)
$\Delta \ln(1 - \text{CITR})$	-0.435	-0.195	-0.300	-0.140
	(1.058)	(0.999)	(0.812)	(0.717)
$\Delta \ln(1 + \text{ITC})$	0.172	-0.083	0.118	-0.034
	(0.737)	(0.688)	(0.993)	(0.689)
$\Delta \ln(1 + \text{RTC})$	0.323	0.311	0.377	0.178
	(0.443)	(0.395)	(0.321)	(0.281)
CZ pair FE	Yes	Yes	No	No
Origin CZ FE and destination CZ FE	No	No	Yes	Yes
Year FE	Yes	No	Yes	No
Region pair × year FE	No	Yes	No	Yes
Observations	4,866	4,866	7,226	7,225
$\overline{R}^2$ (total)	0.893	0.904	0.907	0.917
$\overline{R}^2$ (within)	0.400	0.458	0.411	0.013

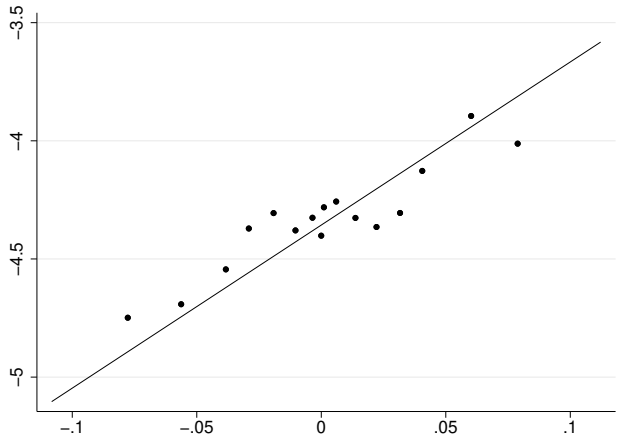
Notes: The dependent variable in each column is the log odds ratio in equation (3). ATR, CITR, ITC, and RTC stand for the individual income average tax rate at the ninety-fifth percentile, corporate income tax rate, investment tax credit, and R&D tax credit, respectively. $\Delta \ln(1 - \text{ATR})$ is defined as $\ln(1 - \text{ATR}_{\sigma(d)t}) - \ln(1 - \text{ATR}_{\sigma(o)t})$. $\Delta \ln(1 - \text{CITR})$, $\Delta \ln(1 + \text{ITC})$, and $\Delta \ln(1 + \text{RTC})$ are defined analogously. Cluster-robust standard errors are in parentheses.

Table 4 shows that the interstate migration results in Moretti and Wilson (2017) can be replicated fairly well at the commuting zone level: The elasticity of the migration of top inventors with respect to the difference in ATRs between origin and destination is positive and significant in all cases. In what follows, we use the result in Column 2 of Table 4 since the specification is most closely related to their baseline case.

[10] We report the results at the ninety-fifth percentile in the main body and the results at the ninety-ninth percentile in Appendix C.1 as robustness checks. In Appendix C.1, we further check the robustness of our results using statutory marginal tax rates (MTRs) and average property tax rates (APTRs).

Figure 3 illustrates a binned scatter plot, where the vertical axis is the log odds ratio of top inventor migrations, $\ln(P_{odt}/P_{oot})$, and the horizontal axis is the difference in ATRs between destination and origin commuting zones, $\Delta \ln(1-\text{ATR}) = \ln(1-\text{ATR}_{\sigma(d)t}) - \ln(1-\text{ATR}_{\sigma(o)t})$.[11] The figure reveals a pronounced tendency for top inventors to migrate from commuting zones with higher individual income tax rates to those with lower rates.

Figure 3: Binned scatter plot of the relationship between top inventor migrations and ATRs.



Notes: This figure illustrates a binned scatter plot, where the vertical axis is the log odds ratio of top inventor migrations, $\ln(P_{odt}/P_{oot})$, and the horizontal axis is the difference in ATRs between destination and origin commuting zones, $\Delta \ln(1-\text{ATR}) = \ln(1-\text{ATR}_{\sigma(d)t}) - \ln(1-\text{ATR}_{\sigma(o)t})$. It is depicted using the Stata package `binsreg` (see Cattaneo et al., 2024), where we incorporate commuting zone pair fixed effects and region pair × year fixed effects, as well as CITRs, ITCs, and RTCs, as covariate adjustments.

4 The migration of top inventors and local patent productivity

To analyze the impact of top inventor inflows on local patent productivity, we first present a new method to construct a Bartik instrument based on the estimated flows of top inventors obtained in Section 3. We then show the main results using a static framework. To check the robustness of our main results, we further consider a dynamic setting in the next section. In both cases, we estimate two types of effects: (a) the productivity gains of all local inventors and (b) those of external inventors. The former effect can be interpreted as local aggregate productivity gains from both internal knowledge sharing and external knowledge spillovers. The latter effect can be viewed as the gains from external knowledge spillovers that go beyond organizational boundaries and co-inventor relationships. Our main focus is on the latter gains since they pertain to what Marshall (1890) referred to as the mysteries of trade in the air.

[11] $\text{ATR}_{\sigma(d)t}$ denotes the ATR in state $\sigma(d)$ in year t. We use similar notation for other state policy variables.

4.1 Bartik instrument

To construct a Bartik instrument, we start with the identity regarding top inventor inflows, $M_{dt} = \sum_{o \neq d} M_{odt}$; i.e., the total number of top inventors migrating to destination commuting zone d in period t equals the sum of the number of top inventors migrating from origin commuting zone o to destination commuting zone d across all $o \neq d$ in period t. Recalling that $P_{odt} = M_{odt}/I_{ot}$, we have

$$M_{dt} = \sum_{o \neq d} P_{odt} I_{ot}, \tag{4}$$

where the right-hand side consists of the share P_{odt} of top inventors migrating from o to d in period t and the shift I_{ot}, i.e., the number of top inventors in o at the beginning of period t.

As shown by Goldsmith-Pinkham et al. (2020) and Borusyak et al. (2022), there are two approaches to ensuring the exogeneity of a Bartik instrument. The former consider share exogeneity, whereas the latter employ shift exogeneity. We rely on the former in this section and check the robustness of the results using the latter in Section 5.5.

Our novelty lies in combining the share exogeneity approach in Goldsmith-Pinkham et al. (2020) with the estimates obtained from the location choice model à la Moretti and Wilson (2017). Let $\{\widehat{\eta}, \widehat{\eta}', \widehat{\gamma}_d, \widehat{\gamma}_o, \widehat{\gamma}_{od}\}$ denote the estimates from (3). Using these estimates, we first compute the predicted probability that a top inventor, who lived in o at the beginning of period t, moves to d by the end of period t as follows (see Appendix B.2 for the derivation):

$$\widehat{P}_{odt} = \frac{\exp\{\widehat{\eta} \ln(1 - \tau_{\sigma(d)t}) + \widehat{\eta}' \ln(1 - \tau'_{\sigma(d)t}) + \widehat{\gamma}_d + \widehat{\gamma}_{od}\}}{\sum_{c \in \mathcal{C}} \exp\{\widehat{\eta} \ln(1 - \tau_{\sigma(c)t}) + \widehat{\eta}' \ln(1 - \tau'_{\sigma(c)t}) + \widehat{\gamma}_c + \widehat{\gamma}_{oc}\}}, \tag{5}$$

where $\mathcal{C}$ is the set of *all* commuting zones including origin commuting zone o and destination commuting zone d.[12] We then construct a Bartik instrument by replacing the share P_{odt} in (4) with the predicted share $\widehat{P}_{odt}$ in (5) as follows:

$$B_{dt} = \sum_{o \neq d} \widehat{P}_{odt} I_{ot}. \tag{6}$$

We use the Bartik instrument (6) in the first-stage regression in Section 4.2 and discuss the share exogeneity in Section 4.3 and the relevance and validity of the instrument in Section 4.5.

[12] When estimating (3), we do not simultaneously use the set of fixed effects $\{\widehat{\gamma}_d, \widehat{\gamma}_o, \widehat{\gamma}_{od}\}$. Recall that in Table 4, we consider $\widehat{\gamma}_{od}$ in Columns 1-2 and adopt $\{\widehat{\gamma}_d, \widehat{\gamma}_o\}$ in Columns 3-4. In what follows, we thus modify the way we incorporate fixed effects into (5) according to empirical specifications.

4.2 Empirical specifications

We start with the fixed effect (FE) model:

$$\ln Y_{dt} = \phi M_{dt} + \xi X_{dt} + \varepsilon_{dt}, \tag{7}$$

where Y_{dt} is the patent productivity of local inventors in commuting zone d in period t (defined as the number of patents by all local inventors or by external inventors), $M_{dt} = \sum_{o \neq d} M_{odt}$ is the number of top inventors who migrate to destination commuting zone d in period t, ϕ stands for the impact of a top inventor inflow on local patent productivity, and ε_{dt} is an i.i.d. shock. In the main analysis, X_{dt} includes commuting zone fixed effects, δ_d, year fixed effects, δ_t, and the ATR at the ninety-fifth percentile of the U.S. income distribution, $\tau_{\sigma(d)t}$.[13] As robustness checks, we incorporate time-varying factors in commuting zone d, as well as other taxes and tax credits in state $\sigma(d)$, into X_{dt}.[14] When estimating (7), we cluster standard errors at the commuting zone level.[15]

However, the top inventor inflows M_{dt} may be endogenous due to reverse causality or the existence of omitted variables that have direct impacts on both top inventor inflows and local patent productivity. Reverse causality arises when greater local patent productivity attracts top inventors, whereas omitted variables exist when there are unobserved consumption and production amenities that have been studied since Roback (1982).

To address these endogeneity issues, we consider an instrumental variable (IV) regression, which consists of the structural equation

$$\ln Y_{dt} = \phi^s M_{dt} + \xi^s X_{dt} + \varepsilon^s_{dt}, \tag{8}$$

and the first-stage regression

$$M_{dt} = \psi^f B_{dt} + \xi^f X_{dt} + \varepsilon^f_{dt}, \tag{9}$$

[13] As robustness checks, we consider the ATR at the ninety-ninth percentile, MTR, APTR, and an alternative way of detecting top inventor migrations in Appendix C.1.

[14] Specifically, we consider manufacturing employment in commuting zone d and the CITR, ITC, RTC, and ATR at the fiftieth percentile in state $\sigma(d)$ in Appendix C.2. We also incorporate other employment variables at the commuting zone level such as "finance and insurance," "professional, scientific, and technical services," and "management of companies and enterprises" into X_{dt} in the specification curve analysis in Figure 4.

[15] When estimating (7), we replace $\ln Y_{dt}$ with $\ln(1 + Y_{dt})$ in the main analysis to accommodate commuting zone $\times$ year observations with no patents. As a robustness check, we drop such observations and estimate (7) while retaining $\ln Y_{dt}$. As shown in Appendix C.3, the results are quite similar to those in the main analysis.

where B_{dt} is the Bartik instrument given by (6). When estimating (8), we cluster standard errors at the commuting zone level.[16] We further consider two variants of the Bartik instrument to assess the sensitivity of our results. One is the prediction of between-state top inventor flows, $B^{\sigma}_{dt} = \sum_{o \notin \sigma(d)} \widehat{P}_{odt} I_{ot}$, to highlight state tax differences. The other is the predicted top inventor flows into commuting zone $\nu(d)$, which is the nearest neighborhood of d, $B^{\nu}_{dt} = \sum_{o \neq d, \nu(d)} \widehat{P}_{o\nu(d)t} I_{ot}$, to take a spatial lag of B_{dt}.

4.3 Potential threats to identification

Recall that the Bartik instrument B_{dt} consists of the shares $\widehat{P}_{odt}$ and the shifts I_{ot}.[17] It is based on the migration identity and predicts the top inventor flows to destination d by the sum of the products of these two elements. In the main analysis, we follow the shares perspective; i.e., it is the shares $\widehat{P}_{odt}$ that provide the exogenous variation satisfying $E(\varepsilon^s_{dt} \widehat{P}_{odt} | X_{dt}) = 0$, and the shifts I_{ot} do not affect the identification of ϕ^s provided that the shares are exogenous (Goldsmith-Pinkham et al., 2020; Borusyak et al., 2025).

We show in Appendix E.1 that the share exogeneity, $E(\varepsilon^s_{dt} \widehat{P}_{odt} | X_{dt}) = 0$, holds under the following two assumptions: (i) ε^s_{dt} is mean zero conditional on X_{dt}, i.e., $E(\varepsilon^s_{dt} | X_{dt}) = 0$, and (ii) ε^s_{dt} and other state taxes $\{\tau_{\sigma(c)t}\}_{c \notin \sigma(d)}$ are independent conditional on X_{dt}, i.e., $\varepsilon^s_{dt} \perp \tau_{\sigma(c)t} | X_{dt}$ for $c \notin \sigma(d)$ and $c, d \in \mathcal{C}$, where we let $X_{dt} = \{\tau_{\sigma(d)t}, \delta_d, \delta_t\}$ in the main analysis.[18]

One may worry that the first assumption, $E(\varepsilon^s_{dt} | X_{dt}) = 0$, may not hold due to a possible correlation between ε^s_{dt} and $\tau_{\sigma(d)t}$ through unobserved state-specific time-varying factors. To alleviate potential concerns that state taxes may respond to local economic conditions or be correlated with local economic policies affecting innovation, we follow Akcigit et al. (2022) and employ alternative specifications with state $\times$ year fixed effects $\delta_{\sigma(d)t}$. Specifically, we replace $X_{dt} = \{\tau_{\sigma(d)t}, \delta_d, \delta_t\}$ and ε^s_{dt} in the structural equation (8) with $X'_{dt} = \{\delta_{\sigma(d)t}, \delta_d, \delta_t\}$ and ζ^s_{dt}, respectively, so that $\delta_{\sigma(d)t}$ subsumes $\tau_{\sigma(d)t}$. We can then recover the share exogeneity $E(\zeta^s_{dt} \widehat{P}_{odt} | X'_{dt}) = 0$ by replacing the assumption $E(\varepsilon^s_{dt} | X_{dt}) = 0$ with $E(\zeta^s_{dt} | X'_{dt}) = 0$ (see Section 4.4 and Appendix E.2).

[16] When estimating (8), we replace $\ln Y_{dt}$ with $\ln(1 + Y_{dt})$ in the main analysis to accommodate commuting zone $\times$ year observations with no patents. As a robustness check, we drop such observations and estimate (8) while retaining $\ln Y_{dt}$. As shown in Appendix C.3, the results are quite similar to those in the main analysis.

[17] In what follows, we refer to the predicted shares $\widehat{P}_{odt}$ as the shares for simplicity when there is no confusion.

[18] $\mathfrak{A} \perp \mathfrak{B} | \mathfrak{C}$ denotes the independence of $\mathfrak{A}$ and $\mathfrak{B}$ conditional on $\mathfrak{C}$.

The second assumption is the exclusion restriction. The shares $\widehat{P}_{odt}$ in (5) depend not only on the destination tax rate $\tau_{\sigma(d)t}$ but also on the distribution of tax rates in all states $\{\tau_{\sigma(c)t}\}_{c\in\mathcal{C}}$. Thus, in the destination specific IV regression model (8) and (9), the state taxes, $\{\tau_{\sigma(c)t}\}_{c\notin\sigma(d)}$, other than that in the destination, $\tau_{\sigma(d)t}$, have an indirect effect on the local patent productivity Y_{dt} only via $\widehat{P}_{odt}$ in the Bartik instrument B_{dt}, given this assumption. Thus, it works as an exclusion restriction for identifying the patent productivity effect ϕ^s.[19,20]

4.4 Main results

Table 5 presents the estimation results for the FE and IV regressions. Column 1 reports the FE case. Columns 2-7 are the results for different IV regressions. Column 2 considers the Bartik instrument $B_{dt} = \sum_{o\neq d} \widehat{P}_{odt} I_{ot}$ in (6). Column 3 adds to Column 2 its variant $B^{\sigma}_{dt} = \sum_{o\notin\sigma(d)} \widehat{P}_{odt} I_{ot}$, which captures top inventor flows only from other states. Column 4 further adds to Column 3 the other variant $B^{\nu}_{dt} = \sum_{o\neq d,\nu(d)} \widehat{P}_{o\nu(d)t} I_{ot}$. It involves top inventor flows from origin commuting zones $o \neq d$ to commuting zone $\nu(d)$, which is the nearest neighborhood of d. Columns 5-7 replace $\ln(1-\text{ATR})$ in Columns 2-4 with state $\times$ year fixed effects, $\delta_{\sigma(d)t}$, to control for time-varying state-specific unobservables, which alleviates the concern that there may be a correlation between ε^s_{dt} and $\tau_{\sigma(d)t}$. In both Panels (a) and (b), we exclude the patents by top inventors who moved in from the dependent variable.

As seen from Table 5, the semi-elasticities of local patent productivity with respect to top inventor inflows, as well as the elasticities of local patent productivity with respect to $1-\text{ATR}$, are virtually identical for all IV regressions within each panel.[21] Panel (a) in Table 5 shows that an inflow of a top inventor raises the patent productivity of all local inventors by approximately 6%, which can be interpreted as local aggregate gains from both knowledge sharing among internal inventors and knowledge spillovers to external inventors. Panel (b) shows that a top inventor inflow raises local patent productivity by approximately 4% when

[19] To address the concern that other state taxes $\{\tau_{\sigma(c)t}\}_{c\notin\sigma(d)}$ may influence the inflows of non-top inventors in d, thereby enhancing Y_{dt}, we estimate (8) and (9) by directly excluding such potential inflows in Section 5.3.

[20] To alleviate the concern that state tax competition may induce correlation between ε^s_{dt} and $\{\tau_{\sigma(c)t}\}_{c\notin\sigma(d)}$, we examine the possibility of strategic interactions among state governments by estimating a reaction function, where the income tax in one state responds to the income taxes in other states (see, e.g., Brueckner, 2003). In line with the exclusion restriction, we do not find strong evidence for state tax competition (see Appendix D).

[21] The semi-elasticities of local patent productivity with respect to top inventor inflows are somewhat smaller for the FE regression. One possible explanation for this is the presence of urban costs such as land rents and commuting costs that are specific to commuting zones and can vary over time (see Duranton and Puga, 2020).

Table 5: The impact of top inventor inflows on local patent productivity.

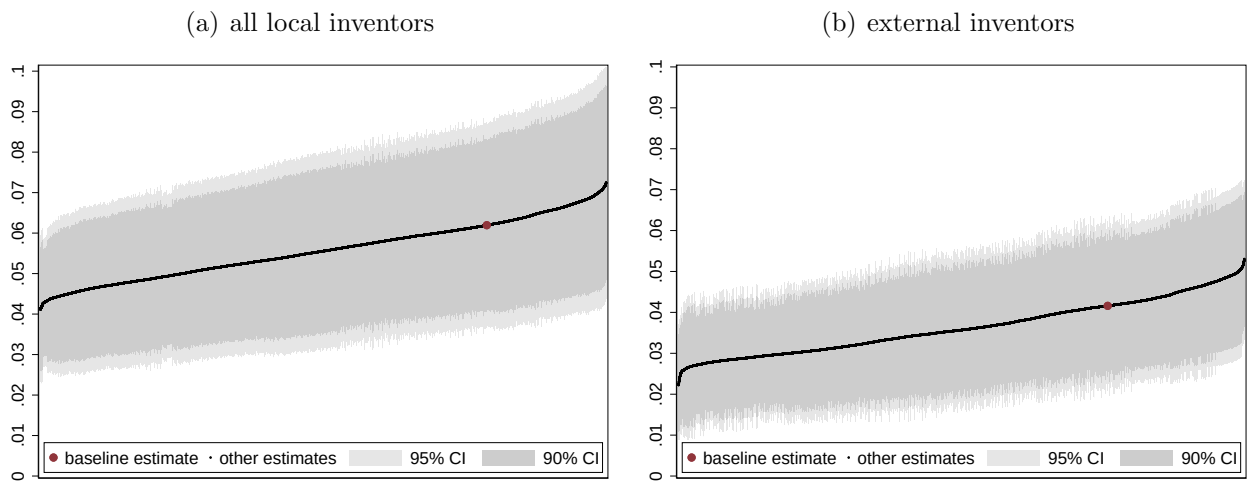
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(a) All local inventors							
Top inventor inflows	0.043	0.062	0.060	0.059	0.066	0.059	0.060
	(0.006)	(0.013)	(0.012)	(0.011)	(0.014)	(0.013)	(0.012)
$\ln(1-\text{ATR})$	6.041	5.915	5.899	6.017			
	(1.038)	(1.040)	(1.038)	(1.038)			
Effective F statistic		37.755	33.377	33.040	51.824	35.130	34.997
$\tau = 5\%$		37.418	31.930	34.734	37.418	31.214	32.989
$\tau = 10\%$		23.109	19.892	21.389	23.109	19.473	20.364
$\tau = 20\%$		15.062	13.094	13.901	15.062	12.839	13.272
$\tau = 30\%$		12.039	10.531	11.093	12.039	10.336	10.610
(b) External inventors							
Top inventor inflows	0.027	0.042	0.040	0.041	0.041	0.036	0.038
	(0.004)	(0.010)	(0.009)	(0.009)	(0.011)	(0.010)	(0.010)
$\ln(1-\text{ATR})$	4.781	4.684	4.641	4.616			
	(0.848)	(0.851)	(0.848)	(0.842)			
Effective F statistic		37.755	33.377	33.040	51.824	35.130	34.997
$\tau = 5\%$		37.418	31.921	34.738	37.418	31.203	32.988
$\tau = 10\%$		23.109	19.887	21.392	23.109	19.467	20.364
$\tau = 20\%$		15.062	13.091	13.902	15.062	12.835	13.272
$\tau = 30\%$		12.039	10.529	11.094	12.039	10.333	10.610
CZ FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	No	No	No
State × year FE	No	No	No	No	Yes	Yes	Yes
Observations	23,628	23,628	23,628	23,463	23,562	23,562	23,397

Notes: The coefficient on top inventor inflows is converted to semi-elasticity. ATR stands for the individual income average tax rate at the ninety-fifth percentile. The coefficient on $\ln(1-\text{ATR})$ is converted to elasticity. Column 1 does not control for the endogeneity of top inventor inflows. Column 2 uses B_{dt} as an instrument. Column 3 uses B_{dt} and B^{σ}_{dt} as instruments. Column 4 uses B_{dt}, B^{σ}_{dt}, and B^{ν}_{dt} as instruments. Columns 5-7 replace $\ln(1-\text{ATR})$ in Columns 2-4 with state × year FE. Cluster-robust standard errors are in parentheses.

we focus on external inventors who are not directly connected to the migrating top inventors. The latter result can be interpreted as evidence for the existence of the mysteries of trade in the air as the number reflects neither knowledge flows within the same assignee nor those between co-inventors. We thus disentangle productivity gains due to external knowledge spillovers ("knowledge in the air") from those due to internal knowledge sharing ("knowledge in the lab").

Our positive causal results differ from Borjas and Doran (2012), who find a negative impact of the inflows of Soviet mathematicians on the productivity of U.S. mathematicians, and from Moser et al. (2014), who summarize that knowledge spillovers from German Jewish émigrés

Figure 4: Specification curve analysis.



Notes: Panels (a) and (b) illustrate the impacts of a top inventor inflow on the patent productivity of all local inventors and that of external inventors, respectively. Each specification curve is depicted using 31,969 alternative specifications, as explained in footnote 23. The vertical axis is the estimated value of $\widehat{\phi}^s$.

to incumbent U.S. inventors are unlikely to have been the main driver of the U.S. patent productivity gains.[22] Even at finer geographical scales, our results from the IV regressions are in contrast to Zacchia (2018), who finds no city-wide spillover effect of inventor inflows. Our main finding—the existence of gains from knowledge in the air (i.e., external inventors' productivity gains due to external knowledge spillovers) at the commuting zone level—also differs from De la Roca and Puga (2017), Moretti (2021), and Prato (2025) in that these studies analyze the impacts on those who migrate themselves or consider internal knowledge sharing through organizations or co-inventor relationships. Furthermore, our 4-6% patent productivity gains for local inventors due to an additional inventor inflow could be compared with the 12% increase in incumbent plants' TFP due to a new plant opening in Greenstone et al. (2010) or with the 62% increase in local patent productivity due to the establishment of a new college in Andrews (2023). However, given the difference between top inventor arrival on the one hand and firm entry and college establishment on the other hand, it is not surprising that the former effect is smaller than the latter effect.

We relegate extensive robustness checks of these main results to Section 5. However, before concluding this subsection, let us briefly illustrate the specification curve analysis as in Simonsohn et al. (2020). We employ different specifications of the IV regressions by consid-

[22]We elaborate on incumbents' productivity gains in Section 5.3.

ering various dimensions.[23] Figure 4 plots the specification curve for $\widehat{\phi}^s$ with 90% and 95% confidence intervals. As seen from Panels (a) and (b), the productivity gains of approximately 6% and 4% are fairly robust for 31,969 alternative specifications, thus verifying that our main estimates do not come from data mining. In what follows, we use the specification in Column 3 of Table 5 as a baseline unless otherwise stated.

4.5 Relevance and validity of Bartik instruments

Our empirical strategy relies on the relevance and validity of the Bartik instruments, which we discuss in what follows.

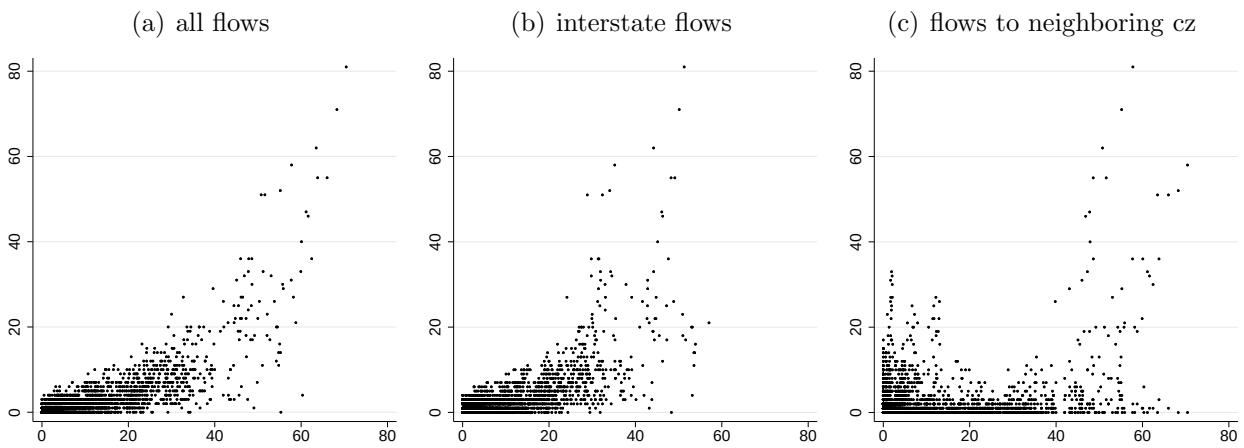
To assess the relevance of the Bartik instruments, we first plot in Figure 5 the relationship between the actual top inventor flows $M_{dt} = \sum_{o \neq d} M_{odt}$ and the Bartik instruments constructed from the predicted top inventor flows. For the latter, we consider $B_{dt} = \sum_{o \neq d} \widehat{P}_{odt} I_{ot}$, $B^{\sigma}_{dt} = \sum_{o \notin \sigma(d)} \widehat{P}_{odt} I_{ot}$, and $B^{\nu}_{dt} = \sum_{o \neq d, \nu(d)} \widehat{P}_{o\nu(d)t} I_{ot}$, in Panels (a), (b), and (c), respectively. There is a positive relationship in each panel, and the correlation coefficients for Panels (a), (b), and (c) are given by 0.78, 0.74, and 0.38, respectively.

We further apply a test for weak instruments developed by Montiel Olea and Pflueger (2013) to these Bartik instruments. The test is robust to heteroskedasticity, autocorrelation, and clustering (see also Andrews et al., 2019). The bottom of each panel in Table 5 reports the effective F statistic, which is a scaled version of the nonrobust first-stage F statistic. Following their baseline, we set the threshold at $\tau = 10\%$ and the significance at 5%. In all cases, the effective F statistic exceeds the critical value reported at $\tau = 10\%$, thus rejecting the null hypothesis of weak instruments.

To address the validity of the Bartik instruments, we follow the shares perspective and focus on the exogeneity of the shares $\widehat{P}_{odt}$. We assess the plausibility of the exogeneity assump-

[23] We consider (i) whether to use the ATR at the ninety-fifth or ninety-ninth percentile; (ii) whether to use $\ln(1+Y_{dt})$ or drop commuting zones with $Y_{dt} = 0$; (iii) whether to use $\{B_{dt}\}$, $\{B_{dt}, B^{\sigma}_{dt}\}$ or $\{B_{dt}, B^{\sigma}_{dt}, B^{\nu}_{dt}\}$ as instruments; (iv) whether to include state $\times$ year fixed effects; (v) whether to use the baseline or alternative detection of top inventor migrations; (vi) whether to use the baseline or alternative definition of local inventors (see the analysis on incumbents' productivity gains in Section 5.3); (vii) whether to use ATRs or MTRs; (viii) whether to include APTRs; and (ix) whether to include each of the other controls (ATR50, CITR, ITC, RTC, manufacturing employment, and other employment variables such as "finance and insurance," "professional, scientific, and technical services," and "management of companies and enterprises" to capture other omitted variables, e.g., access to venture capital). Since the usual caveat on weak instruments is applicable here, we adopt only specifications for which the null hypothesis of weak instruments is rejected.

Figure 5: Actual versus predicted top inventor flows.



Notes: In each panel, the vertical and horizontal axes are the actual top inventor flows and the Bartik instruments constructed from the predicted top inventor flows, respectively. The actual flows in Panels (a), (b), and (c) are defined as M_{dt}. The Bartik instruments in Panels (a), (b), and (c) are B_{dt}, B^{σ}_{dt}, and B^{ν}_{dt}, respectively.

tion in two steps. We first use the decomposition result in Goldsmith-Pinkham et al. (2020) to rewrite the overall estimate of the productivity effect as $\widehat{\phi}^s = \sum_{o \in \mathcal{C}} \widehat{\omega}_o \widehat{\phi}^s_o$, which consists of the origin-specific weight $\widehat{\omega}_o$ and the origin-specific productivity effect $\widehat{\phi}^s_o$. The former is referred to as the Rotemberg weight (Rotemberg, 1983) and measures to what extent the bias originating from commuting zone o contributes to the overall bias. For each of the top five origin commuting zones by Rotemberg weight, we then check the correlation between the predicted probabilities of top inventor migrations to the destination commuting zones and key pre-period destination characteristics. This approach allows us to assess the exogeneity of the shares $\widehat{P}_{odt}$.

We summarize the results in Appendix F. Table F1 presents the summary of the Rotemberg weights. The origin commuting zones with the top five highest weights are Bergen-Essex-Middlesex, Cook-DuPage-Lake, Kings-Queens-NewYork, Philadelphia-Montgomery-Delaware, and Allegheny-Westmoreland-Washington.[24] Table F2 reports the destination commuting zones to which top inventors moved from these five origin commuting zones. The result that origin and destination states differ in almost all cases is in line with the assumption that the main source of identifying variation comes from interstate top inventor migrations induced by personal income tax differences between states. Table F3 reports, for each of the top five origin

[24]The name of each commuting zone shown here is a list of three counties with the largest numbers of inventors (in descending order) in that commuting zone.

commuting zones, the correlation between the predicted shares of top inventor migrations to the destination commuting zones and the log of lagged employment in four sectors: manufacturing; finance and insurance; professional, scientific, and technical services; and management of companies and enterprises. These sectors are chosen as they may correlate with unobservable confounders. The correlations are consistently low, and the coefficients obtained from regressing the predicted shares on the log of lagged sectoral employment are not significant at the 5% level in all specifications, thus supporting the validity of our Bartik instruments.

5 Robustness

In this section, we examine the robustness of our main results in terms of time, space, aggregation, an alternative high-income occupation, and an alternative exogeneity assumption. We first extend our static framework to a dynamic setting, which allows us to assess the impacts on local patent productivity before and after top inventor inflows. We then check the robustness in terms of the geographic extent of productivity gains. We further analyze two disaggregated cases: where local inventors are classified by their patent productivity; and where local inventors differ in terms of how many years they have stayed when top inventors arrive. The latter case assesses whether the productivity gains come from the entry of new local inventors ("extensive margins") or from incumbent local inventors ("intensive margins") by excluding potential inflows of non-top inventors. We also perform a falsification exercise using an alternative high-income occupation. Finally, we consider the shifts perspective (Borusyak et al., 2022) instead of the shares perspective (Goldsmith-Pinkham et al., 2020).

In Appendix C, we conduct additional robustness checks. Specifically, we consider the alternative individual income average tax rate at the ninety-ninth percentile of the U.S. income distribution (ATR99), statutory marginal tax rates (MTRs), average property tax rates (APTRs), and an alternative way of detecting top inventor migrations in Appendix C.1. We also include other controls (ATR50, CITR, ITC, RTC, and manufacturing employment) in Appendix C.2, and drop commuting zone $\times$ year observations with no patents in Appendix C.3. In Appendix C.4, we further use the alternative measure of patent productivity based on patent quality in Kogan et al. (2017). As seen from Tables C2, C4, C6, C8, C9, C10, and C12, all results are consistent with our main findings.

5.1 Dynamic analysis

As a first robustness check, we conduct event study analysis to examine whether our main results are sensitive to the inclusion of lead and lag effects of top inventor inflows. To this end, we rewrite (8) as $\ln Y_{dt} = \sum_{j=\underline{j}+1}^{\overline{j}} \phi_j^{\ell} M_{dt-j} + \xi^{\ell} X_{dt} + \varepsilon_{dt}^{\ell}$, which is a distributed lag model in levels with a binning window $[\underline{j}+1, \overline{j}]$. Thus, when $\underline{j} = -1$ and $\overline{j} = 0$, this extended model degenerates into the static model (8). Schmidheiny and Siegloch (2023) show that the foregoing equation is equivalent to the event study model given by[25]

$$\ln Y_{dt} = \sum_{j=\underline{j}}^{\overline{j}} \mu_j^{es} \Delta M_{dt}^{(j)} + \xi^{es} X_{dt} + \varepsilon_{dt}^{es}, \quad \text{where} \tag{10}$$

$$\Delta M_{dt}^{(j)} = \begin{cases} \sum_{k=-\infty}^{\underline{j}} (M_{dt-k} - M_{dt-k-1}) & \text{if} \quad j = \underline{j} < 0 \\ M_{dt-j} - M_{dt-j-1} & \text{if} \quad \underline{j} < j < \overline{j} \\ \sum_{k=\overline{j}}^{\infty} (M_{dt-k} - M_{dt-k-1}) & \text{if} \quad j = \overline{j} > 0 \end{cases} . \tag{11}$$

Our aim is to estimate $\{\mu_{\underline{j}}^{es}, \mu_{\underline{j}+1}^{es}, ..., \mu_{\overline{j}-1}^{es}, \mu_{\overline{j}}^{es}\}$ with normalization $\mu_{-1}^{es} = 0$. The event study coefficients capture the cumulative effect of the event of top inventor inflows, i.e., $\mu_j^{es} = \mu_{j-1}^{es} + \phi_j^{\ell} = \sum_{h=0}^{j} \phi_h^{\ell}$ for $j = 0, 1, ..., \overline{j}$ and $\mu_j^{es} = \mu_{j+1}^{es} - \phi_{j+1}^{\ell} = -\sum_{h=j+1}^{-1} \phi_h^{\ell}$ for $j = -2, -3, ..., \underline{j}$. Thus, the coefficients for $j \geq 0$ denote cumulative productivity effects from event year 0 (when there are top inventor inflows) to year j. Since the static model abstracts from the lead and lag effects, the baseline model may produce biased estimates of productivity gains.

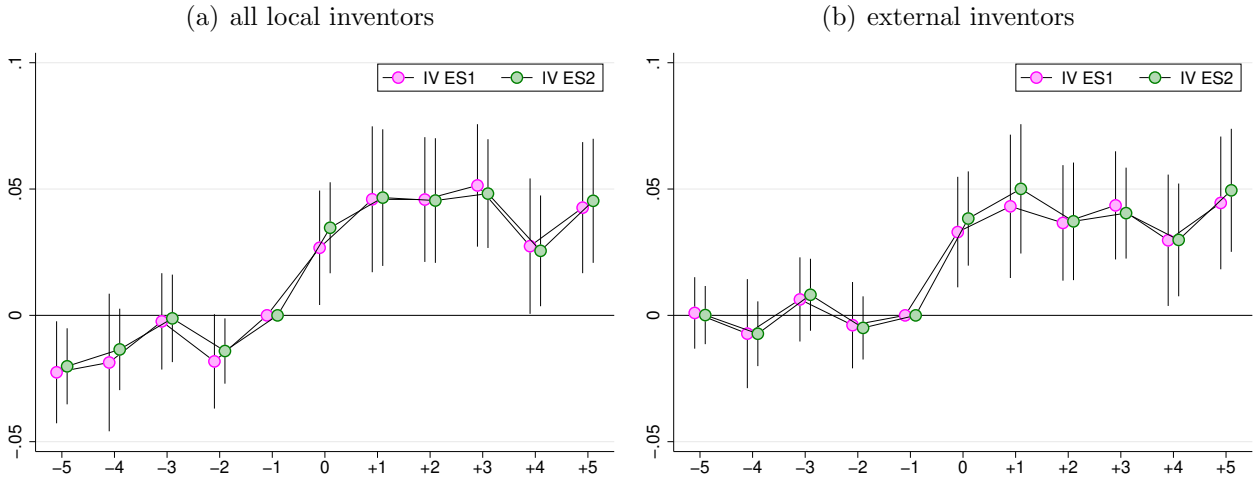
As in the static analysis, we incorporate the Bartik instruments into the event study model. Let $\boldsymbol{\Delta B}_{dt} = [\Delta B_{dt}^{(\underline{j})} \cdots \Delta B_{dt}^{(\overline{j})}]'$ denote a $(\underline{j} + \overline{j} + 1) \times 1$ vector of the first time difference of the IVs, where $\Delta B_{dt}^{(j)}$ is defined in a similar way as in (11). The IV event study model consists of the structural equation (10) and the first-stage regression analogous to (9) as follows[26]

$$\Delta M_{dt}^{(j)} = \boldsymbol{\psi}^{ef(j)} \boldsymbol{\Delta B}_{dt} + \xi^{ef(j)} X_{dt} + \varepsilon_{dt}^{ef(j)}, \tag{12}$$

[25] Unlike in standard event study models with a single treatment of identical intensity, we consider a more general case with multiple treatments of varying intensities. See Schmidheiny and Siegloch (2023) for the detailed classification of event study models.

[26] This robustness check abstracts from the possibility that treatment effects can be heterogeneous. Although several recent papers have explored under what conditions event study models provide valid average treatment effects in the presence of heterogeneous treatment effects (e.g., de Chaisemartin and D'Haultfœuille, 2023), they are not readily applicable to our IV event study model with multiple treatments of varying intensities.

Figure 6: IV event study regressions.



Notes: Panels (a) and (b) illustrate the dynamic impacts of a top inventor inflow on the patent productivity of all local inventors and that of external inventors, respectively. In each panel, IV ES1 uses B_{dt} and B^{σ}_{dt} as instruments and IV ES2 uses B_{dt}, B^{σ}_{dt}, and B^{ν}_{dt} as instruments. In Appendix G, we assess the robustness to possible violations of the parallel trends assumption in both Panels (a) and (b) using the method developed by Rambachan and Roth (2023) and confirm that such violations are unlikely.

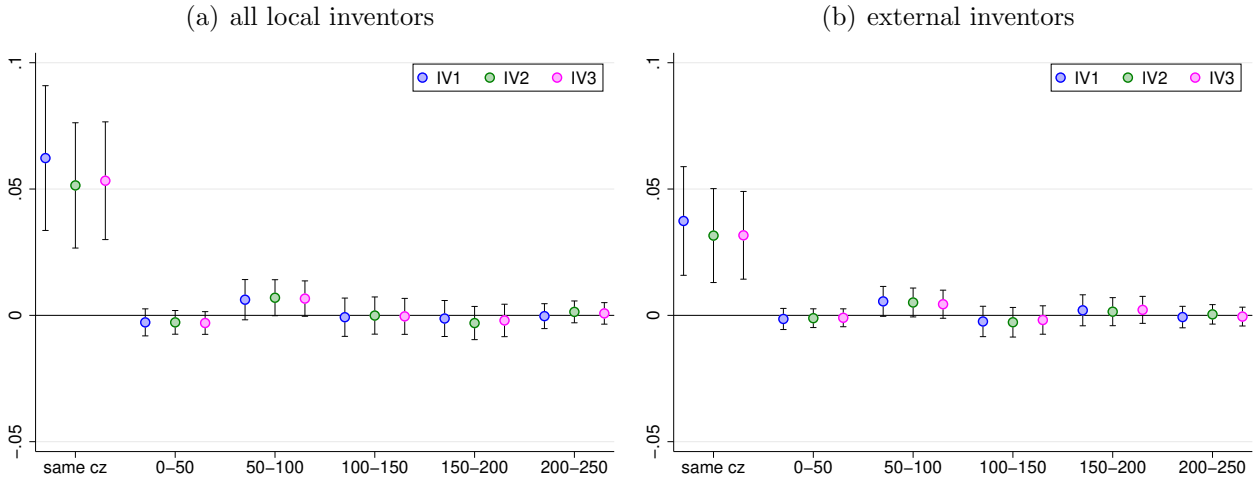
where $\boldsymbol{\psi}^{ef(j)} = [\psi^{ef(j,\underline{j})} \cdots \psi^{ef(j,\bar{j})}]$ is a vector of coefficients.[27]

Figure 6 shows the results for the IV event study regressions.[28] Panel (a) illustrates the dynamic impacts of a top inventor inflow on the patent productivity of all local inventors, which include not only the gains from internal knowledge sharing within the same assignee and between co-inventors but also the gains from external knowledge spillovers. Panel (b) corresponds to the dynamic productivity gains of external inventors, which go beyond organizational boundaries and co-inventor relationships. In both cases, we observe a substantial increase in local patent productivity in event year 0 when there are top inventor inflows. The post-event semi-elasticities go up to approximately 0.05, which ensures our main results in Section 4.4. In contrast, the pre-event semi-elasticities are close to zero in any pre-event year, thus suggesting no productivity gains prior to the event of top inventor migration. In Appendix G, we further assess the robustness to possible violations of the parallel trends assumption in both Panels (a) and (b) using the method developed by Rambachan and Roth (2023) and confirm that such violations are unlikely.

[27]When estimating the event study models with multiple instruments, we set $\boldsymbol{\Delta B}^{\sigma}_{dt} = [\Delta B^{\sigma(\underline{j})}_{dt} \cdots \Delta B^{\sigma(\bar{j})}_{dt}]'$ and $\boldsymbol{\Delta B}^{\nu}_{dt} = [\Delta B^{\nu(\underline{j})}_{dt} \cdots \Delta B^{\nu(\bar{j})}_{dt}]'$ and use $[\boldsymbol{\Delta B}_{dt}{}' \ \boldsymbol{\Delta B}^{\sigma}_{dt}{}']'$ or $[\boldsymbol{\Delta B}_{dt}{}' \ \boldsymbol{\Delta B}^{\sigma}_{dt}{}' \ \boldsymbol{\Delta B}^{\nu}_{dt}{}']'$ as instruments in (12).

[28]We report the numbers used in Figure 6 and the associated first-stage statistics in Tables C13 and C14 in Appendix C.5, respectively.

Figure 7: Distance-ring regression.



Notes: Panels (a) and (b) illustrate the impacts of a top inventor flow to each distance ring $r(d)$ on the patent productivity of all local inventors in d and that of external inventors in d, respectively, where d is the destination commuting zone. In each panel, IV1 uses B_{dt} as an instrument, IV2 uses B_{dt} and B^{σ}_{dt} as instruments, and IV3 uses B_{dt}, B^{σ}_{dt}, and B^{ν}_{dt} as instruments.

5.2 Geographic space

To check the robustness of our main results in terms of the geographic extent of productivity gains, we replace the structural equation (8) with

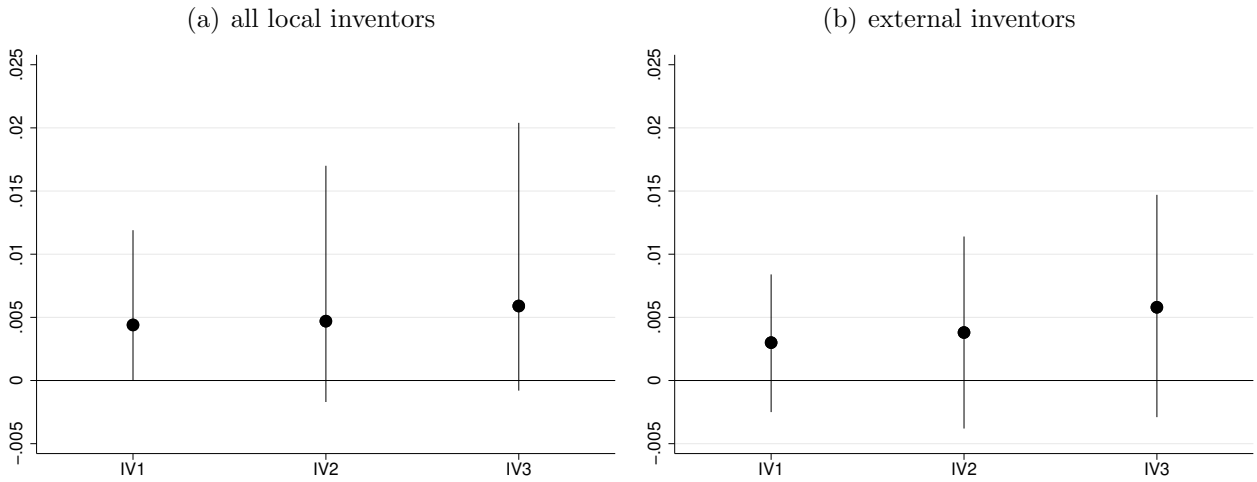
$$\ln Y_{dt} = \sum_{r(d)=1}^{6} \phi^{sr}_{r(d)} M_{r(d)t} + \xi^{sr} X_{dt} + \varepsilon^{sr}_{dt}, \tag{13}$$

where $r(d)$ is the distance ring defined for each destination commuting zone d and $M_{r(d)t}$ is the flows of top inventors in the $r(d)$-th ring. The first ring $r(d) = 1$ stands for destination d itself, whereas $r(d) = 2, ..., 6$ correspond to commuting zones that are 0-50, 50-100, 100-150, 150-200, and 200-250 miles away from commuting zone d.[29]

Figure 7 illustrates the estimated coefficients $\{\widehat{\phi}^{sr}_{r(d)}\}^{6}_{r(d)=1}$. Panels (a) and (b) illustrate the impacts of a top inventor flow to each distance ring $r(d)$ on the patent productivity of all local inventors in d and that of external inventors in d, respectively. In each panel, the impacts are significant only in the first ring for all three different IVs, which implies that top inventor inflows affect patent productivity only in the commuting zone where they enter. Such localized productivity gains are reminiscent of localized knowledge spillovers in Jaffe et al. (1993) and Murata et al. (2014). We will discuss the mechanism of localized productivity gains in terms of localized knowledge spillovers as evidenced by patent citations in Section 6.1.

[29]The distance between any pair of two commuting zones is calculated using the great circle formula.

Figure 8: Placebo.



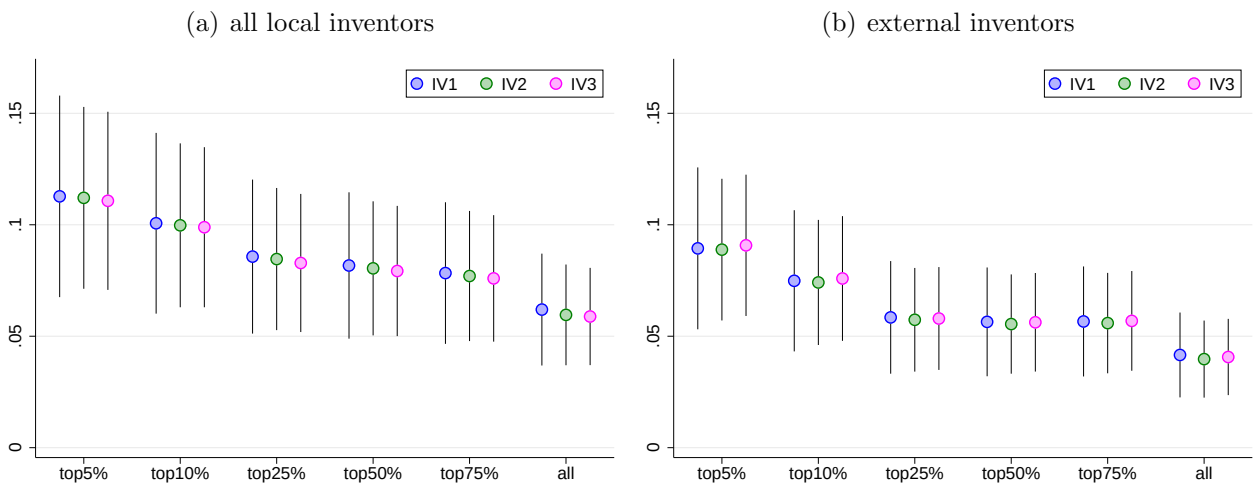
Notes: Panels (a) and (b) illustrate the impacts of a top inventor flow to d on the patent productivity of all local inventors in $R(d)$ and that of external inventors in $R(d)$, respectively, where d is the destination commuting zone and $R(d) \neq d$ is a commuting zone that is randomly drawn from state $\sigma(d)$ to which commuting zone d belongs. In each panel, IV1 uses B_{dt} as an instrument, IV2 uses B_{dt} and B^{σ}_{dt} as instruments, and IV3 uses B_{dt}, B^{σ}_{dt}, and B^{ν}_{dt} as instruments.

We also conduct a permutation-based placebo test to assess the plausibility of our findings that productivity gains are localized within each commuting zone. This is done by examining the impact of top inventor migration into commuting zone d on productivity gains in a randomly drawn commuting zone $R(d) \neq d$ in state $\sigma(d)$. We thus replace the structural equation (8) with

$$\ln Y_{R(d)t} = \phi^{sR} M_{dt} + \xi^{sR} X_{dt} + \varepsilon^{sR}_{dt} \tag{14}$$

and estimate (14) for each IV specification 1000 times with replacement to obtain the distribution of $\{\phi^{sR}_i\}^{1000}_{i=1}$. We then check if the null hypothesis of no productivity gains, $\phi^{sR} = 0$, is rejected. Figure 8 depicts the 95% confidence interval and the mean of the distribution for each IV specification. The results in both Panels (a) and (b) show that top inventor flows into commuting zone d do not significantly change patent productivity in commuting zone $R(d) \neq d$ randomly drawn from state $\sigma(d)$, thus implying that the extent of productivity gains is geographically limited within each commuting zone.

Figure 9: Productivity gains by local inventors' productivity.



Notes: Both panels illustrate the impacts of a top inventor inflow on local patent productivity, where the impacts differ by local inventors' patent productivity. Panels (a) and (b) are the results for all local inventors and for external inventors, respectively. In each panel, IV1 uses B_{dt} as an instrument, IV2 uses B_{dt} and B^{σ}_{dt} as instruments, and IV3 uses B_{dt}, B^{σ}_{dt}, and B^{ν}_{dt} as instruments.

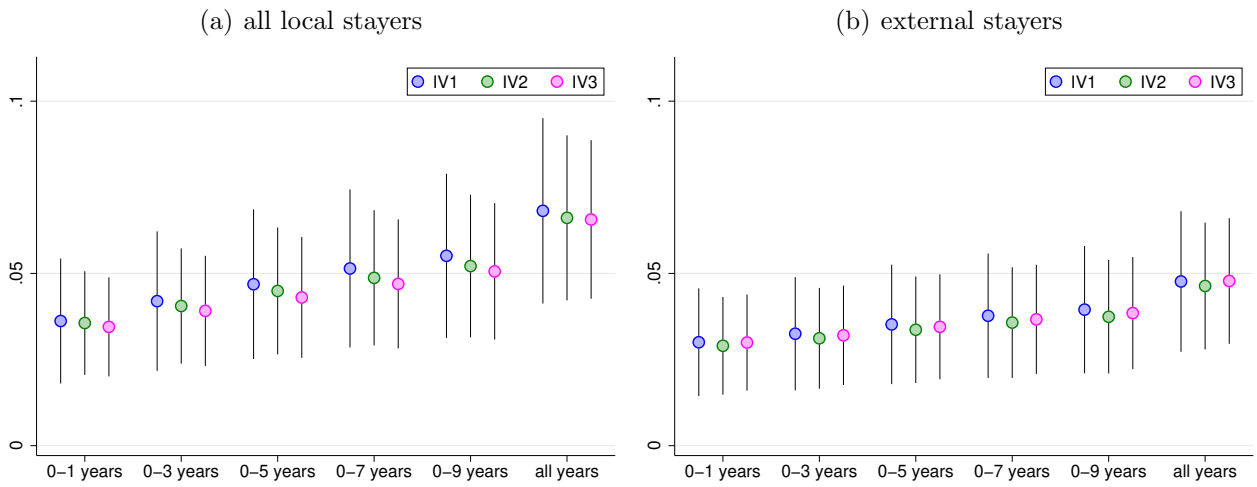
5.3 Productivity gains by local inventor types

We have thus far shown that top inventor inflows enhance patent productivity only for local inventors. We further explore the foregoing result by addressing who gain more from top inventor inflows by allowing for heterogeneity among local inventors.

First, local inventors differ in terms of their patent productivity. We thus consider the top 5%, 10%, 25%, 50%, and 75% of local inventors according to their patent productivity and estimate the causal effect for each productivity group. Panel (a) of Figure 9 illustrates the case with all local inventors. We observe that there are productivity gains for each productivity group and that more-productive local inventors tend to gain more from top inventor inflows. We find a similar pattern in Panel (b). Hence, even when focusing on external inventors who are not directly connected to the migrating top inventors, our result suggests that more-productive local inventors have a tendency to benefit more from their inflows.

Second, local inventors differ in terms of how many years they have stayed when top inventors arrive. To highlight this difference, we focus on local inventors who consistently apply for patents in a single commuting zone during the sample period and refer to them

Figure 10: Productivity gains by local stayers' duration.



Notes: Both panels illustrate the impacts of a top inventor inflow on the patent productivity of local stayers, where the impacts differ by their duration. Panels (a) and (b) are the results for all local stayers and for external stayers, respectively. In each panel, IV1 uses B_{dt} as an instrument, IV2 uses B_{dt} and B^{σ}_{dt} as instruments, and IV3 uses B_{dt}, B^{σ}_{dt}, and B^{ν}_{dt} as instruments.

as *local stayers*, who can be classified into *internal stayers* and *external stayers*.[30] We then disaggregate those local stayers by the length of their stay and estimate the causal effect for each duration. In each panel of Figure 10, the leftmost bars show the impacts of a top inventor inflow on local stayers with a duration of less than one year, which may be viewed as the entry of new local inventors ("extensive margins"). The next bars stand for up to three years of residency, up to five years of residency, and so on, thus capturing not only extensive margins but also the incumbent local inventors ("intensive margins"). The rightmost bars show the overall effect on local stayers, regardless of their duration. These results suggest that the incumbent local inventors tend to contribute more to the overall productivity gains from top inventor inflows when compared with the new local inventors. Recall the concern in Section 4.3, namely that other state taxes may also influence the inflows of non-top inventors, thereby contributing to local patent productivity. This concern is mitigated since the results in Figure 10, which excludes the role of such potential non-top inventor inflows in local patent productivity by definition of local stayers, are quite similar to our main results in Table 5.

[30]The number of local stayers is 1,195,335, which accounts for 93.81% of all 1,274,192 local inventors. Thus, in this robustness check, we exclude 78,857 local inventors (6.19% of all local inventors) who apply for patents in multiple commuting zones. The internal stayers are local stayers who share the same organization as the migrating top inventors and/or if they are co-inventors of the migrating top inventors. All the other local stayers are referred to as external stayers.

5.4 Falsification: The case of top baseball players

To provide further validation of our analysis, we conduct a falsification exercise using an alternative occupation. Following Kleven et al. (2013), we focus on high-income professional sports players who should also respond to tax differentials but should not affect patent productivity. In what follows, we examine to what extent the impact of top inventor inflows on local patent productivity is modified by the inflows of top Major League Baseball (MLB) players or "top players" for short (see Appendix H for the institutional settings and data sources).[31]

Let $\mathcal{I} = \{1, \ldots, I\}$, $\mathcal{J} = \{1, \ldots, J\}$, and $\mathcal{K} = \{1, \ldots, K\}$ denote the set of top players, that of origin teams, and that of destination teams, respectively. The utility of top player $i \in \mathcal{I}$, who played on team $j \in \mathcal{J}$ in period $t-1$ and chooses team $k \in \mathcal{K}$ in period t, is given by $U_{ijkt} = V_{ijkt} + \varepsilon_{ijkt}$, where $V_{ijkt} = \alpha \ln[(1-\tau_{\sigma(k)t})w_{ikt}] + \gamma^h \text{Home}_{ijkt-1} + \gamma^x_k X_{it} - \gamma^c C_{jk} + Z_k$, ε_{ijkt} is independent and identically Gumbel distributed, and $k = 1$ denotes retirement. Top player i's after-tax salary is given by $(1 - \tau_{\sigma(k)t})w_{ikt}$, where we proxy $\tau_{\sigma(k)t}$ by the ATR at the ninety-fifth percentile in state $\sigma(k)$ in year t as before.[32] Home_{ijkt-1}, which takes a value of 1 if $j = k$ and 0 otherwise, is the indicator that captures top player i's preferences for the team in the previous period. X_{it} is a vector of top player i's characteristics and performance. C_{jk} is the cost of migration measured by the geographical distance between teams j and k, and Z_k denotes team fixed effects. The predicted probability that top player i, who played on team j in period $t-1$, chooses team k in period t is then given by

$$\widehat{P}^{\text{ply}}_{ijkt} = \frac{\exp\{\widehat{\alpha} \ln[(1-\tau_{\sigma(k)t})w_{ikt}] + \widehat{\gamma}^h \text{Home}_{ijkt-1} + \widehat{\gamma}^x_k X_{it} - \widehat{\gamma}^c C_{jk} + \widehat{Z}_k\}}{\sum_{k'=1}^{K} \exp\{\widehat{\alpha} \ln[(1-\tau_{\sigma(k')t})w_{ik't}] + \widehat{\gamma}^h \text{Home}_{ijk't-1} + \widehat{\gamma}^x_{k'} X_{it} - \widehat{\gamma}^c C_{jk'} + \widehat{Z}_{k'}\}}. \quad (15)$$

Like top inventors, we find that top players respond to tax differences within a country (see Table H2 in Appendix H), thus showing that the tax-induced international mobility of top players in Kleven et al. (2013) can be applied to a domestic context. These individual-level

[31] We define "top players" as those players whose salaries exceeded the ninety-fifth percentile of the U.S. income distribution in the year they declared free agency after six years of MLB service.

[32] Two remarks are in order. First, since salary data are available only for a subset of all player-year observations, we use a random forest algorithm to estimate salaries based on various player characteristics and performance. The resulting R-squared values are 0.975 for the training set and 0.813 for the test set. Second, since we observe actual salaries only for chosen teams, we need to construct unobserved potential salaries for other teams. Following Kleven et al. (2013), we consider several formulations for counterfactual salaries in the context of professional sports labor markets (see Appendix H).

predicted probabilities are then aggregated to form the predicted top player flows from origin commuting zone o to destination commuting zone d as $\widehat{P}^{\text{ply}}_{odt} = \sum_{k\in\mathcal{K}^d}\sum_{j\in\mathcal{J}^o}\sum_{i\in\mathcal{I}^j}\widehat{P}^{\text{ply}}_{ijkt}$, where $\mathcal{I}^j$ represents the set of top players on team j in period $t-1$; and $\mathcal{J}^o$ and $\mathcal{K}^d$ denote the set of teams located in origin o and destination d, respectively. Letting M^{ply}_{odt} and I^{ply}_{ot} denote the number of top players who migrate from o to d in period t and the number of top players located in origin o in the year prior to t, respectively, we define the Bartik instrument for $M^{\text{ply}}_{dt} = \sum_{o\neq d} M^{\text{ply}}_{odt}$ as $B^{\text{ply}}_{dt} = \sum_{o\neq d}\widehat{P}^{\text{ply}}_{odt} I^{\text{ply}}_{ot}$ to estimate the structural equation as follows:

$$\ln Y_{dt} = \phi^{sF} M_{dt} + \check{\phi}^{sF} M^{\text{ply}}_{dt} + \xi^{sF} X_{dt} + \varepsilon^{sF}_{dt}. \tag{16}$$

Table 6: The impact of top inventor and top player inflows on local patent productivity.

	(1)	(2)	(3)	(4)	(5)	(6)
(a) All local inventors						
Top inventor inflows	0.070	0.065	0.064	0.076	0.062	0.063
	(0.018)	(0.015)	(0.015)	(0.021)	(0.016)	(0.016)
Top player inflows	-0.040	-0.029	-0.029	-0.048	-0.017	-0.017
	(0.032)	(0.027)	(0.027)	(0.036)	(0.027)	(0.027)
$\ln(1-\text{ATR})$	5.946	5.919	6.026			
	(1.041)	(1.038)	(1.039)			
(b) External inventors						
Top inventor inflows	0.049	0.045	0.047	0.050	0.039	0.041
	(0.014)	(0.012)	(0.012)	(0.017)	(0.013)	(0.013)
Top player inflows	-0.037	-0.028	-0.031	-0.040	-0.018	-0.018
	(0.026)	(0.022)	(0.022)	(0.029)	(0.021)	(0.022)
$\ln(1-\text{ATR})$	4.713	4.662	4.627			
	(0.850)	(0.847)	(0.842)			
CZ FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	No	No	No
State × year FE	No	No	No	Yes	Yes	Yes
Observations	23,628	23,628	23,463	23,562	23,562	23,397

Notes: The coefficients on top inventor inflows and top player inflows are converted to semi-elasticities. ATR stands for the individual income average tax rate at the ninety-fifth percentile. The coefficient on $\ln(1-\text{ATR})$ is converted to elasticity. Columns 1 and 4 use B_{dt} as an instrument for top inventor inflows M_{dt}. Columns 2 and 5 use B_{dt} and B^{σ}_{dt} as instruments for M_{dt}. Columns 3 and 6 use B_{dt}, B^{σ}_{dt}, and B^{ν}_{dt} as instruments for M_{dt}. In all cases, B^{ply}_{dt} is used as an instrument for top player inflows M^{ply}_{dt}. Columns 4-6 replace $\ln(1-\text{ATR})$ in Columns 1-3 with state × year FE. Cluster-robust standard errors are in parentheses.

Table 6 presents the results for different specifications.[33] Columns 1-6 show that even in the presence of top player inflows, the impact of top inventor inflows on both all local inventors and external inventors remain consistent with our baseline estimates in Table 5, thus suggesting

[33]We report the associated first-stage statistics in Table H3.

the robustness of our main results. Indeed, the top player inflows have a negligible effect on local patent productivity, indicating that while high-income top players may respond to tax differences, their presence does not significantly alter the geographic distribution of inventive activity across commuting zones. Hence, we may conclude that although both top inventors and top players take into account the variation in tax rates when making their location decisions, the impacts of their inflows on local patent productivity differ substantially.

5.5 Shift exogeneity

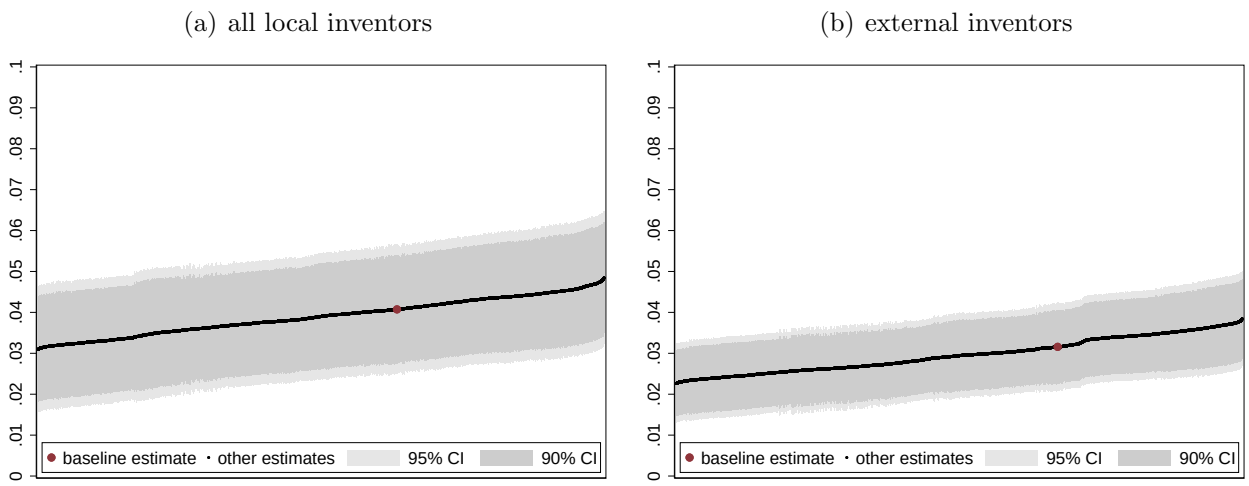
We have so far considered the share exogeneity of the Bartik instruments as in Goldsmith-Pinkham et al. (2020). We now conduct robustness checks from the shifts perspective by reformulating the destination specific structural equation (8) in terms of origin commuting zones (see Appendix I). Borusyak et al. (2022) show that the estimator for ϕ^s obtained from this reformulation converges in probability to the productivity effect ϕ^s in the shares perspective equation (8). Moreover, the standard errors computed in this framework are valid in the presence of exposure-based clustering, as demonstrated by Adão et al. (2019).

Table 7: The impact of top inventor inflows on local patent productivity (shift exogeneity).

	(1)	(2)
(a) All local inventors		
Top inventor inflows	0.041	0.039
	(0.008)	(0.008)
First-stage F statistic	84.781	75.042
(b) External inventors		
Top inventor inflows	0.032	0.028
	(0.005)	(0.006)
First-stage F statistic	84.781	75.042
Destination controls		
$\ln(1 - \text{ATR})$	Yes	No
CZ FE and year FE	Yes	Yes
State × year FE	No	Yes
Observations	2,760	2,760
Number of origin CZs	292	292

Notes: This table reports the estimates $\widehat{\phi}^s$ of the impact of a top inventor inflow on local patent productivity in destination commuting zone d, where we obtain these estimates from equivalent IV regressions regarding origin commuting zone o, instrumented by I_{ot}. The coefficient on top inventor inflows is converted to semi-elasticity. Panel (a) presents the patent productivity gains for all local inventors, whereas Panel (b) focuses on those for external inventors. Column 1 controls for $\ln(1 - \text{ATR})$, commuting zone fixed effects, and year fixed effects. Column 2 replaces $\ln(1 - \text{ATR})$ in Column 1 with state × year FE. Exposure-robust standard errors, clustered at the commuting zone level, are reported in parentheses. All the first-stage F-statistics, which are derived from the equivalent IV regressions instrumented by I_{ot}, exceed 10, thus addressing concerns regarding weak instruments.

Figure 11: Specification curve analysis (shift exogeneity).



Notes: Panels (a) and (b) illustrate the impacts of a top inventor inflow on the patent productivity of all local inventors and that of external inventors, respectively. Each specification curve is depicted using 12,288 alternative specifications, as explained in footnote 34. The vertical axis is the value of $\widehat{\phi}^s$.

Table 7 presents the estimates $\widehat{\phi}^s$ of the impact of a top inventor inflow on local patent productivity in destination commuting zone d, where we obtain these estimates from equivalent IV regressions regarding origin commuting zone o, instrumented by I_{ot}. Panel (a) reports the patent productivity gains for all local inventors, whereas Panel (b) focuses on those for external inventors. Column 1 controls for $\ln(1-\text{ATR})$, commuting zone fixed effects, and year fixed effects, and Column 2 replaces $\ln(1-\text{ATR})$ in Column 1 with state $\times$ year fixed effects.

Reassuringly, the estimates under this alternative identification strategy are not substantially different from our main estimates in Table 5. To assess the robustness of our findings obtained from the shift exogeneity assumption, we conduct a specification curve analysis as in Section 4.4. We employ different specifications of the IV regressions by considering various dimensions.[34] Figure 11 plots the specification curve for $\widehat{\phi}^s$ with 90% and 95% confidence intervals. As seen from Panels (a) and (b), the productivity gains reported in Table 7 are fairly

[34]We consider (i) whether to use the ATR at the ninety-fifth or ninety-ninth percentile; (ii) whether to use $\ln(1+Y_{dt})$ or drop commuting zones with $Y_{dt}=0$; (iii) whether to use the baseline or alternative detection of top inventor migrations; (iv) whether to use the baseline or alternative definition of local inventors (see the analysis of local stayers in Section 5.3); (v) whether to use ATRs or MTRs; (vi) whether to include APTRs; and (vii) whether to include each of the other controls (ATR50, CITR, ITC, RTC, manufacturing employment, and other employment variables such as "finance and insurance," "professional, scientific, and technical services," and "management of companies and enterprises" to capture other omitted variables, e.g., access to venture capital). Since the usual caveat on weak instruments is applicable here, we adopt only specifications for which the null hypothesis of weak instruments is rejected.

robust for 12,288 alternative specifications, thus mitigating concerns about the sensitivity of our main results to different exogeneity assumptions and various choices of control variables.

6 Mechanisms

The productivity gains estimated in the previous sections suggest that local inventors acquire knowledge from migrating top inventors, regardless of whether local inventors are internal or external. We now discuss the underlying mechanisms through which those productivity gains materialize. We first focus on patent citations that have been widely used as proxy for knowledge flows since Jaffe et al. (1993). Specifically, we count how many times local inventors cite incoming top inventors and estimate the percentage change in the number of citations caused by top inventor inflows. Furthermore, Jaffe et al. (1993) recognize the existence of other knowledge flows that cannot be captured by patent citations. We thus complement the foregoing analysis by using state-year variation in legal protection of trade secrets documented by Png (2017a, 2017b) as a quasi-natural experiment. We expect that knowledge flows from migrating top inventors to local external inventors would be greater in states where legal protection of trade secrets is weaker, so that there would be additional local productivity gains in those states. Our results presented below are consistent with Marshall's insight on knowledge spillovers since external inventors can not only learn patentable knowledge but also obtain other forms of knowledge from migrating top inventors as if those were in the air.

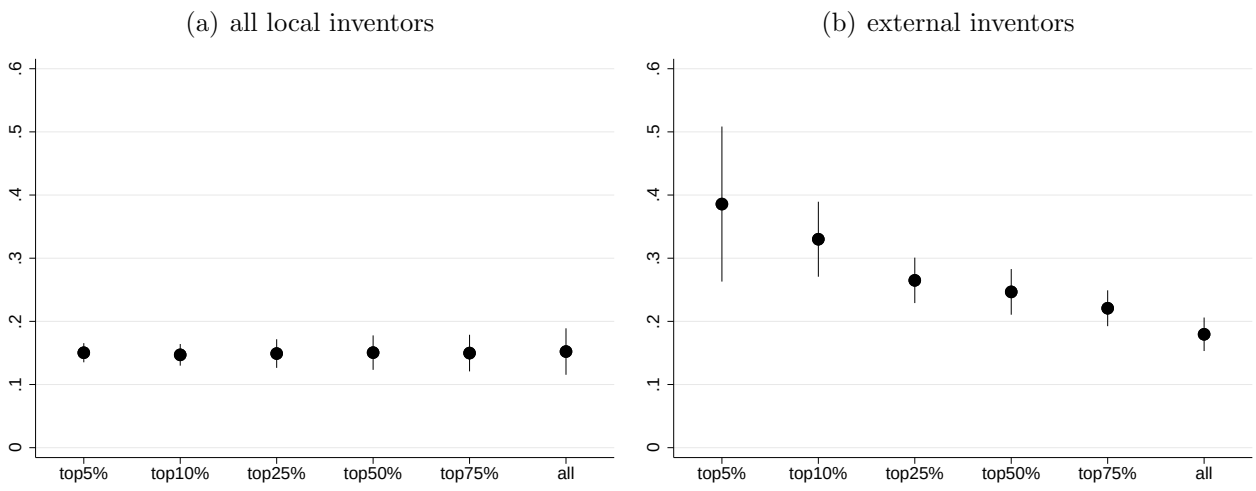
6.1 Patent citations

To see the impact of top inventor inflows on local patent citations, we count how many times the patents of the top inventors who migrated into commuting zone d in year t were cited by the local inventors in commuting zone d in year t and denote it by C_{dt}. When constructing C_{dt}, we focus on the patents that had been applied over the last ten years. Replacing patent productivity Y_{dt} in (8) with the number of citations C_{dt}, we consider the structural equation for citations as follows:

$$\ln C_{dt} = \phi^{sc} M_{dt} + \xi^{sc} X_{dt} + \varepsilon^{sc}_{dt},$$

while retaining the same first-stage equation (9). The coefficient ϕ^{sc} gauges the magnitude of knowledge flows from migrating top inventors to local inventors.

Figure 12: Citations.



Notes: Panels (a) and (b) illustrate the impact of a top inventor inflow on the number of all local inventors' citations to the incoming top inventor and that of external inventors' citations to the incoming top inventor (percentage change in decimal form), respectively. In each panel, we use B_{dt} and B_{dt}^{σ} as instruments.

Panels (a) and (b) of Figure 12 illustrate the estimated coefficients for the case of all local inventors and that of external inventors, respectively. In both panels, we consider the top 5%, 10%, 25%, 50%, and 75% of local inventors according to their patent productivity and estimate the causal effect for each productivity group. In Panel (a), an additional top inventor inflow raises the number of local inventors' citations to the incoming top inventor by 10-20% regardless of the productivity of local inventors. By contrast, in Panel (b), the external inventors, especially those with higher productivity, tend to exhibit a greater percentage change in the number of citations to the top inventors who moved in the same commuting zone. These results imply the existence of knowledge flows from the migrating top inventors to the local inventors, even when we focus on the external inventors who are not directly connected to the migrating top inventors.[35] This existence of external knowledge spillovers as evidenced by the citation flow from each incoming top inventor to each external inventor is related to but differs from Atkin et al. (2022), who find that face-to-face meetings between workers in different establishments enhance between-establishment citations, as the latter aim to capture worker interactions by smartphone data while abstracting from who cites whom.

[35]It is perhaps puzzling that the impact is smaller in Panel (a). This result may stem from the possibility that internal inventors, who account for approximately 40% of all local inventors, had already collaborated with or worked in the same organization as the incoming top inventors and thus had already cited them prior to their migration. In that case, we would expect a smaller percentage change in the number of internal inventors' citations after the arrival of the top inventors.

6.2 Trade secrets

Trade secrets were formerly defined and protected from misappropriation by common law in the United States. However, these definitions and protections have been codified into law with the enactment of federal legislation known as the Uniform Trade Secrets Act (UTSA). While most states had already adopted the UTSA, there had been substantial heterogeneity in the states' approaches to trade secrets due to the differences in the timing of the adoption of the UTSA and the strength of trade secrets protection during the common law era. We exploit the state-year variation in trade secrets protection to uncover productivity gains through knowledge flows that cannot be captured by patent citations.

Given the heterogeneity in legal protection of trade secrets, top inventors who migrate to a state with weaker protection would exchange knowledge more frequently with other inventors beyond organizational boundaries and co-inventor relationships, thereby bringing about additional productivity gains to local external inventors. In contrast, the change in legal protection of trade secrets would not influence knowledge sharing through organizations or co-inventor relationships, thus leaving the productivity gains of local internal inventors unaffected.

We examine those differential impacts of top inventor inflows on local patent productivity by using the state-level index of trade secrets in Png (2017a, 2017b).[36] This index captures both legal protection under common law and the enactment of the UTSA and ranges between 0 and 1, where a higher score implies stronger legal protection. Let $S_{\sigma(d)t}$ denote an indicator variable for whether the trade secrets index in state $\sigma(d)$ in year t is below the median of the trade secrets index distribution. If $S_{\sigma(d)t} = 1$, the degree of trade secrets protection is low in commuting zone d in year t, so that we expect higher patent productivity of external inventors due to a greater amount of knowledge brought about by top inventor inflows.

Let $\boldsymbol{\phi}^s = [\phi_0^s \ \phi_1^s \ \phi_2^s]$ and $E_{dt} = [M_{dt} \ S_{\sigma(d)t} \ M_{dt}S_{\sigma(d)t}]'$ denote a vector of coefficients and a vector of endogenous variables. The structural equation is then given by

$$\ln Y_{dt} = \boldsymbol{\phi}^s E_{dt} + \xi^s X_{dt} + \varepsilon_{dt}^s. \tag{17}$$

Our interest is in the coefficient ϕ_2^s on $M_{dt}S_{\sigma(d)t}$ when Y_{dt} is measured by the patent pro-

[36] Png (2017a) provides the index for the years 1979 to 1998, and Png (2017b) extends it to the years 1970 to 2010.

ductivity of external inventors. If $\phi_2^s > 0$, the top inventor inflows lead to higher patent productivity of external inventors in commuting zones with weaker trade secrets protection, which suggests that the knowledge brought about by the top inventor inflows is more likely to spill over to external inventors in commuting zones with weaker trade secrets protection.

As before, we use the Bartik instruments for the top inventor inflows M_{dt} to mitigate the endogeneity concern. To address the potential endogeneity of the trade secrets indicator $S_{\sigma(d)t}$, we follow Png (2017b) who argues that the enactment of the UTSA is related to the enactment of other state-level uniform laws such as the Uniform Determination of Death Act (UDDA), Uniform Federal Lien Registration Act (UFLRA), and Uniform Fraudulent Transfer Act (UFTA) because these laws were introduced to harmonize state laws. Since the three laws are unlikely to be associated with local patent productivity shocks, we use them as instruments for the trade secrets indicator. The first-stage equation that accompanies (17) is thus given by

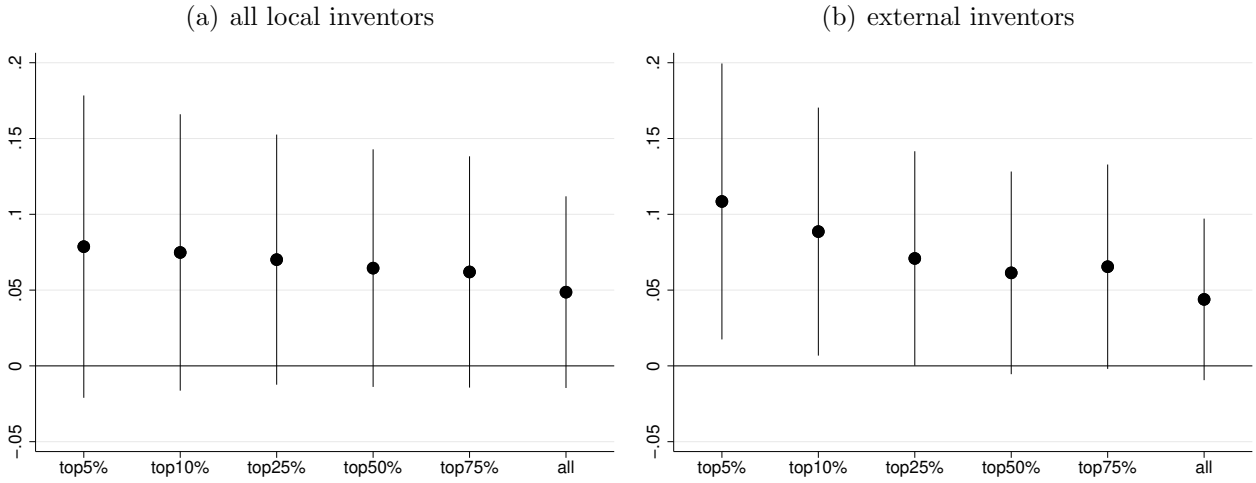
$$E_{dt} = \boldsymbol{\psi}^f Z_{dt} + \xi^f X_{dt} + \varepsilon_{dt}^f, \tag{18}$$

where $\boldsymbol{\psi}^f$ is a vector of coefficients and $Z_{dt} = [(B_{dt}, B_{dt}^{\sigma})\#\#(U_{\sigma(d)t}^{\text{DDA}}, U_{\sigma(d)t}^{\text{FLRA}}, U_{\sigma(d)t}^{\text{FTA}})]'$ is a vector of IVs, with $U_{\sigma(d)t}^{\ell}$ indicating whether uniform law $\ell = \{\text{DDA}, \text{FLRA}, \text{FTA}\}$ was in effect in state $\sigma(d)$ in year t, and $\#\#$ denotes an interaction-term operator.[37]

Panels (a) and (b) in Figure 13 illustrate the results for all local inventors and those for external inventors, respectively. In both panels, we consider the top 5%, 10%, 25%, 50%, and 75% of local inventors according to their patent productivity and estimate the causal effect for each productivity group. Since the strength of trade secrets protection is unlikely to affect internal knowledge sharing within the same assignee and between co-inventors, it is not surprising that the overall impact in Panel (a) is insignificant, regardless of the productivity of local inventors. In contrast, in Panel (b) the impacts for the top 5%, 10%, and 25% of inventors (top 50% and 75% of inventors) are significant at the 5% (10 %) level. Hence, top inventor migration tends to enhance the patent productivity of external inventors in commuting zones with weaker trade secrets protection.

[37]The interaction-term operator $\#\#$ generates all possible combinations of elements for a given pair of sets. For example, let $\mathfrak{S}_1 = \{A, B\}$ and $\mathfrak{S}_2 = \{C, D\}$, where each set $\mathfrak{S}_i$ has two elements. Then, $[\mathfrak{S}_1\#\#\mathfrak{S}_2]' = [\{A, B\}\#\#\{C, D\}]' = [A\ B\ C\ D\ AC\ AD\ BC\ BD]'$.

Figure 13: Trade secrets.



Notes: Panels (a) and (b) illustrate the coefficients ϕ_2^s on the interaction term $M_{dt}S_{\sigma(d)t}$ in (17) for all local inventors and for external inventors, respectively. In each panel, we use B_{dt} and B_{dt}^{σ} as instruments.

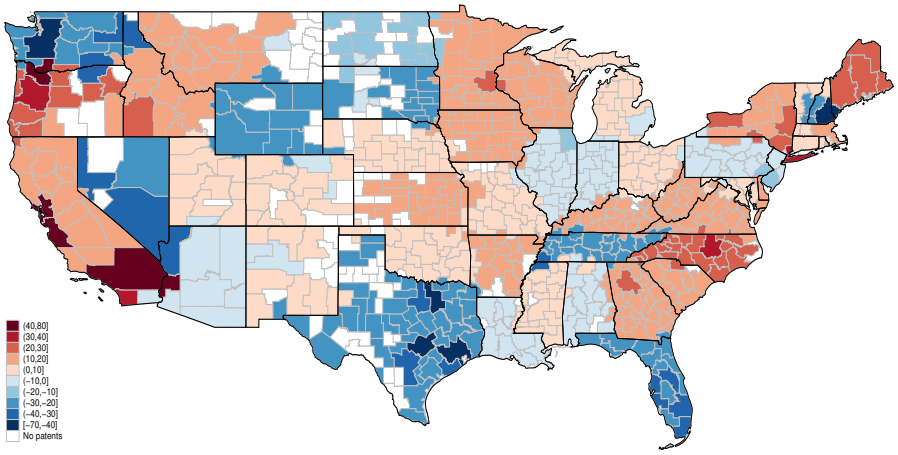
7 A counterfactual experiment

We now conduct a counterfactual experiment. Using the baseline specification in Section 4, we consider what happens to the geographic distribution of patent productivity if all state individual income taxes are set to their average. This experiment is useful for assessing to what extent state tax differences contribute to patent productivity differences across space.

Recalling that the changes in state taxes affect the choice probabilities $\widehat{P}_{odt}$ in (5) and the Bartik instruments B_{dt} in (6), as well as B_{dt}^{σ}, the procedure of the counterfactual analysis can be summarized as follows. We first derive the counterfactual probabilities $\widetilde{P}_{odt}$ to construct the counterfactual Bartik instruments $\{\widetilde{B}_{dt}, \widetilde{B}_{dt}^{\sigma}\}$, which allow us to estimate the counterfactual top inventor flows $\widetilde{M}_{dt}$ via the first-stage regression (9). We then define the counterfactual changes in the top inventor flows as $\widetilde{\Delta}M_{dt} = \left(\frac{\widetilde{M}_{dt}-\widehat{M}_{dt}}{\widehat{M}_{dt}}\right) M_{dt}$, where M_{dt}, $\widehat{M}_{dt}$, and $\widetilde{M}_{dt}$ are the actual, fitted, and counterfactual flows, respectively.[38] We finally apply $\widetilde{\Delta}M_{dt}$ to the structural equation (8) to construct the counterfactual changes in the log patent productivity $\widetilde{\Delta}\ln Y_{dt} = \widehat{\phi}^s\widetilde{\Delta}M_{dt} + \widehat{\xi}^s\widetilde{\Delta}\ln(1-\mathrm{ATR}_{\sigma(d)t})$, where $\widetilde{\Delta}\ln(1-\mathrm{ATR}_{\sigma(d)t}) = \ln(1-\widetilde{\mathrm{ATR}}_{\sigma(d)t}) - \ln(1-\mathrm{ATR}_{\sigma(d)t})$

[38] If the actual and fitted flows coincide, the definition reduces to $\widetilde{\Delta}M_{dt} = \widetilde{M}_{dt} - M_{dt}$. However, since the actual and fitted flows generally differ, we compute the percentage change in the top inventor flows $\left(\frac{\widetilde{M}_{dt}-\widehat{M}_{dt}}{\widehat{M}_{dt}}\right)$ based on the fitted and counterfactual flows in the tax-induced migration model and then multiply it by the actual flows M_{dt}.

Figure 14: Counterfactual experiment (setting state taxes to their average).



Notes: This figure illustrates the counterfactual percentage change in the number of patents at the commuting zone level when all state individual income taxes in 2009 are set to their average. We use B_{dt} and B_{dt}^{σ} as instruments.

captures the counterfactual tax changes.[39] Thus, the overall impact of tax changes on $\widetilde{\Delta} \ln Y_{dt}$ can be decomposed into two: the direct effect from the tax changes, $\widehat{\xi}^{s} \widetilde{\Delta} \ln(1 - \mathrm{ATR}_{\sigma(d)t})$, and the indirect effect via the changes in top inventor flows induced by the tax changes, $\widehat{\phi}^{s} \widetilde{\Delta} M_{dt}$. The indirect effect can be further decomposed into two: productivity gains due to internal knowledge sharing and those due to external knowledge spillovers.

Figure 14 illustrates the percentage change in local patent productivity when state taxes are set to their average. The overall impact tends to be large in commuting zones in California, Oregon, North Carolina, and New York, where state taxes and initial patent productivity are high.[40] Table 8 summarizes the top 10 commuting zones by patent productivity gains. For instance, if state taxes were equal, the number of patents in Santa Clara–Monterey–Santa Cruz (which is the commuting zone with the highest patent productivity in Table 2) would be larger by 72.3%. In contrast, the overall impact tends to be small in commuting zones in Texas, Washington, Florida, and New Hampshire, where state taxes are low and initial patent productivity is high. Table 9 summarizes the bottom 10 commuting zones by patent productivity gains. For instance, if state taxes were equal, the number of patents in King–Pierce–Snohomish (which is ranked as the tenth most productive commuting zone in Table 2)

[39] When computing the counterfactual change, we replace $\widetilde{\Delta} \ln Y_{dt}$ with $\widetilde{\Delta} \ln(1 + Y_{dt})$ as before to accommodate commuting zone × year observations with no patents.

[40] Notably, the counterfactual changes are heterogeneous even within states, although we equalize taxes between states. The reason is that the counterfactual choice probabilities $\widetilde{P}_{odt}$, which are obtained by setting state taxes equal in (5), include fixed effects at the commuting zone level.

Table 8: Top 10 commuting zones by patent productivity gains (%).

rank	cz number	counties	state	gains (%)
1	37500	Santa Clara–Monterey–Santa Cruz	CA	72.291
2	37800	Alameda–Contra Costa–San Francisco	CA	53.377
3	38300	Los Angeles–Orange–San Bernardino	CA	47.115
4	38801	Multnomah–Washington–Clackamas	OR	46.022
5	38000	San Diego	CA	38.042
6	1701	Wake–Durham–Orange	NC	35.253
7	19400	Kings–Queens–New York	NY	35.080
8	38901	Lane–Marion–Linn	OR	31.310
9	35801	Ada–Canyon–Elmore	ID	29.754
10	39203	Deschutes–Crook–Jefferson	OR	29.206

Notes: Patent productivity gains are defined as the percentage change in the number of patents when all state individual income taxes in 2009 are set to their average. We use B_{dt} and B_{dt}^{σ} as instruments.

Table 9: Bottom 10 commuting zones by patent productivity gains (%).

rank	cz number	counties	state	gains (%)
1	39400	King–Pierce–Snohomish	WA	-64.777
2	31201	Travis–Williamson–Hays	TX	-50.710
3	32000	Harris–Fort Bend–Galveston	TX	-49.002
4	33100	Dallas–Denton–Collin	TX	-46.491
5	20600	Hillsborough–Rockingham–York	NH	-41.326
6	7100	Palm Beach–St. Lucie–Martin	FL	-38.194
7	7400	Orange–Seminole–Lake	FL	-35.219
8	5202	Shelby–DeSoto–Tipton	TN	-35.020
9	6900	Sarasota–Manatee–Charlotte	FL	-34.585
10	7000	Dade–Broward–Monroe	FL	-34.561

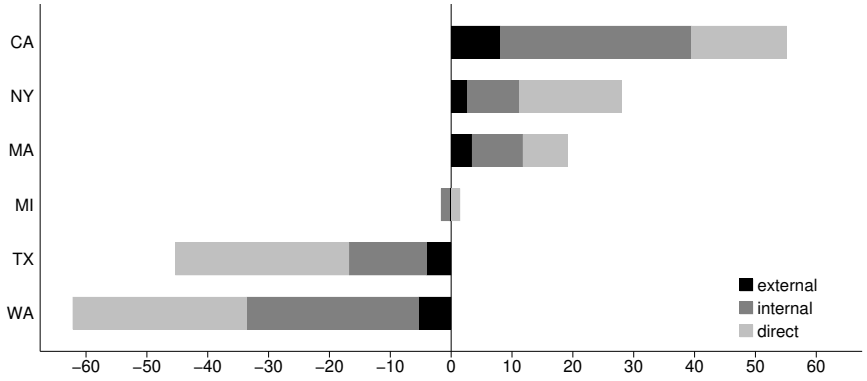
Notes: Patent productivity gains are defined as the percentage change in the number of patents when all state individual income taxes in 2009 are set to their average. We use B_{dt} and B_{dt}^{σ} as instruments.

would be smaller by 64.8%. These results suggest that the presence of state tax differences significantly distorts the spatial distribution of inventive activity.

To see which states are most affected by the presence of tax differences, we first define, for each commuting zone d, the counterfactual change in the number of patents $\widetilde{\Delta}Y_{dt} = \left(\frac{\widetilde{Y}_{dt}-\widehat{Y}_{dt}}{\widehat{Y}_{dt}}\right) Y_{dt}$ in the same way as $\widetilde{\Delta}M_{dt}$, where Y_{dt}, $\widehat{Y}_{dt}$, and $\widetilde{Y}_{dt}$ are the actual, fitted, and counterfactual numbers of patents in d, respectively. We then aggregate $\widetilde{\Delta}Y_{dt}$ within each state σ to obtain the counterfactual changes in the number of patents $\widetilde{\Delta}Y_{\sigma t} = \sum_{d\in\sigma} \widetilde{\Delta}Y_{dt}$. Denoting by $Y_{\sigma t} = \sum_{d\in\sigma} Y_{dt}$ the actual number of patents at the state level, we finally compute the percentage change in the number of patents at the state level $\frac{\widetilde{\Delta}Y_{\sigma t}}{Y_{\sigma t}}$.[41]

[41] Observe that $\frac{\widetilde{\Delta}Y_{\sigma t}}{Y_{\sigma t}} = \frac{\sum_{d\in\sigma}\widetilde{\Delta}Y_{dt}}{\sum_{d\in\sigma}Y_{dt}} = \sum_{d\in\sigma}\left(\frac{\widetilde{Y}_{dt}-\widehat{Y}_{dt}}{\widehat{Y}_{dt}}\right)\frac{Y_{dt}}{\sum_{d\in\sigma}Y_{dt}}$. The latter is the weighted average of the percentage change $\left(\frac{\widetilde{Y}_{dt}-\widehat{Y}_{dt}}{\widehat{Y}_{dt}}\right)$ with weight being the share of actual number of patents $\frac{Y_{dt}}{\sum_{d\in\sigma}Y_{dt}}$. If the actual and fitted numbers of patents coincide, the percentage change in the number of patents at the state level reduces to $\frac{\widetilde{\Delta}Y_{\sigma t}}{Y_{\sigma t}} = \frac{\sum_{d\in\sigma}\widetilde{Y}_{dt}-\sum_{d\in\sigma}Y_{dt}}{\sum_{d\in\sigma}Y_{dt}}$.

Figure 15: Counterfactual experiment (setting state taxes to their average).



Notes: This figure illustrates the counterfactual percentage change in the number of patents for the states of California, New York, Massachusetts, Michigan, Texas, and Washington when all state individual income taxes in 2009 are set to their average. The overall change for each state is decomposed into the direct effect via the change in state taxes and the indirect effect via the tax-induced top inventor migration. The latter consists of the internal knowledge sharing effect and the external knowledge spillover effect. We use B_{dt} and B_{dt}^{σ} as instruments.

Figure 15 illustrates the percentage change in the number of patents for selected states when state taxes are set to their average. For instance, if state taxes were equal, the number of patents in California (where state taxes and patent productivity are high) would be greater by 55.1%, which can be decomposed into the direct effect via the reduction in California state taxes (15.6%) and the indirect effect via the tax-induced top inventor migration (39.5%). The indirect effect can be further decomposed into the internal sharing effect (31.4%) and the external spillover effect (8.1%). In contrast, the number of patents in Texas (where state taxes are low and patent productivity is high) would be smaller by 45.3%, which can be decomposed into the direct effect via the rise in Texas state taxes (28.5%) and the indirect effect via the tax-induced migration (16.8%). The indirect effect can be further decomposed into the internal sharing effect (12.7%) and the external spillover effect (4.1%).

These results suggest that the indirect effect via the tax-induced migration of top inventors can be substantial. To see the relative importance of the direct and indirect effects on patent productivity at the national level, we aggregate those changes in the number of patents across all commuting zones in all states that we consider in the paper. We find that the share of the indirect effect is 0.725, while that of the direct effect is 0.275. Our results thus complement Akcigit et al. (2022) who assess the direct impact of state taxes on innovation.

8 Concluding remarks

In this paper, we have uncovered the idea-generating process described by Marshall (1890) using Bartik (1991) instruments. We have identified a significant causal effect of a top inventor inflow on the patent productivity of all local inventors. Even when we focus on local external inventors who are not directly connected to incoming top inventors through organizations or co-inventor relationships, the effect remains significant and is approximately 4%, thus implying that the mysteries of the trade are in the air.

We have disentangled productivity gains due to external knowledge spillovers from those due to internal knowledge sharing. Thus, our findings are consistent with the partially nonexcludable good nature of knowledge, whose implications have been explored theoretically in the technology and growth literature. Since the existence of the productivity gains from external knowledge spillovers leads to market failures and constitutes a rationale for spatial agglomeration of inventive activity, our analysis would be useful for innovation policies that consider both the benefits and costs of entrepreneurial clusters.

Our counterfactual experiment suggests that the presence of tax differences across states distort the spatial distribution of inventive activity up to -64.8% to 72.3%, with considerable spatial heterogeneity. The decomposition of those gains and losses reveals that not only the direct gains from tax changes but also the indirect gains via the top inventor migration driven by tax changes are important.

While this paper has considered the tax-induced domestic migration of top inventors, our model-based Bartik instruments can be used in any setting where origin-destination flows are affected by changes in location-specific policies. Thus, our framework would be applicable to various settings where the movement of goods, people, and ideas across space is influenced by policy differences between locations.

References

[1] Adão, Rodrigo, Michal Kolesár, Eduardo Morales (2019). Shift-Share Designs: Theory and Inference. *Quarterly Journal of Economics* 134(4): 1949–2010.

[2] Aghion, Philippe, and Xavier Jaravel (2015). Knowledge Spillovers, Innovation and Growth. *Economic Journal* 125 (March): 533–573.

[3] Akcigit, Ufuk, John Grigsby, Tom Nicholas, Stefanie Stantcheva (2022). Taxation and Innovation in the Twentieth Century. *Quarterly Journal of Economics* 137(1): 329–385.

[4] Akcigit, Ufuk, Stefanie Stantcheva (2022). Taxation and Innovation: What Do We Know? In Goolsbee, Austan, Benjamin Jones (Eds.) Innovation and Public Policy. *University of Chicago Press.*

[5] Andrews, Isaiah, James H. Stock, Liyang Sun (2019). Weak Instruments in Instrumental Variables Regression: Theory and Practice. *Annual Review of Economics* 11: 727–53.

[6] Andrews, Michael J. (2023). How Do Institutions of Higher Education Affect Local Invention? Evidence from the Establishment of US Colleges. *American Economic Journal: Economic Policy* 15(2): 1–41.

[7] Atkin, David, Keith Chen, Anton Popov (2022). The Returns to Face-to-Face Interactions: Knowledge Spillovers in Silicon Valley. *NBER Working Paper* #30147.

[8] Azoulay, Pierre, Joshua S. Graff Zivin, Jialan Wang (2010). Superstar Extinction. *Quarterly Journal of Economics* 125(2): 549–589.

[9] Azoulay, Pierre, Christian Fons-Rosen, Joshua S. Graff Zivin (2019). Does Science Advance One Funeral at a Time? *American Economic Review* 109(8): 2889–2920.

[10] Bartik, Timothy J. (1991). Who Benefits from State and Local Economic Development Policies? Kalamazoo, MI: W.E. Upjohn Institute.

[11] Borjas, George J., Kirk B. Doran (2012). The Collapse of the Soviet Union and the Productivity of American Mathematicians. *Quarterly Journal of Economics* 127(3): 1143–1203.

[12] Borusyak, Kirill, Peter Hull, Xavier Jaravel (2025). A Practical Guide to Shift-Share Instruments. *Journal of Economic Perspectives* 39(1): 181–204.

[13] Borusyak, Kirill, Peter Hull, Xavier Jaravel (2022). Quasi-Experimental Shift-Share Research Designs. *Review of Economic Studies* 89(1): 181–213.

[14] Brueckner, Jan K. (2003). Strategic Interaction among Governments: An Overview of Empirical Studies. *International Regional Science Review* 26(2): 175–188.

[15] Carlino, Gerald, William R. Kerr (2015). Agglomeration and Innovation. *Handbook of Regional and Urban Economics*, Volume 5, 349–404.

[16] Cattaneo, Matias D., Richard K. Crump, Max H. Farrell, and Yingjie Feng (2024). On Binscatter. *American Economic Review* 114(5): 1488–1514.

[17] de Chaisemartin, Clément, Xavier D'Haultfœuille (2023). Two-Way Fixed Effects and Differences-in-Differences with Heterogeneous Treatment Effects: A Survey. *Econometrics Journal* 26(3): C1–C30.

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

[19] Goldsmith-Pinkham, Paul, Isaac Sorkin, Henry Swift (2020). Bartik Instruments: What, When, Why, and How. *American Economic Review* 110(8): 2586–2624.

[20] Greenstone, Michael, Richard Hornbeck, Enrico Moretti (2010). Identifying Agglomeration Spillovers: Evidence from Winners and Losers of Large Plant Openings. *Journal of Political Economy*, 118(3): 536–598.

[21] Griliches, Zvi (1979). Issues in Assessing the Contribution of Research and Development to Productivity Growth. *Bell Journal of Economics* 10(1): 92–116.

[22] Jaffe, Adam B., Manuel Trajtenberg, Rebecca Henderson (1993). Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations. *Quarterly Journal of Economics* 108(3): 577–598.

[23] Kerr, William R., Frederic Robert-Nicoud (2020). Tech Clusters. *Journal of Economic Perspectives* 34(3): 50–76.

[24] Kleven, Henrik, Camille Landais, Mathilde Muñoz, Stefanie Stantcheva (2020). Taxation and Migration: Evidence and Policy Implications. *Journal of Economic Perspectives* 34(2): 119–142.

[25] Kleven, Henrik Jacobsen, Camille Landais, Emmanuel Saez (2013). Taxation and International Migration of Superstars: Evidence from the European Football Market. *American Economic Review* 103(5): 1892–1924.

[26] Kogan, Leonid, Dimitris Papanikolaou, Amit Seru, Noah Stoffman (2017). Technological Innovation, Resource Allocation, and Growth. *Quarterly Journal of Economics* 132(2): 665–712.

[27] Marshall, Alfred (1890). Principles of Economics. Macmillan, London.

[28] Montiel Olea, José Luis, Carolin Pflueger (2013). A Robust Test for Weak Instruments. *Journal of Business & Economic Statistics* 31(3): 358–369.

[29] Moretti, Enrico (2021). The Effect of High-Tech Clusters on the Productivity of Top Inventors. *American Economic Review* 111(10): 3328–3375.

[30] Moretti, Enrico, Daniel J. Wilson (2017). The Effect of State Taxes on the Geographical Location of Top Earners: Evidence from Star Scientists. *American Economic Review* 107(7): 1858–1903.

[31] Moser, Petra, Alessandra Voena, Fabian Waldinger (2014). German Jewish Émigrés and US Invention. *American Economic Review* 104(10): 3222–3255.

[32] Murata, Yasusada, Ryo Nakajima (2023). Marshall meets Bartik: Revisiting the mysteries of the trade. *Keio-IES Discussion Paper Series* DP2023–015.

[33] Murata, Yasusada, Ryo Nakajima, Ryosuke Okamoto, Ryuichi Tamura (2014). Localized Knowledge Spillovers and Patent Citations: A Distance-Based Approach. *Review of Economics and Statistics* 96(5): 967–985.

[34] Png, I. P. L. (2017a). Law and Innovation: Evidence from State Trade Secrets Laws. *Review of Economics and Statistics* 99(1): 167–179.

[35] Png, I. P. L. (2017b). Secrecy and Patents: Theory and Evidence from the Uniform Trade Secrets Act. *Strategy Science* 2(3): 176–193.

[36] Prato, Marta (2025). The Global Race for Talent: Brain Drain, Knowledge Transfer, and Growth. *Quarterly Journal of Economics* 140(1): 165–238.

[37] Rambachan, Ashesh, Jonathan Roth (2023). A More Credible Approach to Parallel Trends. *Review of Economic Studies* 90(5): 2555–2591.

[38] Roback, Jennifer (1982). Wages, Rents, and the Quality of Life. *Journal of Political Economy* 90(6): 1257–1278.

[39] Robinson, Sarah, Alisa Tazhitdinova (2025). One Hundred Years of U.S. State Taxation. *Journal of Public Economics* 241: 105273.

[40] Romer, Paul M. (1990). Endogenous Technological Change. *Journal of Political Economy* 98(5): S71–S102.

[41] Rotemberg, Julio J. (1983). Instrumental Variable Estimation of Misspecified Models. *MIT Sloan Working Paper* 1508–83.

[42] Schmidheiny, Kurt, Sebastian Siegloch (2023). On Event Studies and Distributed-lags in Two-way Fixed Effects Models: Identification, Equivalence, and Generalization. *Journal of Applied Econometrics* 38(5): 695–713.

[43] Simonsohn, Uri, Joseph P. Simmons, Leif D. Nelson (2020). Specification curve analysis. *Nature Human Behaviour* 4, 1208–1214.

[44] Stantcheva, Stefanie (2021). The Effects of Taxes on Innovation: Theory and Empirical Evidence. *NBER Working Paper* #29359.

[45] Zacchia, Paolo (2018). Benefiting colleagues but not the city: Localized effects from the relocation of superstar inventors. *Research Policy* 47, 992–1005.

Appendix A Data appendix

A.1 Data sources and construction

Patent data. The main data come from USPTO PatentsView (`https://patentsview.org/`). It includes data on patents, inventors, inventors' addresses, assignees, and patent citations and provides data files regarding the disambiguation of inventor and assignee names (`https://patentsview.org/disambiguation/`). Additional procedures are used to allocate inventors' addresses to commuting zones. We first use the latitude and longitude of each inventor's address (which are taken from USPTO PatentsView) to identify his/her county of residence. We then relate it to the commuting zone in which the inventor resides based on the correspondence table between counties and commuting zones in 1990 on the IPUMS USA website (`https://usa.ipums.org/usa/volii/1990lma.shtml`).

The disambiguation algorithm adopted in PatentsView is known to be highly accurate, as it copes with the two problems involving false positives and nagatives pointed out by Trajtenberg et al. (2006) and Monath et al. (2021). One is multiple names for the same entity (assignee, inventor, or location), e.g., the spelling of an assignee's name may differ from one patent to another. The other is multiple different entities with the same name, e.g., different inventors might share the exact same name (also known as the "John Smith" problem).

However, the disambiguation process is not error-free (see Toole et al., 2021). To address the former problem, we check the most typical assignee names: International Business Machines Corporation and IBM Corporation. In our dataset, the number of applications for International Business Machines Corporation is $207,139$, whereas that for IBM Corporation is $1,074$, which implies that $\frac{207,139}{207,139+1,074} \times 100 \approx 99.484\%$ of the IBM-related applications are classified into the same assignee. The latter problem implies that different inventors with the same name might be mistakenly recognized as the same inventor, which could generate seemingly frequent moves. To cope with possible overdetection of moves, we focus on the top inventors who moved fewer than eight times. This still leaves us with $59,770$ out of $60,294$ unique top inventors—more than $\frac{59,770}{60,294} \times 100 \approx 99.131\%$ of the inventors who applied for patents between two consecutive years *and* who qualified as top inventors at least once.

State taxes and tax credits. Data on U.S. state taxes and tax credits for 1976-2019 are obtained from openICPSR (https://www.openicpsr.org/openicpsr/project/113057/version/V1/view). Summary statistics are presented in the Online Appendix for Moretti and Wilson (2017).

Employment data. The employment data are taken from the County Business Patterns (CBP) database (http://fpeckert.me/cbp/). Eckert et al. (2021) provide a detailed description of the data. Since the original employment data are at the county level, we aggregate them at the commuting zone level. We use the 2012 NAICS codes 31-33 to obtain the number of employees for manufacturing, code 52 for finance and insurance, code 54 for professional, scientific, and technical services, and code 55 for management of companies and enterprises.

Trade secrets index and state-level uniform laws data. The trade secrets index is compiled by Png (2017a, b). Each state has six binary scores regarding the strength of legal protection of trade secrets under the common law and the Uniform Trade Secrets Act (UTSA). The trade secrets index used in our analysis is the sum of the six scores divided by six, which takes a value between zero and one. Data on state-level uniform laws such as the Uniform Determination of Death Act (UDDA), Uniform Federal Lien Registration Act (UFLRA), and Uniform Fraudulent Transfer Act (UFTA) are provided by Png (2017b).

Data for state tax competition analysis. Data on the socio-politico-economic characteristics used in the state tax competition analysis are taken from multiple sources. The data on population for various age and race groups are from the "U.S. Intercensal County Population Data by Age, Sex, Race, and Hispanic Origin" web page (https://www.nber.org/research/data/us-intercensal-county-population-data-age-sex-race-and-hispanic-origin) operated by NBER. The key economic and state finance data are from the "State Economic and Government Finance Data" web page (https://doi.org/10.7910/DVN/CJBTGD) provided by Klarner (2015). The political party affiliation of each state governor is from the National Governors Association web page (https://www.nga.org/governors/).

A.2 Other summary statistics

Table A1: Summary statistics at the commuting zone level (other variables).

	mean	sd	min	max
ATR	0.238	0.030	0.164	0.330
ATR99	0.315	0.032	0.244	0.410
ATR50	0.108	0.027	0.033	0.169
CITR	0.064	0.027	0.000	0.138
ITC	0.009	0.023	0.000	0.100
RTC	0.022	0.038	0.000	0.250
TSI	0.340	0.235	0.000	0.767
UDDA	0.490	0.500	0.000	1.000
UFLRA	0.596	0.491	0.000	1.000
UFTA	0.522	0.500	0.000	1.000
MFG	22,398.524	61,797.154	0.000	1,152,493.572
FIN	7,304.904	27,786.213	0.000	541,668.019
PRO	7,342.944	31,249.756	0.000	616,664.500
MNG	3,463.924	13,608.328	0.000	168,097.910
CITES ALL	21.966	347.065	0.000	17,344.000
CITES EXT	2.098	37.218	0.000	2,215.000
Number of observations				23,628
Number of commuting zones				716
Number of years				33

Notes: Summary statistics are based on the data described in Section 2 for the years 1977 to 2009. ATR (ATR99, ATR50), CITR, ITC, RTC, TSI, UDDA, UFLRA, and UFTA stand for the individual income average tax rate at the ninety-fifth (ninety-ninth, fiftieth) percentile of the U.S. income distribution, corporate income tax rate, investment tax credits, R&D tax credits, trade secrets index, Uniform Determination of Death Act, Uniform Federal Lien Registration Act, and Uniform Fraudulent Transfer Act, respectively. MFG, FIN, PRO, and MNG denote the employment in "manufacturing," "finance and insurance," "professional, scientific, and technical services," and "management of companies and enterprises." CITES ALL and CITES EXT are the number of citations by all local inventors and the number of citations by external inventors.

Table A2: Summary statistics at the state level (other variables).

	mean	sd	min	max
ATR	0.240	0.030	0.164	0.330
ATR99	0.317	0.032	0.244	0.410
ATR50	0.108	0.026	0.033	0.169
CITR	0.067	0.028	0.000	0.138
ITC	0.009	0.022	0.000	0.100
RTC	0.024	0.044	0.000	0.250
TSI	0.339	0.227	0.000	0.767
UDDA	0.522	0.500	0.000	1.000
UFLRA	0.580	0.494	0.000	1.000
UFTA	0.522	0.500	0.000	1.000
Number of observations				1,584
Number of states				48
Number of years				33

Notes: Summary statistics are based on the data described in Section 2 for the years 1977 to 2009. ATR (ATR99, ATR50), CITR, ITC, RTC, TSI, UDDA, UFLRA, and UFTA stand for the individual income average tax rate at the ninety-fifth (ninety-ninth, fiftieth) percentile of the U.S. income distribution, corporate income tax rate, investment tax credits, R&D tax credits, trade secrets index, Uniform Determination of Death Act, Uniform Federal Lien Registration Act, and Uniform Fraudulent Transfer Act, respectively.

Appendix B Derivation

B.1 Derivation of equation (3)

To derive (3), we first solve (2) for $\ln w_{dt} - \ln w_{ot}$ as follows

$$\ln w_{dt} - \ln w_{ot} = \beta[\ln(1-\tau'_{\sigma(d)t}) - \ln(1-\tau'_{\sigma(o)t})] + [Z'_d - Z'_o] - C'_{od} - \ln(P'_{odt}/P'_{oot}).$$

Plugging this expression into (1) and setting $\ln(P'_{odt}/P'_{oot}) = \ln(P_{odt}/P_{oot})$, we obtain

$$\begin{aligned}\ln(P_{odt}/P_{oot}) &= \alpha[\ln(1-\tau_{\sigma(d)t}) - \ln(1-\tau_{\sigma(o)t})] \\ &\quad +\alpha\{\beta[\ln(1-\tau'_{\sigma(d)t}) - \ln(1-\tau'_{\sigma(o)t})] + [Z'_d - Z'_o] - C'_{od} - \ln(P_{odt}/P_{oot})\} \\ &\quad +[Z_d - Z_o] - C_{od},\end{aligned}$$

which yields

$$\begin{aligned}(1+\alpha)\ln(P_{odt}/P_{oot}) &= \alpha[\ln(1-\tau_{\sigma(d)t}) - \ln(1-\tau_{\sigma(o)t})] \\ &\quad +\alpha\{\beta[\ln(1-\tau'_{\sigma(d)t}) - \ln(1-\tau'_{\sigma(o)t})] + [Z'_d - Z'_o] - C'_{od}\} \\ &\quad +[Z_d - Z_o] - C_{od}.\end{aligned}$$

We thus have

$$\begin{aligned}\ln(P_{odt}/P_{oot}) &= \tfrac{\alpha}{1+\alpha}[\ln(1-\tau_{\sigma(d)t}) - \ln(1-\tau_{\sigma(o)t})] + \tfrac{\alpha\beta}{1+\alpha}[\ln(1-\tau'_{\sigma(d)t}) - \ln(1-\tau'_{\sigma(o)t})] \\ &\quad +\tfrac{1}{1+\alpha}[Z_d - Z_o] + \tfrac{\alpha}{1+\alpha}[Z'_d - Z'_o] - \tfrac{1}{1+\alpha}[C_{od} + \alpha C'_{od}] \\ &= \tfrac{\alpha}{1+\alpha}[\ln(1-\tau_{\sigma(d)t}) - \ln(1-\tau_{\sigma(o)t})] + \tfrac{\alpha\beta}{1+\alpha}[\ln(1-\tau'_{\sigma(d)t}) - \ln(1-\tau'_{\sigma(o)t})] \\ &\quad +\tfrac{1}{1+\alpha}[Z_d + \alpha Z'_d] - \tfrac{1}{1+\alpha}[Z_o + \alpha Z'_o] - \tfrac{1}{1+\alpha}[C_{od} + \alpha C'_{od}].\end{aligned}$$

Setting $\eta = \frac{\alpha}{1+\alpha}$, $\eta' = \frac{\alpha\beta}{1+\alpha}$, $\gamma_d = \frac{1}{1+\alpha}[Z_d + \alpha Z'_d]$, $\gamma_o = -\frac{1}{1+\alpha}[Z_o + \alpha Z'_o]$, and $\gamma_{od} = -\frac{1}{1+\alpha}[C_{od} + \alpha C'_{od}]$, the foregoing equation can be rewritten as

$$\begin{aligned}\ln(P_{odt}/P_{oot}) &= \eta[\ln(1-\tau_{\sigma(d)t}) - \ln(1-\tau_{\sigma(o)t})] + \eta'[\ln(1-\tau'_{\sigma(d)t}) - \ln(1-\tau'_{\sigma(o)t})] \\ &\quad +\gamma_d + \gamma_o + \gamma_{od}. \qquad (19)\end{aligned}$$

Adding an error term u_{odt} to the right-hand side of (19), we obtain the expression in (3).

B.2 Derivation of equation (5)

We derive the predicted probability $\widehat{P}_{odt}$ in (5) that top inventors migrate from o to d in year t by the following three steps. First, equation (19) implies that for any pair of commuting zones c and d, P_{oct} and P_{odt} must satisfy

$$\frac{P_{oct}}{P_{odt}} = \frac{\exp\{\eta \ln(1-\tau_{\sigma(c)t}) + \eta' \ln(1-\tau'_{\sigma(c)t}) + \gamma_c + \gamma_{oc}\}}{\exp\{\eta \ln(1-\tau_{\sigma(d)t}) + \eta' \ln(1-\tau'_{\sigma(d)t}) + \gamma_d + \gamma_{od}\}}.$$

Second, let $\mathcal{C}$ denote the set of *all* commuting zones including origin commuting zone o and destination commuting zone d. Since $\sum_{c\in\mathcal{C}} P_{oct} = P_{oot} + P_{odt} + \sum_{c\in\mathcal{C}, c\neq\{o,d\}} P_{oct} = 1$ holds, we have

$$\frac{\sum_{c\in\mathcal{C}} P_{oct}}{P_{odt}} = \frac{1}{P_{odt}} = \frac{\sum_{c\in\mathcal{C}} \exp\{\eta \ln(1-\tau_{\sigma(c)t}) + \eta' \ln(1-\tau'_{\sigma(c)t}) + \gamma_c + \gamma_{oc}\}}{\exp\{\eta \ln(1-\tau_{\sigma(d)t}) + \eta' \ln(1-\tau'_{\sigma(d)t}) + \gamma_d + \gamma_{od}\}},$$

so that

$$P_{odt} = \frac{\exp\{\eta \ln(1-\tau_{\sigma(d)t}) + \eta' \ln(1-\tau'_{\sigma(d)t}) + \gamma_d + \gamma_{od}\}}{\sum_{c\in\mathcal{C}} \exp\{\eta \ln(1-\tau_{\sigma(c)t}) + \eta' \ln(1-\tau'_{\sigma(c)t}) + \gamma_c + \gamma_{oc}\}}.$$

Finally, replacing the parameters with the estimates, $\widehat{\eta}$, $\widehat{\eta}'$, and $\{\widehat{\gamma}_c, \widehat{\gamma}_{oc}\}_{c\in\mathcal{C}}$, from (3) yields (5).

Appendix C Additional robustness checks

C.1 Alternative tax rates and alternative detection of top inventor migrations

In this subsection, we conduct four sets of robustness checks. We first consider an alternative case, where we use the ATR for a hypothetical taxpayer at the ninety-ninth percentile of the U.S. income distribution. We then use the statutory marginal tax rates (MTRs) and turn to the case of the average property tax rates (APTRs). We finally consider alternative detection of top inventor migrations.

Alternative average tax rates (ATRs). In the log odds regressions in Section 3, $\tau_{\sigma(d)t}$ is proxied by the ATR for a hypothetical taxpayer at the ninety-fifth percentile of the U.S. income distribution. Table C1 reports an alternative case where we use the ninety-ninth percentile. Using the specification in Column 2 of Table C1, we construct the three types of Bartik instruments as in Section 4 and check the robustness of our main results in Table 5. As seen from Table C2, the semi-elasticities are virtually identical to those in Table 5: Panel

(a) shows that an inflow of a top inventor raises the patent productivity of all local inventors by approximately 6%, and Panel (b) shows that a top inventor inflow enhances local patent productivity by approximately 4% when we focus on external inventors.

Table C1: The impact of tax differences on the migration of top inventors (ATR99).

	(1)	(2)	(3)	(4)
$\Delta \ln(1 - \text{ATR99})$	2.340	2.504	2.382	2.455
	(1.315)	(1.198)	(1.150)	(0.966)
$\Delta \ln(1 - \text{CITR})$	-0.440	-0.251	-0.147	-0.018
	(1.115)	(1.063)	(0.841)	(0.750)
$\Delta \ln(1 + \text{ITC})$	-0.700	-0.913	-0.549	-0.711
	(0.719)	(0.672)	(0.948)	(0.666)
$\Delta \ln(1 + \text{RTC})$	0.223	0.234	0.319	0.124
	(0.445)	(0.392)	(0.324)	(0.285)
CZ pair FE	Yes	Yes	No	No
Origin CZ FE and destination CZ FE	No	No	Yes	Yes
Year FE	Yes	No	Yes	No
Region pair × year FE	No	Yes	No	Yes
Observations	4,866	4,866	7,226	7,225
$\overline{R}^2$ (total)	0.892	0.902	0.906	0.916
$\overline{R}^2$ (within)	0.392	0.451	0.405	0.003

Notes: The dependent variable in each column is the log odds ratio in equation (3). ATR99, CITR, ITC, and RTC stand for the individual income average tax rate at the ninety-ninth percentile, corporate income tax rate, investment tax credits, and R&D tax credits, respectively. $\Delta \ln(1 - \text{ATR99})$ is defined as $\ln(1 - \text{ATR99}_{\sigma(d)t}) - \ln(1 - \text{ATR99}_{\sigma(o)t})$. $\Delta \ln(1 - \text{CITR})$, $\Delta \ln(1 + \text{ITC})$, and $\Delta \ln(1 + \text{RTC})$ are defined analogously. Cluster-robust standard errors are in parentheses.

Using this alternative ATR, we also conduct the dynamic analysis in Figure C1. Panel (a) illustrates the dynamic impact of a top inventor inflow on the patent productivity of all local inventors, which includes not only the gains from internal knowledge sharing within assignees and between co-inventors but also the gains from external knowledge spillovers. Panel (b) focuses on the latter gains. In both cases, the results are quite similar to those in Figure 6: the pre-event semi-elasticities are close to zero and the post-event semi-elasticities increase to approximately 0.05, which ensures the main results from our static model. We report the numbers that we use in Figure C1 and the associated first-stage statistics in Tables C13 and C14, respectively.

Statutory marginal tax rates (MTRs). In Section 3, we examine the effect of ATRs on the migration of top inventors as in Moretti and Wilson (2017). In Section 4, we further

Table C2: The impact of top inventor inflows on local patent productivity (ATR99).

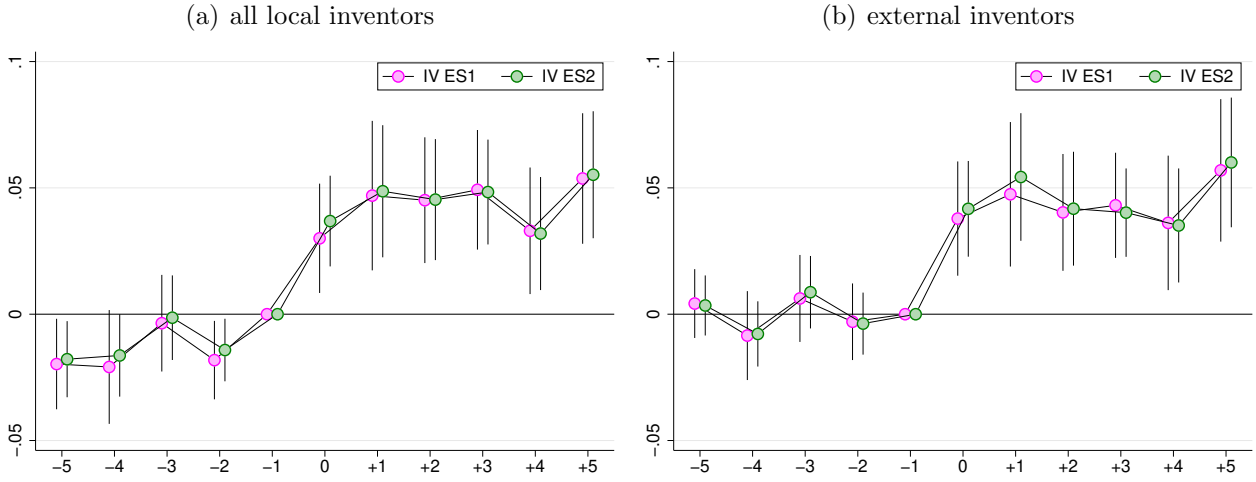
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(a) All local inventors							
Top inventor inflows	0.044	0.064	0.062	0.062	0.065	0.059	0.060
	(0.007)	(0.013)	(0.012)	(0.012)	(0.014)	(0.013)	(0.012)
$\ln(1 - \text{ATR99})$	2.580	2.625	2.604	2.658			
	(0.693)	(0.699)	(0.697)	(0.699)			
Effective F statistic		36.605	33.598	33.179	52.203	35.170	35.141
$\tau = 5\%$		37.418	31.742	34.735	37.418	30.971	33.112
$\tau = 10\%$		23.109	19.781	21.385	23.109	19.330	20.431
$\tau = 20\%$		15.062	13.025	13.894	15.062	12.751	13.309
$\tau = 30\%$		12.039	10.478	11.086	12.039	10.269	10.637
(b) External inventors							
Top inventor inflows	0.027	0.044	0.043	0.044	0.041	0.035	0.038
	(0.004)	(0.010)	(0.009)	(0.009)	(0.011)	(0.010)	(0.010)
$\ln(1 - \text{ATR99})$	2.142	2.178	2.136	2.160			
	(0.617)	(0.621)	(0.618)	(0.619)			
Effective F statistic		36.605	33.598	33.179	52.203	35.170	35.141
$\tau = 5\%$		37.418	31.713	34.738	37.418	30.948	33.112
$\tau = 10\%$		23.109	19.764	21.386	23.109	19.318	20.431
$\tau = 20\%$		15.062	13.016	13.895	15.062	12.744	13.309
$\tau = 30\%$		12.039	10.471	11.086	12.039	10.263	10.636
CZ FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	No	No	No
State × year FE	No	No	No	No	Yes	Yes	Yes
Observations	23,628	23,628	23,628	23,463	23,562	23,562	23,397

Notes: The coefficient on top inventor inflows is converted to semi-elasticity. ATR99 stands for the individual income average tax rate at the ninety-ninth percentile. The coefficient on $\ln(1 - \text{ATR99})$ is converted to elasticity. Column 1 does not control for the endogeneity of top inventor inflows. Column 2 uses B_{dt} as an instrument. Column 3 uses B_{dt} and B^{σ}_{dt} as instruments. Column 4 uses B_{dt}, B^{σ}_{dt}, and B^{ν}_{dt} as instruments. Columns 5-7 replace $\ln(1 - \text{ATR99})$ in Columns 2-4 with state × year FE. Cluster-robust standard errors are in parentheses.

derive the Bartik instruments from the migration identity to investigate the impact of top inventor inflows on local patent productivity.

To assess the robustness of our findings, we replace the ATRs with the statutory marginal tax rates (MTRs) at the ninety-fifth percentile of the U.S. income distribution and reproduce the migration analysis in Table 4 and the productivity analysis in Table 5. As seen from Table C3, the effect of tax differences on the migration of top inventors is consistent with the results in Tables 4. Table C4 further shows that the impact of a top inventor inflow on local patent productivity remains stable, thus suggesting that our main results in Table 5 are robust regardless of whether we use the average or marginal tax rates.

Figure C1: IV event study regressions (ATR99).



Notes: Panels (a) and (b) illustrate the dynamic impacts of a top inventor inflow on the patent productivity of all local inventors and that of external inventors, respectively. In each panel, IV ES1 uses B_{dt} and B^{σ}_{dt} as instruments and IV ES2 uses B_{dt}, B^{σ}_{dt}, and B^{ν}_{dt} as instruments.

Table C3: The impact of tax differences on the migration of top inventors (MTR).

	(1)	(2)	(3)	(4)
$\Delta \ln(1 - \text{MTR})$	6.842	6.623	5.937	5.979
	(1.557)	(1.245)	(1.158)	(1.000)
$\Delta \ln(1 - \text{CITR})$	-0.990	-0.674	-0.576	-0.394
	(1.044)	(0.981)	(0.801)	(0.705)
$\Delta \ln(1 + \text{ITC})$	-0.329	-0.518	-0.273	-0.422
	(0.691)	(0.642)	(0.950)	(0.653)
$\Delta \ln(1 + \text{RTC})$	0.162	0.204	0.326	0.156
	(0.441)	(0.384)	(0.320)	(0.277)
CZ pair FE	Yes	Yes	No	No
Origin CZ FE and destination CZ FE	No	No	Yes	Yes
Year FE	Yes	No	Yes	No
Region pair × year FE	No	Yes	No	Yes
Observations	4,866	4,866	7,226	7,225
$\overline{R}^2$ (total)	0.894	0.904	0.908	0.917
$\overline{R}^2$ (within)	0.404	0.461	0.413	0.016

Notes: The dependent variable in each column is the log odds ratio in equation (3). MTR, CITR, ITC, and RTC stand for statutory marginal tax rate at the ninety-fifth percentile, corporate income tax rate, investment tax credit, and R&D tax credit, respectively. $\Delta \ln(1-\text{MTR})$ is defined as $\ln(1-\text{MTR}_{\sigma(d)t}) - \ln(1-\text{MTR}_{\sigma(o)t})$. $\Delta \ln(1 - \text{CITR})$, $\Delta \ln(1 + \text{ITC})$, and $\Delta \ln(1 + \text{RTC})$ are defined analogously. Cluster-robust standard errors are in parentheses.

Table C4: The impact of top inventor inflows on local patent productivity (MTR).

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(a) All local inventors							
Top inventor inflows	0.042	0.061	0.060	0.059	0.063	0.058	0.059
	(0.007)	(0.013)	(0.012)	(0.011)	(0.014)	(0.012)	(0.012)
$\ln(1-\text{MTR})$	3.368	3.134	3.229	3.268			
	(0.783)	(0.790)	(0.781)	(0.781)			
Effective F statistic		39.387	35.309	35.008	50.760	37.522	37.501
$\tau = 5\%$		37.418	31.707	34.385	37.418	30.993	32.442
$\tau = 10\%$		23.109	19.761	21.177	23.109	19.344	20.035
$\tau = 20\%$		15.062	13.014	13.766	15.062	12.760	13.065
$\tau = 30\%$		12.039	10.470	10.987	12.039	10.276	10.448
(b) External inventors							
Top inventor inflows	0.027	0.041	0.040	0.041	0.036	0.038	0.037
	(0.005)	(0.010)	(0.009)	(0.009)	(0.010)	(0.010)	(0.009)
$\ln(1-\text{MTR})$	2.925	2.743	2.779	2.756			
	(0.678)	(0.684)	(0.681)	(0.680)			
Effective F statistic		39.387	35.309	35.008	50.760	37.514	37.489
$\tau = 5\%$		37.418	31.707	34.385	37.418	30.996	32.440
$\tau = 10\%$		23.109	19.761	21.177	23.109	19.345	20.034
$\tau = 20\%$		15.062	13.014	13.766	15.062	12.761	13.064
$\tau = 30\%$		12.039	10.470	10.987	12.039	10.276	10.448
CZ FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	No	No
State × year FE	No	No	No	No	Yes	Yes	Yes
Observations	23,628	23,628	23,628	23,463	23,562	23,562	23,397

Notes: The coefficient on top inventor inflows is converted to semi-elasticity. MTR stands for the statutory marginal tax rate at the ninety-fifth percentile. The coefficient on $\ln(1-\text{MTR})$ is converted to elasticity. Column 1 does not control for the endogeneity of top inventor inflows. Column 2 uses B_{dt} as an instrument. Column 3 uses B_{dt} and B^{σ}_{dt} as instruments. Column 4 uses B_{dt}, B^{σ}_{dt}, and B^{ν}_{dt} as instruments. Columns 5-7 replace $\ln(1-\text{MTR})$ in Columns 2-4 with state × year FE. Cluster-robust standard errors are in parentheses.

Average property tax rates (APTRs). Since top inventors may consider not only individual income taxes but also property taxes, we check the robustness of our results by incorporating both taxes into our location choice model. However, as pointed out by Moretti and Wilson (2017), there are no comprehensive panel data on effective property taxes. We thus follow their approach and measure the property taxes for top inventors by the average property tax rates (APTRs). The APTR in a county represents the proportion of income that an average individual residing in that county pays for property taxes. We compute this by dividing the total property tax revenue of a county by the total personal income of its residents. The county-level property tax revenue data are obtained from the Government

Finance Database (see Pierson et al. 2015).[42] The personal income data are sourced from the U.S. Bureau of Economic Analysis (BEA). We calculate the APTR for each commuting zone by weighting the county-level APTRs by their respective population shares within each commuting zone.

To estimate the impact of the property tax rates on top inventor migrations, we replace the income tax difference $\eta[\ln(1-\tau_{\sigma(d)t})-\ln(1-\tau_{\sigma(o)t})]$ in equation (3) with $\eta[\ln(1-\tau_{\sigma(d)t}-\tau^p_{dt})-\ln(1-\tau_{\sigma(o)t}-\tau^p_{ot})]$, where τ^p_{dt} and τ^p_{ot} denote the property tax rates in destination and origin commuting zones in period t, respectively. We assume that $\tau_{\sigma(d)t}$ and $\tau_{\sigma(o)t}$ are approximated by the ATRs paid by individuals at the ninety-fifth percentile of the U.S. income distribution, and that τ^p_{dt} and τ^p_{ot} are proxied by the APTRs, as in Moretti and Wilson (2017).

Table C5: The impact of tax differences on the migration of top inventors (APTR).

	(1)	(2)	(3)	(4)
$\Delta\ln(1-\text{ATR}-\text{APTR})$	6.503	6.219	5.553	5.729
	(1.477)	(1.322)	(1.244)	(1.063)
$\Delta\ln(1-\text{CITR})$	-0.720	-0.479	-0.510	-0.355
	(1.072)	(1.017)	(0.805)	(0.714)
$\Delta\ln(1+\text{ITC})$	-0.127	-0.354	-0.126	-0.282
	(0.726)	(0.678)	(0.980)	(0.680)
$\Delta\ln(1+\text{RTC})$	0.243	0.251	0.309	0.107
	(0.443)	(0.393)	(0.323)	(0.281)
CZ pair FE	Yes	Yes	No	No
Origin CZ FE and destination CZ FE	No	No	Yes	Yes
Year FE	Yes	No	Yes	No
Region pair × year FE	No	Yes	No	Yes
Observations	4,852	4,852	7,204	7,203
$\overline{R}^2$ (total)	0.893	0.903	0.907	0.917
$\overline{R}^2$ (within)	0.398	0.456	0.409	0.010

Notes: The dependent variable in each column is the log odds ratio in equation (3). ATR, CITR, ITC, RTC, and APTR stand for the individual income average tax rate at the ninety-fifth percentile, corporate income tax rate, investment tax credit, R&D tax credit, and average property tax rate, respectively. $\Delta\ln(1-\text{ATR}-\text{APTR})$ is defined as $\ln(1-\text{ATR}_{\sigma(d)t}-\text{APTR}_{dt})-\ln(1-\text{ATR}_{\sigma(o)t}-\text{APTR}_{ot})$. $\Delta\ln(1-\text{CITR})$, $\Delta\ln(1+\text{ITC})$, and $\Delta\ln(1+\text{RTC})$ are defined analogously. Cluster-robust standard errors are in parentheses.

Table C5 presents the estimation results. The result for each specification shows that top inventors respond to the tax difference, $\Delta\ln(1-\text{ATR}-\text{APTR})$. The coefficient on this term is similar to the coefficient on $\Delta\ln(1-\text{ATR})$ in Table 4, which ensures our previous results.

[42]The database is available at `https://willamette.edu/mba/research-impact/public-datasets/`. When the property tax revenue data for a specific county and year is missing, we use an interpolated value based on data from other years for that county.

Table C6: The impact of top inventor inflows on local patent productivity (APTR).

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(a) All local inventors							
Top inventor inflows	0.043	0.064	0.062	0.062	0.065	0.059	0.060
	(0.007)	(0.013)	(0.012)	(0.012)	(0.014)	(0.012)	(0.012)
$\ln(1-\text{ATR}-\text{APTR})$	1.638	1.564	1.574	1.625			
	(0.936)	(0.920)	(0.923)	(0.936)			
Effective F statistic		36.829	31.766	31.443	50.188	34.187	34.145
$\tau = 5\%$		37.418	30.970	34.518	37.418	31.322	32.916
$\tau = 10\%$		23.109	19.328	21.241	23.109	19.535	20.324
$\tau = 20\%$		15.062	12.749	13.794	15.062	12.876	13.249
$\tau = 30\%$		12.039	10.266	11.003	12.039	10.364	10.593
(b) External inventors							
Top inventor inflows	0.027	0.043	0.042	0.044	0.041	0.036	0.038
	(0.005)	(0.010)	(0.009)	(0.009)	(0.011)	(0.010)	(0.010)
$\ln(1-\text{ATR}-\text{APTR})$	0.947	0.889	0.887	0.888			
	(0.750)	(0.738)	(0.740)	(0.749)			
Effective F statistic		36.829	31.766	31.443	50.188	34.187	34.145
$\tau = 5\%$		37.418	30.971	34.522	37.418	31.284	32.915
$\tau = 10\%$		23.109	19.328	21.243	23.109	19.514	20.323
$\tau = 20\%$		15.062	12.749	13.795	15.062	12.864	13.249
$\tau = 30\%$		12.039	10.266	11.004	12.039	10.355	10.593
CZ FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	No	No	No
State × year FE	No	No	No	No	Yes	Yes	Yes
Observations	23,584	23,584	23,584	23,419	23,562	23,562	23,397

Notes: The coefficient on top inventor inflows is converted to semi-elasticity. ATR and APTR stand for the individual income average tax rate at the ninety-fifth percentile and average property tax rate, respectively. The coefficient on $\ln(1-\text{ATR}-\text{APTR})$ is converted to elasticity. Column 1 does not control for the endogeneity of top inventor inflows. Column 2 uses B_{dt} as an instrument. Column 3 uses B_{dt} and B^{σ}_{dt} as instruments. Column 4 uses B_{dt}, B^{σ}_{dt}, and B^{ν}_{dt} as instruments. Columns 5-7 replace $\ln(1-\text{ATR}-\text{APTR})$ in Columns 2-4 with state × year FE. Cluster-robust standard errors are in parentheses.

Table C6 presents the estimated effects of top inventor inflows on local patent productivity, assuming that the property taxes paid by these top inventors can be proxied by the APTRs. The results are nearly identical to the ones in Table 5 obtained from the specifications without APTRs, thus showing the robustness of our main findings.

Alternative detection of top inventor migrations. As explained in Section 2, we detect the migration of a top inventor if the commuting zone of residence of that top inventor recorded in the patent application data differs between two consecutive years. Since the status of a top inventor varies from year to year, we consider the migration of inventors who qualify as top inventors in the first year as a baseline case. However, this does not imply that the migrating inventors retain the status of a top inventor in the year following their move. Consequently, our baseline detection of top inventor migrations includes the cases where inventors experience a decline in patent productivity after migration. Although such instances make up less than 5% of all top inventor migrations, inventors whose productivity declines may have a different impact on local inventors compared to other inventors whose productivity remains high.

Table C7: The impact of tax differences on the migration of top inventors (alternative detection of top inventor migrations).

	(1)	(2)	(3)	(4)
$\Delta \ln(1 - \text{ATR})$	7.384	6.933	6.127	6.373
	(1.655)	(1.445)	(1.317)	(1.126)
$\Delta \ln(1 - \text{CITR})$	0.046	0.471	-0.089	0.202
	(1.147)	(1.084)	(0.875)	(0.754)
$\Delta \ln(1 + \text{ITC})$	0.253	-0.017	0.197	0.016
	(0.757)	(0.706)	(1.018)	(0.696)
$\Delta \ln(1 + \text{RTC})$	0.338	0.352	0.341	0.156
	(0.452)	(0.404)	(0.329)	(0.286)
CZ pair FE	Yes	Yes	No	No
Origin CZ FE and destination CZ FE	No	No	Yes	Yes
Year FE	Yes	No	Yes	No
Region pair × year FE	No	Yes	No	Yes
Observations	4,702	4,702	6,969	6,968
$\overline{R}^2$ (total)	0.893	0.904	0.906	0.916
$\overline{R}^2$ (within)	0.398	0.461	0.406	0.012

Notes: The dependent variable in each column is the log odds ratio in equation (3). ATR, CITR, ITC, and RTC stand for the individual income average tax rate at the ninety-fifth percentile, corporate income tax rate, investment tax credit, and R&D tax credit, respectively. $\Delta \ln(1 - \text{ATR})$ is defined as $\ln(1 - \text{ATR}_{\sigma(d)t}) - \ln(1 - \text{ATR}_{\sigma(o)t})$. $\Delta \ln(1 - \text{CITR})$, $\Delta \ln(1 + \text{ITC})$, and $\Delta \ln(1 + \text{RTC})$ are defined analogously. Cluster-robust standard errors are in parentheses.

To address this issue, we redefine the migration of top inventors by focusing only on those who retain the status of a top inventor for two consecutive years before and after their migrations. We then re-estimate our models using this more stringent criterion to examine how top inventor inflows affect local patent productivity. Table C7 presents the results of

the tax-induced top inventor migrations, which replicate those in Table 4 reasonably well. Table C8 reports the impacts of top inventor inflows on local patent productivity, which are quite similar to those in Table 5. Thus, we may conclude that our main results are robust to this alternative detection of top inventor migrations.

Table C8: The impact of top inventor inflows on local patent productivity (alternative detection of top inventor migrations).

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(a) All local inventors							
Top inventor inflows	0.044	0.065	0.063	0.062	0.070	0.063	0.064
	(0.007)	(0.013)	(0.012)	(0.012)	(0.015)	(0.013)	(0.013)
$\ln(1-\text{ATR})$	6.030	5.898	5.876	5.992			
	(1.040)	(1.042)	(1.040)	(1.040)			
Effective F statistic		39.878	31.699	31.269	55.801	33.862	33.393
$\tau = 5\%$		37.418	32.620	34.839	37.418	32.003	33.012
$\tau = 10\%$		23.109	20.297	21.470	23.109	19.935	20.397
$\tau = 20\%$		15.062	13.342	13.965	15.062	13.121	13.307
$\tau = 30\%$		12.039	10.721	11.149	12.039	10.552	10.644
(b) External inventors							
Top inventor inflows	0.029	0.045	0.043	0.044	0.046	0.040	0.042
	(0.004)	(0.010)	(0.009)	(0.009)	(0.012)	(0.011)	(0.010)
$\ln(1-\text{ATR})$	4.772	4.667	4.612	4.587			
	(0.864)	(0.867)	(0.863)	(0.856)			
Effective F statistic		39.878	31.699	31.269	55.801	33.862	33.393
$\tau = 5\%$		37.418	32.614	34.840	37.418	31.987	33.012
$\tau = 10\%$		23.109	20.293	21.471	23.109	19.927	20.397
$\tau = 20\%$		15.062	13.340	13.966	15.062	13.116	13.307
$\tau = 30\%$		12.039	10.719	11.150	12.039	10.548	10.644
CZ FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	No	No	No
State × year FE	No	No	No	No	Yes	Yes	Yes
Observations	23,628	23,628	23,628	23,463	23,562	23,562	23,397

Notes: The coefficient on top inventor inflows is converted to semi-elasticity. ATR stands for the individual income average tax rate at the ninety-fifth percentile. The coefficient on $\ln(1-\text{ATR})$ is converted to elasticity. Column 1 does not control for the endogeneity of top inventor inflows. Column 2 uses B_{dt} as an instrument. Column 3 uses B_{dt} and B_{dt}^{σ} as instruments. Column 4 uses B_{dt}, B_{dt}^{σ}, and B_{dt}^{ν} as instruments. Columns 5-7 replace $\ln(1-\text{ATR})$ in Columns 2-4 with state × year FE. Cluster-robust standard errors are in parentheses.

C.2 Including other controls in our baseline specification

We further check the robustness of our main results in Table 5 by including additional controls such as the ATR at the fiftieth percentile (ATR50), CITR, ITC, and RTC, as well as manufacturing employment (MFG) at the commuting zone level. Table C9 shows that the results for the cases with these additional controls are virtually identical to those in Table 5 for all specifications including both the FE and IV cases. In particular, the IV regressions reported in Panel (a) show that an inflow of a top inventor raises the patent productivity of all local inventors by approximately 6%. The IV regressions reported in Panel (b) show that a top inventor inflow enhances local patent productivity by approximately 4% when we focus on external inventors. These results suggest that our main results in Table 5 are robust to the inclusion of these additional controls.

C.3 Dropping commuting zone $\times$ year observations with no patents

We check the robustness of our main results in Table 5 by dropping commuting zone $\times$ year observations with no patents while retaining $\ln Y_{dt}$, instead of using $\ln(1 + Y_{dt})$ as in the baseline case. Table C10 shows that the results are qualitatively similar to those in Table 5 for all specifications including both the FE and IV cases. In particular, the IV regressions reported in Panel (a) show that an inflow of a top inventor raises the patent productivity of all local inventors by approximately 5-6%. The IV regressions reported in Panel (b) show that a top inventor inflow enhances the patent productivity of external inventors by approximately 3-4%. These results can be viewed as constituting the lower bounds of productivity gains as we focus on intensive margins by abstracting from extensive margins (i.e., by excluding the possibility that commuting zones with no patents could potentially gain from top inventor inflows, if any).

Table C9: The impact of top inventor inflows on local patent productivity (other controls).

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(a) All local inventors								
Top inventor inflows	0.043	0.063	0.060	0.060	0.044	0.066	0.064	0.064
	(0.006)	(0.013)	(0.012)	(0.011)	(0.007)	(0.013)	(0.012)	(0.011)
$\ln(1-\text{ATR})$	6.159	6.175	6.137	6.217	6.245	6.267	6.201	6.256
	(1.179)	(1.182)	(1.179)	(1.183)	(1.147)	(1.146)	(1.144)	(1.148)
$\ln(1-\text{ATR50})$	-0.181	-0.285	-0.325	-0.234	-0.385	-0.510	-0.544	-0.437
	(1.388)	(1.376)	(1.377)	(1.379)	(1.372)	(1.358)	(1.360)	(1.363)
$\ln(1-\text{CITR})$	0.021	-0.176	-0.090	-0.134	-0.129	-0.354	-0.245	-0.288
	(0.645)	(0.647)	(0.640)	(0.654)	(0.637)	(0.638)	(0.632)	(0.646)
$\ln(1+\text{ITC})$	0.127	0.160	0.133	0.120	0.257	0.300	0.262	0.254
	(0.409)	(0.408)	(0.407)	(0.407)	(0.404)	(0.403)	(0.402)	(0.402)
$\ln(1+\text{RTC})$	-0.100	-0.212	-0.174	-0.172	-0.095	-0.219	-0.171	-0.165
	(0.251)	(0.252)	(0.248)	(0.253)	(0.250)	(0.251)	(0.247)	(0.251)
$\ln(1+\text{MFG})$					0.118	0.124	0.121	0.122
					(0.025)	(0.025)	(0.025)	(0.025)
Effective F statistic		38.554	33.272	32.878		38.751	33.214	32.814
$\tau = 5\%$		37.418	31.894	34.721		37.418	31.953	34.724
$\tau = 10\%$		23.109	19.871	21.381		23.109	19.905	21.384
$\tau = 20\%$		15.062	13.081	13.895		15.062	13.102	13.898
$\tau = 30\%$		12.039	10.521	11.088		12.039	10.537	11.091
(b) External inventors								
Top inventor inflows	0.027	0.042	0.040	0.041	0.028	0.044	0.042	0.044
	(0.004)	(0.010)	(0.009)	(0.009)	(0.004)	(0.010)	(0.009)	(0.009)
$\ln(1-\text{ATR})$	4.790	4.802	4.756	4.690	4.847	4.864	4.805	4.734
	(0.978)	(0.983)	(0.980)	(0.977)	(0.958)	(0.961)	(0.958)	(0.958)
$\ln(1-\text{ATR50})$	-0.127	-0.207	-0.255	-0.102	-0.264	-0.359	-0.418	-0.269
	(1.181)	(1.176)	(1.175)	(1.176)	(1.170)	(1.164)	(1.163)	(1.165)
$\ln(1-\text{CITR})$	0.325	0.174	0.228	0.153	0.225	0.055	0.124	0.047
	(0.543)	(0.544)	(0.541)	(0.551)	(0.542)	(0.543)	(0.540)	(0.550)
$\ln(1+\text{ITC})$	0.272	0.298	0.277	0.293	0.359	0.392	0.364	0.387
	(0.395)	(0.395)	(0.394)	(0.393)	(0.393)	(0.393)	(0.392)	(0.392)
$\ln(1+\text{RTC})$	-0.050	-0.137	-0.103	-0.102	-0.047	-0.141	-0.100	-0.098
	(0.218)	(0.221)	(0.218)	(0.221)	(0.218)	(0.221)	(0.218)	(0.222)
$\ln(1+\text{MFG})$					0.079	0.084	0.082	0.083
					(0.020)	(0.020)	(0.020)	(0.020)
Effective F statistic		38.554	33.272	32.878		38.751	33.214	32.814
$\tau = 5\%$		37.418	31.884	34.726		37.418	31.941	34.729
$\tau = 10\%$		23.109	19.865	21.384		23.109	19.899	21.387
$\tau = 20\%$		15.062	13.078	13.897		15.062	13.098	13.900
$\tau = 30\%$		12.039	10.519	11.089		12.039	10.534	11.093
CZ FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State × year FE	No	No	No	No	No	No	No	No
Observations	23,628	23,628	23,628	23,463	23,628	23,628	23,628	23,463

Notes: The coefficient on top inventor inflows is converted to semi-elasticity. The other coefficients are converted to elasticities. ATR (ATR50), CITR, ITC, RTC, and MFG stand for the individual income average tax rate at the ninety-fifth (fiftieth) percentile, corporate income tax rate, investment tax credit, R&D tax credit, and manufacturing employment, respectively. Column 1 does not control for the endogeneity of top inventor inflows. Column 2 uses B_{dt} as an instrument. Column 3 uses B_{dt} and B^{σ}_{dt} as instruments. Column 4 uses B_{dt}, B^{σ}_{dt}, and B^{ν}_{dt} as instruments. Columns 5-8 repeat the same specifications as Columns 1-4 with $\ln(1+\text{MFG})$ at the commuting zone level. Cluster-robust standard errors are in parentheses.

Table C10: The impact of top inventor inflows on local patent productivity (dropping zeros).

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(a) All local inventors							
Top inventor inflows	0.039	0.054	0.051	0.050	0.059	0.051	0.052
	(0.006)	(0.012)	(0.010)	(0.010)	(0.013)	(0.012)	(0.011)
$\ln(1-\text{ATR})$	6.934	6.838	6.841	7.017			
	(1.257)	(1.260)	(1.259)	(1.255)			
Effective F statistic		37.096	32.976	32.657	49.965	34.258	34.170
$\tau = 5\%$		37.418	31.930	34.704	37.418	31.378	33.007
$\tau = 10\%$		23.109	19.892	21.372	23.109	19.569	20.377
$\tau = 20\%$		15.062	13.094	13.890	15.062	12.897	13.282
$\tau = 30\%$		12.039	10.531	11.085	12.039	10.380	10.619
CZ FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	No	No	No
State × year FE	No	No	No	No	Yes	Yes	Yes
Observations	20,038	20,038	20,038	19,941	19,972	19,972	19,875
(b) External inventors							
Top inventor inflows	0.023	0.034	0.032	0.033	0.035	0.029	0.031
	(0.004)	(0.009)	(0.008)	(0.008)	(0.011)	(0.009)	(0.009)
$\ln(1-\text{ATR})$	5.450	5.380	5.357	5.340			
	(1.054)	(1.056)	(1.054)	(1.044)			
Effective F statistic		37.072	32.971	32.651	52.203	34.229	34.137
$\tau = 5\%$		37.418	31.917	34.706	37.418	31.350	32.995
$\tau = 10\%$		23.109	19.885	21.373	23.109	19.553	20.370
$\tau = 20\%$		15.062	13.090	13.891	15.062	12.888	13.278
$\tau = 30\%$		12.039	10.528	11.085	12.039	10.263	10.615
CZ FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	No	No	No
State × year FE	No	No	No	No	Yes	Yes	Yes
Observations	19,903	19,903	19,903	19,806	19,837	19,837	19,740

Notes: The coefficient on top inventor inflows is semi-elasticity. ATR stands for the individual income average tax rate at the ninety-fifth percentile. The coefficient on $\ln(1-\text{ATR})$ is elasticity. Column 1 does not control for the endogeneity of top inventor inflows. Column 2 uses B_{dt} as an instrument. Column 3 uses B_{dt} and B^{σ}_{dt} as instruments. Column 4 uses B_{dt}, B^{σ}_{dt}, and B^{ν}_{dt} as instruments. Columns 5-7 replace $\ln(1-\text{ATR})$ in Columns 2-4 with state × year FE. Cluster-robust standard errors are in parentheses.

C.4 An alternative measure of patent productivity: Patent quality

To examine the robustness of our main results, we further consider an alternative measure of patent productivity. We adopt patent quality measured by patent market values in Kogan et al. (2017). This approach allows us to consider alternative definitions of top inventors and local patent productivity.

We first identify the top 5% of inventors based on this measure over the last ten years and refer to them as top inventors as before. We then examine the effect of ATRs on their migrations as in Table 4 and obtain qualitatively similar results (see Table C11). Although the effect is somewhat weaker, those estimates are sufficient for constructing Bartik instruments that are strong enough to pass the first stage tests.

Using the Bartik instruments thus obtained, we estimate the impact of a top inventor inflow on local patent productivity as in Table 5 and find that it remains positive and statistically significant both for all local inventors and for external inventors (see Table C12). Hence, we confirm the robustness of our main findings to this alternative measure of patent productivity.

Table C11: The impact of tax differences on the migration of top inventors (patent quality).

	(1)	(2)	(3)	(4)
$\Delta \ln(1 - \text{ATR})$	3.934	3.334	3.614	4.228
	(2.145)	(2.032)	(1.512)	(1.410)
$\Delta \ln(1 - \text{CITR})$	-0.648	-0.719	-0.316	-0.232
	(1.410)	(1.298)	(1.006)	(0.906)
$\Delta \ln(1 + \text{ITC})$	-0.508	-1.251	-0.784	-1.149
	(0.718)	(0.660)	(1.030)	(0.811)
$\Delta \ln(1 + \text{RTC})$	0.621	0.321	0.802	0.594
	(0.578)	(0.517)	(0.393)	(0.363)
CZ pair FE	Yes	Yes	No	No
Origin CZ FE and destination CZ FE	No	No	Yes	Yes
Year FE	Yes	No	Yes	No
Region pair × year FE	No	Yes	No	Yes
Observations	3,582	3,582	5,146	5,138
$\overline{R}^2$ (total)	0.868	0.882	0.883	0.896
$\overline{R}^2$ (within)	0.331	0.402	0.349	0.007

Notes: The dependent variable in each column is the log odds ratio in equation (3). ATR, CITR, ITC, and RTC stand for the individual income average tax rate at the ninety-fifth percentile, corporate income tax rate, investment tax credit, and R&D tax credit, respectively. $\Delta \ln(1 - \text{ATR})$ is defined as $\ln(1 - \text{ATR}_{\sigma(d)t}) - \ln(1 - \text{ATR}_{\sigma(o)t})$. $\Delta \ln(1 - \text{CITR})$, $\Delta \ln(1 + \text{ITC})$, and $\Delta \ln(1 + \text{RTC})$ are defined analogously. Cluster-robust standard errors are in parentheses.

Table C12: The impact of top inventor inflows on local patent productivity (patent quality).

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(a) All local inventors							
Top inventor inflows	0.070	0.113	0.114	0.116	0.115	0.120	0.122
	(0.010)	(0.024)	(0.022)	(0.022)	(0.027)	(0.024)	(0.023)
$\ln(1-\text{ATR})$	9.840	9.707	9.723	9.897			
	(2.116)	(2.107)	(2.107)	(2.120)			
Effective F statistic		39.877	32.635	32.148	41.100	32.290	31.865
$\tau = 5\%$		37.418	28.588	34.316	37.418	30.199	33.395
$\tau = 10\%$		23.109	17.931	21.047	23.109	18.871	20.549
$\tau = 20\%$		15.062	11.895	13.621	15.062	12.465	13.346
$\tau = 30\%$		12.039	9.614	10.843	12.039	10.047	10.648
CZ FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	No	No	No
State × year FE	No	No	No	No	Yes	Yes	Yes
Observations	23,628	23,628	23,628	23,463	23,562	23,562	23,397
(b) External inventors							
Top inventor inflows	0.062	0.120	0.133	0.137	0.115	0.126	0.133
	(0.014)	(0.030)	(0.024)	(0.024)	(0.034)	(0.029)	(0.029)
$\ln(1-\text{ATR})$	9.581	9.404	9.401	9.515			
	(1.995)	(1.983)	(1.984)	(1.996)			
Effective F statistic		39.877	32.635	32.148	41.100	32.290	31.865
$\tau = 5\%$		37.418	28.626	34.317	37.418	30.048	33.371
$\tau = 10\%$		23.109	17.953	21.047	23.109	18.786	20.535
$\tau = 20\%$		15.062	11.908	13.621	15.062	12.417	13.339
$\tau = 30\%$		12.039	9.623	10.843	12.039	10.012	10.642
CZ FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	No	No	No
State × year FE	No	No	No	No	Yes	Yes	Yes
Observations	23,628	23,628	23,628	23,463	23,562	23,562	23,397

Notes: The coefficient on top inventor inflows is converted to semi-elasticity. ATR stands for the individual income average tax rate at the ninety-fifth percentile. The coefficient on $\ln(1-\text{ATR})$ is converted to elasticity. Column 1 does not control for the endogeneity of top inventor inflows. Column 2 uses B_{dt} as an instrument. Column 3 uses B_{dt} and B^{σ}_{dt} as instruments. Column 4 uses B_{dt}, B^{σ}_{dt}, and B^{ν}_{dt} as instruments. Columns 5-7 replace $\ln(1-\text{ATR})$ in Columns 2-4 with state × year FE. Cluster-robust standard errors are in parentheses.

C.5 Dynamic analysis

Tables C13 and C14 report the estimates from the IV event study regressions specified in (10) and (11), where the first-stage regression is given in (12). Table C13 presents the estimated coefficients $\{\widehat{\mu}^{es}_j\}_{j=-5}^{-2}$ and $\{\widehat{\mu}^{es}_j\}_{j=0}^{+5}$, which correspond to those illustrated in Figure 6. Table C14 reports the associated first-stage statistics.

Table C13: The dynamic impact of top inventor inflows on local patent productivity.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Top inventor inflows ($j = -5$)	-0.023	-0.020	0.001	0.000	-0.020	-0.018	0.004	0.003
	(0.010)	(0.008)	(0.007)	(0.006)	(0.009)	(0.008)	(0.007)	(0.006)
Top inventor inflows ($j = -4$)	-0.019	-0.013	-0.007	-0.007	-0.021	-0.016	-0.008	-0.008
	(0.014)	(0.008)	(0.011)	(0.007)	(0.011)	(0.008)	(0.009)	(0.007)
Top inventor inflows ($j = -3$)	-0.002	-0.001	0.006	0.008	-0.004	-0.001	0.006	0.009
	(0.010)	(0.009)	(0.008)	(0.007)	(0.010)	(0.009)	(0.009)	(0.007)
Top inventor inflows ($j = -2$)	-0.018	-0.014	-0.004	-0.005	-0.018	-0.014	-0.003	-0.004
	(0.010)	(0.007)	(0.009)	(0.006)	(0.008)	(0.006)	(0.008)	(0.006)
Top inventor inflows ($j = 0$)	0.027	0.035	0.033	0.038	0.030	0.037	0.038	0.042
	(0.012)	(0.009)	(0.011)	(0.010)	(0.011)	(0.009)	(0.012)	(0.010)
Top inventor inflows ($j = +1$)	0.046	0.047	0.043	0.050	0.047	0.049	0.047	0.054
	(0.015)	(0.014)	(0.014)	(0.013)	(0.015)	(0.013)	(0.015)	(0.013)
Top inventor inflows ($j = +2$)	0.046	0.045	0.037	0.037	0.045	0.045	0.040	0.042
	(0.013)	(0.013)	(0.012)	(0.012)	(0.013)	(0.012)	(0.012)	(0.011)
Top inventor inflows ($j = +3$)	0.051	0.048	0.044	0.040	0.049	0.048	0.043	0.040
	(0.012)	(0.011)	(0.011)	(0.009)	(0.012)	(0.011)	(0.011)	(0.009)
Top inventor inflows ($j = +4$)	0.027	0.026	0.030	0.030	0.033	0.032	0.036	0.035
	(0.014)	(0.011)	(0.013)	(0.011)	(0.013)	(0.011)	(0.014)	(0.012)
Top inventor inflows ($j = +5$)	0.043	0.045	0.045	0.049	0.054	0.055	0.057	0.060
	(0.013)	(0.013)	(0.013)	(0.012)	(0.013)	(0.013)	(0.014)	(0.013)
Observations	15,752	15,642	15,752	15,642	15,752	15,642	15,752	15,642

Notes: Odd number columns use B_{dt} and B^{σ}_{dt} as instruments. Even number columns use B_{dt}, B^{σ}_{dt}, and B^{ν}_{dt} as instruments. Columns 1-4 use ATR (the individual income average tax rate at the ninety-fifth percentile), whereas Columns 5-8 use ATR99 (the individual income average tax rate at the ninety-ninth percentile) as robustness checks. The dependent variable in Columns 1, 2, 5, and 6 is the patent productivity for all local inventors, whereas that in Columns 3, 4, 7, and 8 is the patent productivity for external inventors. Cluster-robust standard errors are in parentheses.

Table C14: First-stage statistics for the dynamic analysis.

	(1)	(2)	(3)	(4)
(a) Sanderson-Windmeijer (under identification)				
Top inventor inflows ($j = -5$)	340.216	575.561	391.350	671.175
Top inventor inflows ($j = -4$)	156.024	284.008	160.880	248.067
Top inventor inflows ($j = -3$)	177.020	308.252	217.101	325.561
Top inventor inflows ($j = -2$)	163.037	223.759	211.554	254.252
Top inventor inflows ($j = 0$)	257.894	446.932	204.805	397.033
Top inventor inflows ($j = +1$)	442.020	477.121	372.415	463.968
Top inventor inflows ($j = +2$)	208.149	421.991	194.555	381.307
Top inventor inflows ($j = +3$)	255.615	465.129	277.758	503.246
Top inventor inflows ($j = +4$)	126.082	266.307	130.618	229.608
Top inventor inflows ($j = +5$)	271.808	417.144	216.519	338.164
(b) Sanderson-Windmeijer (weak instruments)				
Top inventor inflows ($j = -5$)	30.805	27.280	35.435	31.812
Top inventor inflows ($j = -4$)	14.127	13.461	14.567	11.758
Top inventor inflows ($j = -3$)	16.028	14.610	19.658	15.431
Top inventor inflows ($j = -2$)	14.762	10.606	19.155	12.051
Top inventor inflows ($j = 0$)	23.351	21.183	18.544	18.818
Top inventor inflows ($j = +1$)	40.023	22.614	33.721	21.991
Top inventor inflows ($j = +2$)	18.847	20.001	17.616	18.073
Top inventor inflows ($j = +3$)	23.145	22.046	25.150	23.852
Top inventor inflows ($j = +4$)	11.416	12.622	11.827	10.883
Top inventor inflows ($j = +5$)	24.611	19.771	19.605	16.028

Notes: The Sanderson-Windmeijer tests are used for assessing our instruments since the IV event study model involves multiple endogenous variables (Sanderson and Windmeijer, 2016). Panel (a) presents the chi-squared test statistics for under-identification, while Panel (b) presents the F-test statistics for weak instruments. Odd number columns use B_{dt} and B^{σ}_{dt} as instruments. Even number columns use B_{dt}, B^{σ}_{dt}, and B^{ν}_{dt} as instruments. Columns 1 and 2 use ATR (the individual income average tax rate at the ninety-fifth percentile), whereas Columns 3 and 4 use ATR99 (the individual income average tax rate at the ninety-ninth percentile) as robustness checks. In all specifications, the test statistic for each endogenous variable exceeds 10, which is the commonly used rule-of-thumb value (Staiger and Stock, 1997).

Appendix D State tax competition

We examine the possibility of strategic interactions among state governments by estimating a reaction function, where the income tax in one state responds to the income taxes in other states as follows (see, e.g., Brueckner, 2003):

$$\tau_{\sigma t} = \rho \sum_{\sigma' \neq \sigma} \omega_{\sigma\sigma'} \tau_{\sigma' t} + \boldsymbol{\beta} \mathbf{X}_{\sigma t-1} + \chi_\sigma + \chi_t + \varepsilon_{\sigma t}. \tag{20}$$

The dependent variable $\tau_{\sigma t}$ is the tax rate in state σ in year t, and the term $\sum_{\sigma' \neq \sigma} \omega_{\sigma\sigma'} \tau_{\sigma' t}$ on the right-hand side is the weighted sum of the tax rates in the neighboring states with the weight being $\omega_{\sigma\sigma'}$. $\mathbf{X}_{\sigma t-1}$ denotes a vector of socio-politico-economic characteristics for state σ in year $t-1$ (see Table D1).[43] χ_σ and χ_t stand for state and year fixed effects, respectively, and $\varepsilon_{\sigma t}$ is the error term.

Table D1: Summary statistics (state tax competition).

	mean	sd	min	max	obs
log patent productivity	6.285	1.422	2.881	10.201	1,617
log population	15.011	1.009	12.951	17.430	1,617
share of black or African American population	0.114	0.121	0.002	0.705	1,617
share of population younger than 20	0.292	0.028	0.205	0.414	1,617
share of population older than 64	0.124	0.018	0.075	0.184	1,617
unemployment rate	5.919	2.046	2.342	17.350	1,617
log total income	18.015	1.185	15.010	21.186	1,617
log gross state product	18.191	1.170	15.106	21.403	1,617
democrat governor (dummy variable)	0.511	0.500	0.000	1.000	1,617
log tax revenue	15.291	1.159	12.318	18.581	1,584
log government debt	15.302	1.253	10.872	18.819	1,584
log government revenue	16.110	1.134	13.301	19.511	1,584

Notes: Summary statistics are based on the data described in Appendix A.1 for the years 1977 to 2009.

Our null hypothesis is that $\rho = 0$. If $\rho \neq 0$ were to hold, there would be strategic tax competition between state governments, which would induce interstate correlation between taxes and productivity. In that case, the exclusion restrictions would be violated.

However, measuring ρ is challenging because state tax decisions are simultaneous. To address the potential endogeneity problem that the main regressor $\sum_{\sigma' \neq \sigma} \omega_{\sigma\sigma'} \tau_{\sigma' t}$ is correlated with the error term $\varepsilon_{\sigma t}$, we follow Kelejian and Prucha (1998) and use the weighted sum of neighboring states' socio-politico-economic characteristics as an instrument. Let $\mathbf{X}_{\sigma t-1} =$

[43] When the state governments choose a tax rate, they observe the information on the socio-politico-economic conditions in the previous year. We thus use the lagged variables $\mathbf{X}_{\sigma t-1}$.

$[x_{1,\sigma t-1}, \cdots, x_{K,\sigma t-1}]'$, where $x_{k,\sigma t-1}$ is the k-th characteristic. We first generate the weighted sum $\sum_{\sigma' \neq \sigma} \omega_{\sigma\sigma'} x_{k,\sigma' t-1}$ for each characteristic $\{x_{k,\sigma t-1}\}_{k=1}^{K}$. These K instruments are then used to estimate the predicted value of the weighted sum of neighboring states' tax rates, $\sum_{\sigma' \neq \sigma} \omega_{\sigma\sigma'} \tau_{\sigma' t}$, in the first-stage regression.

Estimating $\omega_{\sigma\sigma'}$ is difficult due to lack of degrees of freedom. We thus consider several alternative weights and examine the sensitivity of the estimates. In the baseline cases, we use the following two types of weights. One is the first-order contiguity weight: $\omega_{\sigma\sigma'} = 1$ if state σ' is contiguous with state σ and $\omega_{\sigma\sigma'} = 0$ otherwise. The other is constructed such that it is proportional to top inventor flows from state σ' to state σ.[44]

Table D2 reports the regression results for the sample period 1977-2009.[45] Columns 1-2 (Columns 3-4) use the first-order contiguity (top-inventor-inflow) weights, where Columns 2 and 4 include the lagged own-state patent productivity. In all cases, the estimated values of ρ are not significantly different from zero at the conventional 5% level. These results imply that there is neither strategic tax competition between states nor a direct tax response to previous patent productivity within each state. We further apply a test by Montiel Olea and Pflueger (2013) to each specification in Table D2. The effective F-statistics indicate that in all cases we can reject the null hypothesis of a weak instrument at the conventional level.

We check the robustness of the results using the specification curve analysis as in Simonsohn et al. (2020). We consider different specifications of equation (20) by focusing on three dimensions. First, we include three other weights discussed in footnote 44 in the specification curve analysis. Second, we estimate the model for every possible combination of the variables listed in Table D1. Last, since the model is overidentified, we use different combinations of neighboring states' characteristics as instruments, which allows us to explore the sensitivity of the estimates.[46]

[44] In the specification curve analysis below, we consider three other weights: (i) the second-order contiguity weight, i.e., $\omega_{\sigma\sigma'} = 1$ if state σ' is contiguous with state σ or is contiguous with the states that are contiguous with state σ; (ii) the inverse-distance weight, i.e., $\omega_{\sigma\sigma'}$ is inversely related to the geographical distance between σ and σ'; and (iii) the inverse-distance weight with a cutoff distance, i.e., the interstate effect is assumed to be zero beyond 1000 miles. When estimating (20), all weights are normalized such that $\sum_{\sigma' \neq \sigma} \omega_{\sigma\sigma'} = 1$ for any σ.

[45] We proxy $\tau_{\sigma t}$ by the ATR at the ninety-fifth percentile in state σ in year t. We drop Washington D.C. from the sample because some state characteristics are unavailable.

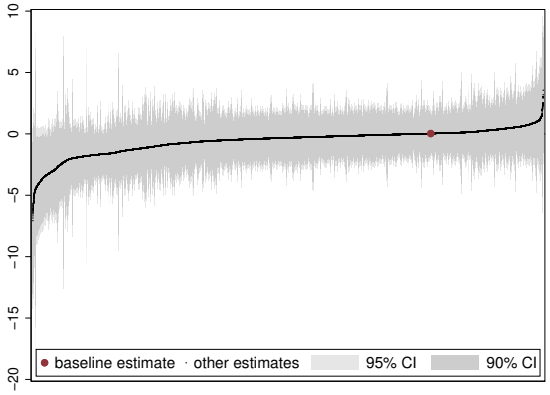
[46] Since the usual caveat on weak instruments is applicable here, we adopt only specifications for which the null hypothesis of weak instruments is rejected.

Table D2: State tax competition.

	(1)	(2)	(3)	(4)
$\sum_{\sigma' \neq \sigma} \omega_{\sigma\sigma'} \tau_{\sigma' t}$	-0.163	0.027	-1.364	-1.071
	(0.314)	(0.260)	(0.822)	(0.797)
log patent productivity		0.005		0.005
		(0.003)		(0.003)
log population	0.016	0.016	0.017	0.013
	(0.038)	(0.033)	(0.029)	(0.029)
share of black or African American population	0.193	0.236	0.111	0.165
	(0.107)	(0.117)	(0.120)	(0.124)
share of population younger than 20	-0.175	-0.160	-0.192	-0.167
	(0.179)	(0.168)	(0.165)	(0.160)
share of population older than 64	0.024	0.029	0.051	0.088
	(0.287)	(0.249)	(0.222)	(0.212)
unemployment rate	-0.000	-0.000	-0.000	-0.000
	(0.001)	(0.001)	(0.001)	(0.000)
log total income	0.003	-0.005	-0.009	-0.013
	(0.032)	(0.031)	(0.028)	(0.029)
log gross state product	0.029	0.028	0.036	0.036
	(0.015)	(0.016)	(0.014)	(0.014)
democrat governor	0.001	0.002	0.001	0.002
	(0.001)	(0.001)	(0.001)	(0.001)
log tax revenue	-0.024	-0.027	-0.023	-0.025
	(0.012)	(0.011)	(0.011)	(0.010)
log government debt	0.000	-0.000	-0.000	-0.000
	(0.002)	(0.002)	(0.002)	(0.002)
log government revenue	-0.002	0.001	-0.002	-0.000
	(0.004)	(0.003)	(0.004)	(0.003)
Effective F statistic	15.924	18.517	50.742	50.307
$\tau = 5\%$	26.392	26.689	25.159	25.143
$\tau = 10\%$	15.173	15.368	14.444	14.355
$\tau = 20\%$	9.191	9.320	8.745	8.639
$\tau = 30\%$	7.055	7.156	6.715	6.612
Observations	1,584	1,584	1,584	1,584

Notes: Columns 1-2 (Columns 3-4) use the first-order contiguity (top-inventor-inflow) weights. Columns 2 and 4 include the lagged own-state patent productivity $\ln Y_{\sigma t-1}$.

Figure D1: Specification curve (state tax competition).



Notes: The specification curve is depicted using 30,843 alternative specifications, as explained in Appendix D. The vertical axis is the value of $\widehat{\rho}$.

Figure D1 plots the specification curve for $\widehat{\rho}$ using $30,843$ alternative specifications. The figure shows that the great majority of the estimated coefficients are not significantly different from zero at the 5% level. Hence, we do not find strong evidence of strategic tax competition between states during the sample period.

Appendix E Share exogeneity

E.1 Two main assumptions to ensure share exogeneity

We elaborate on the two main assumptions to ensure the share exogeneity. Recall that we let $X_{dt} = \{\tau_{\sigma(d)t}, \delta_d, \delta_t\}$ in the main analysis, where $\tau_{\sigma(d)t}$ is the tax rate in state $\sigma(d)$ in year t, δ_d is the destination commuting zone fixed effect, and δ_t is the year fixed effect. We can then rewrite (8) and (9) as

$$\ln Y_{dt} = \phi^s M_{dt} + \xi^s \tau_{\sigma(d)t} + \delta_d + \delta_t + \varepsilon^s_{dt} \tag{21}$$

$$M_{dt} = \psi^f B_{dt} + \xi^f \tau_{\sigma(d)t} + \delta_d + \delta_t + \varepsilon^f_{dt}, \tag{22}$$

where ε^s_{dt} and ε^f_{dt} are error terms.

We impose the following two assumptions to ensure the share exogeneity, $E(\varepsilon^s_{dt}\widehat{P}_{odt}|X_{dt}) = 0$, for the IV regression (21) and (22). First, ε^s_{dt} is mean zero conditional on X_{dt}, i.e.,

$$E(\varepsilon^s_{dt}|\tau_{\sigma(d)t}, \delta_d, \delta_t) = 0. \tag{23}$$

This relies on $E(\varepsilon^s_{dt}) = 0$ and $\varepsilon^s_{dt} \perp \{\tau_{\sigma(d)t}, \delta_d, \delta_t\}$, both of which are assumed to hold.[47] Second, ε^s_{dt} and *other state taxes* $\{\tau_{\sigma(c)t}\}_{c \notin \sigma(d)}$ are independent conditional on X_{dt}, i.e.,[48]

$$\varepsilon^s_{dt} \perp \tau_{\sigma(c)t} | \{\tau_{\sigma(d)t}, \delta_d, \delta_t\} \quad \text{for} \quad c \notin \sigma(d) \quad \text{and} \quad c, d \in \mathcal{C}, \tag{24}$$

which is the exclusion restriction when using B_{dt} as a Bartik instrument for the endogenous variable M_{dt}. This assumption implies that, once we account for the local tax rate, the local commuting zone fixed effect, and the year fixed effect, the local patent productivity shock should be unaffected by the tax policies of other states. This would be satisfied in a situation where local patent productivity primarily responds to its own local factors and conditions.

[47] $\mathfrak{A} \perp \mathfrak{B}$ denote the independence of $\mathfrak{A}$ and $\mathfrak{B}$.

[48] $\mathfrak{A} \perp \mathfrak{B}|\mathfrak{C}$ denote the independence of $\mathfrak{A}$ and $\mathfrak{B}$ conditional on $\mathfrak{C}$.

Under these assumptions, we can show that the share exogeneity, $E(\varepsilon^s_{dt}\widehat{P}_{odt}|\tau_{\sigma(d)t},\delta_d,\delta_t) = 0$, holds as follows. By the property of conditional expectations, we have

$$E(\varepsilon^s_{dt}\widehat{P}_{odt}|\tau_{\sigma(d)t},\delta_d,\delta_t) = E[E(\varepsilon^s_{dt}\widehat{P}_{odt}|\{\tau_{\sigma(c)t}\}_{\forall c\in\mathcal{C}},\delta_d,\delta_t)|\tau_{\sigma(d)t},\delta_d,\delta_t].$$

The right-hand side of the above equation becomes:

$$\begin{aligned}
& E[E(\varepsilon^s_{dt}\widehat{P}_{odt}|\{\tau_{\sigma(c)t}\}_{\forall c\in\mathcal{C}},\delta_d,\delta_t)|\tau_{\sigma(d)t},\delta_d,\delta_t] \\
= \; & E[\widehat{P}_{odt}E(\varepsilon^s_{dt}|\{\tau_{\sigma(c)t}\}_{\forall c\in\mathcal{C}},\delta_d,\delta_t)|\tau_{\sigma(d)t},\delta_d,\delta_t] \\
= \; & E[\widehat{P}_{odt}E(\varepsilon^s_{dt}|\tau_{\sigma(d)t},\delta_d,\delta_t,\{\tau_{\sigma(c)t}\}_{c\notin\sigma(d)})|\tau_{\sigma(d)t},\delta_d,\delta_t] \\
= \; & E[\widehat{P}_{odt}E(\varepsilon^s_{dt}|\tau_{\sigma(d)t},\delta_d,\delta_t)|\tau_{\sigma(d)t},\delta_d,\delta_t] \\
= \; & E[\widehat{P}_{odt}\cdot 0|\tau_{\sigma(d)t},\delta_d,\delta_t] = 0.
\end{aligned}$$

The first equality holds because $\widehat{P}_{odt}$ is a function of $\{\tau_{\sigma(c)t}\}_{\forall c\in\mathcal{C}}$.[49] The second equality holds because $\{\tau_{\sigma(c)t}\}_{\forall c\in\mathcal{C}} = \{\tau_{\sigma(d)t},\{\tau_{\sigma(c)t}\}_{c\notin\sigma(d)}\}$. The third equality is due to the conditional independence assumption (24). The last equality comes from the conditional mean assumption (23).

E.2 Alternative assumptions to ensure share exogeneity

As discussed in Section 4.3, the assumption in (23) may not hold due to a possible correlation between ε^s_{dt} and $\tau_{\sigma(d)t}$ through unobserved state-specific time-varying factors. To alleviate potential concerns that state taxes may respond to local economic conditions or be correlated with local economic policies affecting innovation, we follow Akcigit et al. (2022) and employ alternative specifications with state × year fixed effects. Specifically, we replace $X_{dt} = \{\tau_{\sigma(d)t},\delta_d,\delta_t\}$ with $X'_{dt} = \{\delta_{\sigma(d)t},\delta_d,\delta_t\}$ and consider

$$\ln Y_{dt} = \phi^s M_{dt} + \delta_{\sigma(d)t} + \delta_d + \delta_t + \zeta^s_{dt} \tag{25}$$

$$M_{dt} = \psi^f B_{dt} + \delta_{\sigma(d)t} + \delta_d + \delta_t + \zeta^f_{dt}, \tag{26}$$

where ζ^s_{dt} and ζ^f_{dt} are error terms. We then impose the following two assumptions to obtain the share exogeneity, $E(\zeta^s_{dt}\widehat{P}_{odt}|X'_{dt}) = 0$, for the IV regression (25) and (26). First, ζ^s_{dt} is

[49] One may worry that $\widehat{P}_{odt}$ in (5) depends not only on $\{\tau_{\sigma(c)t}\}_{\forall c\in\mathcal{C}}$ but also on the set of fixed effects $\{\widehat{\gamma}_c,\widehat{\gamma}_{oc}\}_{\forall c\in\mathcal{C}}$ obtained from the log odds regression in (3). We show in Appendix E.3 that, even in that case, a similar procedure can be used to establish the share exogeneity by imposing an additional assumption.

mean zero conditional on X'_{dt}, i.e.,

$$E(\zeta^s_{dt}|\delta_{\sigma(d)t}, \delta_d, \delta_t) = 0.$$

Second, ζ^s_{dt} and *other state taxes* $\{\tau_{\sigma(c)t}\}_{c\notin\sigma(d)}$ are independent conditional on X'_{dt}, i.e.,

$$\zeta^s_{dt} \perp \tau_{\sigma(c)t}|\{\delta_{\sigma(d)t}, \delta_d, \delta_t\} \quad \text{for} \quad c \notin \sigma(d) \quad \text{and} \quad c, d \in \mathcal{C}.$$

Under these assumptions, we can show that the share exogeneity, $E(\zeta^s_{dt}\widehat{P}_{odt}|\delta_{\sigma(d)t}, \delta_d, \delta_t) = 0$, holds in the same way as above.

E.3 Additional assumptions to ensure share exogeneity

One may worry that $\widehat{P}_{odt}$ in (5) depends not only on $\{\tau_{\sigma(c)t}\}_{\forall c\in\mathcal{C}}$ but also on the set of fixed effects $\{\widehat{\gamma}_c, \widehat{\gamma}_{oc}\}_{\forall c\in\mathcal{C}}$ obtained from the log odds regression in (3). Recall that when estimating (3) we do not simultaneously use $\{\widehat{\gamma}_c, \widehat{\gamma}_{oc}\}_{\forall c\in\mathcal{C}}$ (see footnote 12). Since we consider $\{\widehat{\gamma}_{oc}\}_{\forall c\in\mathcal{C}}$ in the baseline specification, we focus on $\{\widehat{\gamma}_{oc}\}_{\forall c\in\mathcal{C}}$ below.[50] We now show that, even if we take into account these fixed effects, a similar procedure as in Appendix E.1 can be used to establish the share exogeneity by imposing an additional assumption.

We assume that the shock ε^s_{dt} is independent of fixed effects γ_{oc} that are specific to each pair of commuting zones conditional on $\{\{\tau_{\sigma(c)t}\}_{\forall c\in\mathcal{C}}, \delta_d, \delta_t\}$:

$$\varepsilon^s_{dt} \perp \gamma_{oc} \,|\{\{\tau_{\sigma(c)t}\}_{\forall c\in\mathcal{C}}, \delta_d, \delta_t\} \quad \text{for } o, c, d \in \mathcal{C}. \tag{27}$$

This would be satisfied in a situation where patent productivity shocks are unrelated to factors affecting inventors' migration costs or firms' relocation costs between origin and destination commuting zones, after controlling for time invariant factors specific to the destination commuting zone and for contemporaneous factors common to all commuting zones.

Using this additional assumption, as well as the conditional mean assumption (23) and the conditional independence assumption (24) in Appendix E.1, we can show that the share exogeneity, $E(\varepsilon^s_{dt}\widehat{P}_{odt}|\tau_{\sigma(d)t}, \delta_d, \delta_t) = 0$, holds as follows. By the property of conditional expectations, we have

$$E(\varepsilon^s_{dt}\widehat{P}_{odt}|\tau_{\sigma(d)t}, \delta_d, \delta_t) = E[E(\varepsilon^s_{dt}\widehat{P}_{odt}|\{\gamma_{oc}\}_{\forall o,c\in\mathcal{C}}, \{\tau_{\sigma(c)t}\}_{\forall c\in\mathcal{C}}, \delta_d, \delta_t)|\tau_{\sigma(d)t}, \delta_d, \delta_t].$$

[50] The following discussion can be readily modified to accommodate the case with $\{\widehat{\gamma}_c\}_{\forall c\in\mathcal{C}}$.

The right-hand side of the above equation becomes:

$$
\begin{aligned}
& E[E(\varepsilon^s_{dt}\widehat{P}_{odt}|\{\gamma_{oc}\}_{\forall o,c\in\mathcal{C}},\{\tau_{\sigma(c)t}\}_{\forall c\in\mathcal{C}},\delta_d,\delta_t)|\tau_{\sigma(d)t},\delta_d,\delta_t] \\
= \; & E[\widehat{P}_{odt}E(\varepsilon^s_{dt}|\{\gamma_{oc}\}_{\forall o,c\in\mathcal{C}},\{\tau_{\sigma(c)t}\}_{\forall c\in\mathcal{C}},\delta_d,\delta_t)|\tau_{\sigma(d)t},\delta_d,\delta_t] \\
= \; & E[\widehat{P}_{odt}E(\varepsilon^s_{dt}|\{\tau_{\sigma(c)t}\}_{\forall c\in\mathcal{C}},\delta_d,\delta_t)|\tau_{\sigma(d)t},\delta_d,\delta_t] \\
= \; & E[\widehat{P}_{odt}E(\varepsilon^s_{dt}|\tau_{\sigma(d)t},\{\tau_{\sigma(c)t}\}_{c\in\mathcal{C},c\notin\sigma(d)},\delta_d,\delta_t,)|\tau_{\sigma(d)t},\delta_d,\delta_t] \\
= \; & E[\widehat{P}_{odt}E(\varepsilon^s_{dt}|\tau_{\sigma(d)t},\delta_d,\delta_t)|\tau_{\sigma(d)t},\delta_d,\delta_t] \\
= \; & E[\widehat{P}_{odt}\cdot 0|\tau_{\sigma(d)t},\delta_d,\delta_t]=0.
\end{aligned}
$$

The sequence of equalities can be explained as follows: The first equality holds because $\widehat{P}_{odt}$ is a function of $\{\gamma_{oc}\}_{\forall o,c\in\mathcal{C}}$ and $\{\tau_{\sigma(c)t}\}_{\forall c\in\mathcal{C}}$. The second equality is derived from assumption (27). The last two equalities follow from assumptions (24) and (23).

Appendix F Decomposition of the Bartik estimator

To assess the validity of the Bartik estimator, we follow Goldsmith-Pinkham et al. (2020) and decompose it into the weighted sum of just-identified instrumental variable estimators, $\widehat{\phi}^s = \sum_{o\in\mathcal{C}}\widehat{\omega}_o\widehat{\phi}^s_o$, where $\widehat{\omega}_o$ and $\widehat{\phi}^s_o$ are the Rotemberg weight and the origin-specific productivity effect, respectively.

Table F1 presents the summary statistics.[51] Panel A shows that the Rotemberg weights are positive in almost all cases. Panel B shows that the weights are highly correlated with the variances of the shares $\text{var}(\widehat{P}_o)$, where the variances are taken across destination commuting zones d and years t. Panel C reports origin commuting zones with the top five highest Rotemberg weights. These commuting zones account for 26.3% of the total share.

Panel D provides the heterogeneous effects interpretation of the Bartik instrument. The Bartik estimator can be rewritten as $\widehat{\phi}^s \approx \sum_d \phi^s_d \sum_o \omega_o v_{od}$, where ϕ^s_d is the destination-specific productivity effect and $v_{od}\geq 0$ is defined in the same way as in Proposition 4 in Goldsmith-Pinkham et al. (2020). Since negative Rotemberg weights for some origin commuting zones o may make $\sum_o \omega_o v_{od}$ negative, the Bartik estimator $\widehat{\phi}^s$ may become a *nonconvex* combination. However, since Panel D shows that the positive part $\sum_{o|\widehat{\omega}_o>0}\widehat{\omega}_o\widehat{\phi}^s_o$ is much larger than the

[51] Since the decomposition is applicable to a single estimator, we focus on $B_{dt}=\sum_{o\neq d}\widehat{P}_{odt}I_{ot}$.

negative part $\sum_{o|\widehat{\omega}_o<0} \widehat{\omega}_o \widehat{\phi}_o^s$, the negative Rotemberg weights are unlikely to be a problem for the LATE (local average treatment effect)-like interpretation of the productivity effect.

We further assess the validity of the identification assumption. For each of the top five *origin* commuting zones in Panel C of Table F1, Table F2 lists the top five *destination* commuting zones by the predicted top inventor migration probability. The result that origin and destination states differ in almost all cases is line with the assumption that the main source of identifying variation comes from interstate top inventor migrations induced by individual income tax differences across states.

We finally examine the relationship between the shares $\widehat{P}_{odt}$ associated with top five origin commuting zones on the one hand and the location-specific characteristics that may be correlated with the outcome Y_{dt} on the other hand as suggested by Goldsmith-Pinkham et al. (2020). For the shares $\widehat{P}_{odt}$ to satisfy the share exogeneity assumption in Section 4.3, they should not be correlated with destination commuting zone characteristics. In our analysis, we use the log of lagged employment in four sectors—"manufacturing," "finance and insurance," "professional, scientific, and technical services," and "management of companies and enterprises"—in commuting zone d as such characteristics.

Table F3 presents the results of this analysis, where we take the time difference of the variables to control for commuting zone fixed effects. The first panel reports the correlation between the shares related to the highest-weight origins and the log of lagged sectoral employment, which shows that the correlations are low in all cases. The second panel reports the coefficients from regressing the shares related to the highest-weight origins on the lagged sectoral employment while controlling for year fixed effects. The coefficients are not statistically significant at the conventional 5% level.[52] Hence, the results in Table F3 suggest that there is no compelling evidence of a significant correlation between the shares related to the highest-weight origins and the log of lagged employment across sectors and commuting zones.

[52]The absence of significance at the 5% level does not necessarily imply that the true coefficient is zero. Indeed, in one commuting zone, the null hypothesis is marginally rejected at the 10% level for some sectors. However, since our analysis involves multiple regressions—each estimating the effect of lagged sectoral employment on the shares related to the highest-weight origin, the rejection of the null hypothesis, if any, might be an artifact of multiple hypothesis testing rather than a genuine effect. To address spurious rejections arising from multiple hypothesis testing, we employ two approaches. First, we report p-values adjusted to control the family-wise error rate (FWER) using the Romano-Wolf multiple hypothesis correction method. Second, we compute the sharpened false discovery rate (FDR) q-values using the code provided by Anderson (2008).

Table F1: Summary of the decomposition of the Bartik estimator.

Panel A: Negative and positive weights

	sum	mean	share
Negative	-0.0002	-0.0000	0.0002
Positive	1.0002	0.0099	0.9998

Panel B: Correlations

	$\widehat{\omega}_o$	$\overline{I}_o$	$\widehat{\phi}_o^s$	$\widehat{F}_o$	var($\widehat{P}_o$)
$\widehat{\omega}_o$	1.0000				
$\overline{I}_o$	0.9204	1.0000			
$\widehat{\phi}_o^s$	-0.0478	-0.0365	1.0000		
$\widehat{F}_o$	0.0372	-0.0603	-0.0268	1.0000	
var($\widehat{P}_o$)	0.7865	0.7801	-0.0417	-0.0845	1.0000

Panel C: Top five Rotemberg weight commuting zones

		$\widehat{\omega}_o$	$\overline{I}_o$	$\widehat{\phi}_o^s$	95% CI
19600	Bergen-Essex-Middlesex (NJ)	0.0838	186	0.0507	(0.033, 0.097)
24300	Cook-DuPage-Lake (IL)	0.0535	113	0.0617	(0.037, 0.170)
19400	Kings-Queens-New York (NY)	0.0448	63	0.0402	(0.011, 0.065)
19700	Philadelphia-Montgomery-Delaware (PA)	0.0414	68	0.0512	(0.033, 0.170)
16300	Allegheny-Westmoreland-Washington (PA)	0.0396	48	0.0358	(0.018, 0.059)

Panel D: Estimates of $\widehat{\phi}_o^s$ for positive and negative weights

	$\widehat{\omega}$-weightd sum	share of overall $\widehat{\phi}^s$	mean
Negative	0.0001	0.0023	-0.5575
Positive	0.0417	0.9977	0.0478

Notes: Panel A presents the sum, mean, and share of the positive and negative Rotemberg weights. Panel B reports correlations between the Rotemberg weights ($\widehat{\omega}_o$), the number of top inventors ($\overline{I}_o = (1/|\mathcal{S}|)\sum_{t\in\mathcal{S}} I_{ot}$), the just-identified coefficient estimates ($\widehat{\phi}_o^s$), the first stage F-statistic of the share ($\widehat{F}_o$), and the variance of the shares across destinations and years (var($\widehat{P}_o$)). Panel C reports the origin commuting zones with the top five highest Rotemberg weights. The state of the representative county of each commuting zone is in parentheses, where the representative county is the one with the largest number of inventors. The 95% confidence interval is the weak instrument robust confidence interval as in Chernozhukhov and Hansen (2008) over a range from 0 to 0.5. In Panel D "$\widehat{\omega}$-weighted sum" reports $\sum_{o|\widehat{\omega}_o<0} \widehat{\omega}_o\widehat{\phi}_o^s$ for negative and $\sum_{o|\widehat{\omega}_o>0} \widehat{\omega}_o\widehat{\phi}_o^s$ for positive cases, and "share of overall $\widehat{\phi}^s$" reports $(1/\widehat{\phi}^s)\sum_{o|\widehat{\omega}_o<0} \widehat{\omega}_o\widehat{\phi}_o^s$ for negative and $(1/\widehat{\phi}^s)\sum_{o|\widehat{\omega}_o>0} \widehat{\omega}_o\widehat{\phi}_o^s$ for positive cases.

Table F2: Destinations to which top inventors migrated from the highest-weight origins.

	predicted migration probability	state of the representative county
(A) 19600 Bergen-Essex-Middlesex (the highest weight)		NJ
38000 San Diego	0.0078	CA
37500 Santa Clara-Monterey-Santa Cruz	0.0073	CA
39400 King-Pierce-Snohomish	0.0071	WA
35801 Ada-Canyon-Elmore	0.0070	ID
37800 Alameda-Contra Costa-San Francisco	0.0069	CA
(B) 24300 Cook-DuPage-Lake (the second highest weight)		IL
7100 Palm Beach-St. Lucie-Martin	0.0113	FL
9100 Fulton-DeKalb-Cobb	0.0091	GA
11304 Fairfax-Montgomery-Prince George's	0.0081	MD
37000 Stanislaus-Merced-Tuolumne	0.0079	CA
37500 Santa Clara-Monterey-Santa Cruz	0.0079	CA
(C) 19400 Kings-Queens-New York (the third highest weight)		NY
37500 Santa Clara-Monterey-Santa Cruz	0.0118	CA
18600 Albany-Saratoga-Rensselaer	0.0114	NY
37800 Alameda-Contra Costa-San Francisco	0.0111	CA
38801 Multnomah-Washington-Clackamas	0.0100	OR
39400 King-Pierce-Snohomish	0.0097	WA
(D) 19700 Philadelphia-Montgomery-Delaware (the fourth highest weight)		PA
39400 King-Pierce-Snohomish	0.0151	WA
28900 Denver-Jefferson-Arapahoe	0.0143	CO
38000 San Diego	0.0131	CA
38801 Multnomah-Washington-Clackamas	0.0116	OR
37800 Alameda-Contra Costa-San Francisco	0.0113	CA
(E) 16300 Allegheny-Westmoreland-Washington (the fifth highest weight)		PA
37500 Santa Clara-Monterey-Santa Cruz	0.0217	CA
37800 Alameda-Contra Costa-San Francisco	0.0211	CA
38300 Los Angeles-Orange-San Bernardino	0.0195	CA
21501 Hennepin-Ramsey-Dakota	0.0180	MN
7000 Dade-Broward-Monroe	0.0175	FL

Notes: The top five destination commuting zones are defined by the predicted migration probability of the top inventors from each origin commuting zone given in the first column. The second column reports the state of the representative county of each commuting zone, where the representative county is the one with the largest number of inventors.

Table F3: Relationship between the shares and the lagged employment.

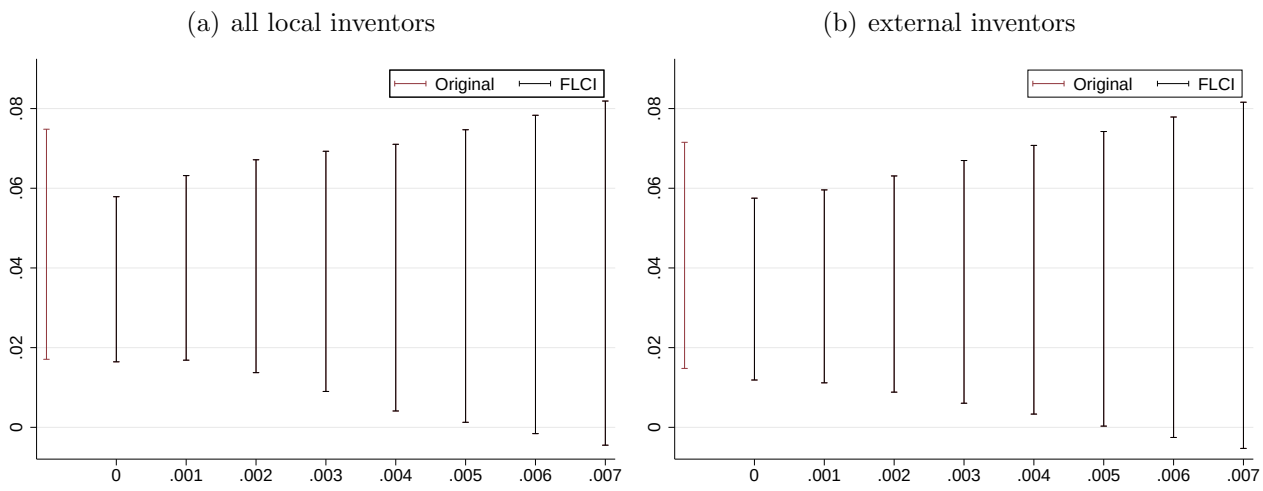
	share: the predicted probability of top inventors from				
	19600 Bergen-Essex-Middlesex	24300 Cook-DuPage-Lake	19400 Kings-Queens-New York	19700 Philadelphia-Montgomery-Delaware	16300 Allegheny-Westmoreland-Washington
Correlation coefficients					
Manufacturing	-0.007549	0.003165	-0.005326	-0.002585	-0.006714
Finance and insurance	-0.002242	-0.005738	-0.000737	0.000201	-0.008787
Professional, scientific, and technical services	-0.002235	-0.002048	-0.001205	-0.001759	-0.000779
Management of companies and enterprises	0.002552	-0.001077	-0.000365	0.000917	-0.001256
Regression coefficients					
Manufacturing	-0.000007	0.000012	-0.000011	0.000005	-0.000017
	(0.000011)	(0.000014)	(0.000012)	(0.000010)	(0.000010)
	[0.542358]	[0.401073]	[0.368871]	[0.648052]	[0.094328]
	{0.989000}	{0.615400}	{0.989000}	{1.000000}	{0.347700}
	<1.000000>	<1.000000>	<1.000000>	<1.000000>	<1.000000>
Finance and insurance	-0.000013	-0.000043	-0.000010	-0.000001	-0.000078
	(0.000018)	(0.000031)	(0.000021)	(0.000022)	(0.000042)
	[0.460584]	[0.174263]	[0.647973]	[0.963237]	[0.060422]
	{0.965000}	{0.615400}	{0.989000}	{1.000000}	{0.200800}
	<1.000000>	<1.000000>	<1.000000>	<1.000000>	<1.000000>
Professional, scientific, and technical services	0.000000	-0.000009	0.000001	-0.000001	0.000005
	(0.000007)	(0.000009)	(0.000007)	(0.000007)	(0.000010)
	[0.950435]	[0.281902]	[0.882331]	[0.838880]	[0.574786]
	{1.000000}	{0.615400}	{1.000000}	{1.000000}	{0.200800}
	<1.000000>	<1.000000>	<1.000000>	<1.000000>	<1.000000>
Management of companies and enterprises	0.000002	-0.000002	-0.000000	0.000001	-0.000002
	(0.000002)	(0.000003)	(0.000001)	(0.000002)	(0.000002)
	[0.178380]	[0.544422]	[0.872453]	[0.386848]	[0.344007]
	{0.615400}	{0.989000}	{1.000000}	{1.000000}	{0.200800}
	<1.000000>	<1.000000>	<1.000000>	<1.000000>	<1.000000>

Notes: The first panel reports the correlation coefficient between the shares related to the highest-weight origin and the log of lagged sectoral employment. The second panel reports the coefficients obtained from regressing the shares related to the highest-weight origin on the lagged sectoral employment. The parentheses contain the conventional standard errors, while the square brackets show the p-values. The braces present the family-wise error rate (FWER) adjusted p-values, computed by using the Romano-Wolf multiple hypothesis correction method, and the angle brackets present the sharpened false discovery rate (FDR) q-values, calculated by using the code provided by Anderson (2008).

Appendix G Robustness to possible violations of the parallel trends assumption

We examine the robustness to possible violations of the parallel trends assumption for the event study analysis in Section 5. Following Rambachan and Roth (2023) we specify, for each event study regression, a set $\Upsilon^{\mathrm{SD}}(M) = \{\upsilon : |(\upsilon_{t+1} - \upsilon_t) - (\upsilon_t - \upsilon_{t-1})| \leq M\}$ to bound the degree to which the slope of differential trend υ can vary between consecutive periods. We use the default setting in the R package `HonestDiD` provided by Rambachan and Roth (2023), namely that the value of M ranges from 0, which corresponds to a linear trend, to half a standard deviation of the parameter of interest.

Figure G1: Robustness to possible violations of parallel trends assumption.



Notes: Panels (a) and (b) illustrate the sensitivity analysis for the event study model in Section 5, where we consider all local inventors and external inventors, respectively. In each panel, the leftmost bar is the 95% confidence interval of the estimate $\widehat{\mu}_1^{es}$ and the other bars are fixed length confidence intervals (FLCIs) proposed by Rambachan and Roth (2023).

Figure G1 illustrates the robustness test result, where we evaluate $\Upsilon^{\text{SD}}(M)$ at $t = 0$ by normalizing $\upsilon_0 = 0$ to conduct a sensitivity analysis for the first-period effect $\widehat{\mu}_1^{es}$. In Panel (a) (Panel (b)), the "break-down" value of M, at which the null hypothesis that the first-period effect is zero can no longer be rejected, is 0.006 (0.006). Since the value is 40.00% (42.90%) of the standard error of the estimated effect $\widehat{\mu}_1^{es}$, the parallel trends assumption holds for a reasonable deviation from a linear trend.[53]

We further perform a permutation-based placebo analysis to see how likely the "break-down" value of $M = 0.006$ is to occur. Specifically, we first estimate each event study model 200 times by randomly reshuffling the commuting zones to which top inventors moved and then apply the sensitivity analysis proposed by Rambachan and Roth (2023) mentioned above to the estimates.

Table G1 shows the summary statistics of the simulated "break-down" values of M. In both Panels (a) and (b), the values are zero in 95% of cases. These results indicate that the deviation from the linear trend up to $M = 0.006$ occurs extremely rarely. Therefore, we may conclude that the parallel trends assumption is unlikely to be violated.

[53]We examine the parallel trends assumption for the IV event study regression using B_{dt} and B_{dt}^{σ} as instruments, which corresponds to the IV ES1 case presented in Figure 5. The standard error of the first-period effect $\widehat{\mu}_1^{es}$ for all local inventors (for external inventors) is 0.015 (0.014).

Table G1: Summary statistics of the "break-down" values for the placebo simulations.

	95th percentile	99th percentile
(a) all local inventors	0.0000	0.0010
(b) external inventors	0.0000	0.0010

Notes: Panels (a) and (b) present the distribution of the "break-down" values of M obtained from the permutation-based placebo analysis for the event study model, where we consider all local inventors and external inventors, respectively. The "break-down" value is defined as the value of M at which the null hypothesis that the first period effect is zero can no longer be rejected. The summary statistics are for 200 simulation results.

Appendix H Falsification: The case of top baseball players

In this section, we first estimate the effect of state income tax differences on top MLB player migrations to quantify how location decisions of high-income individuals other than top inventors are affected by tax incentives. We then verify that the tax-induced migration of top MLB players does not affect local patent productivity. This falsification test allows us to isolate the causal effect of top inventor inflows on local patent productivity from other high-income migrants and mitigate potential threats to internal validity, thereby showing the robustness of our main findings.

Institutional Background. MLB players have limited opportunities for team selection. The most significant avenue to team selection for experienced players is free agency. After accruing six years of MLB service time, players become eligible for free agency, which allows them to maximize their market value and pursue optimal financial and career opportunities by negotiating contracts with any team. Free agency serves as a key mechanism for the migration of MLB players. We assume that MLB players are subject to the income tax rate of their home team's state. While this assumption does not fully capture the complexity of players' actual tax environment, it provides a reasonable approximation for examining tax differentials across team locations and their potential influence on players' migrations.[54]

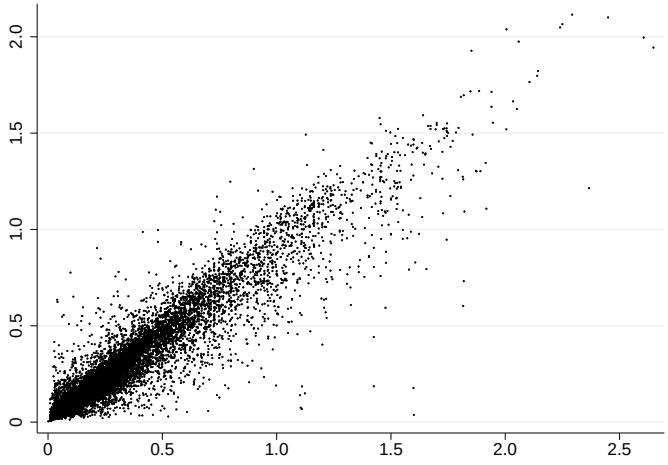
Data. Our analysis uses data from multiple sources to compile information on MLB players and team attributes. The primary source is the Sean Lahman Baseball Database, from which

[54]MLB players are in a complex tax environment, which is primarily determined by two key concepts: duty days and the jock tax. The former represent the total number of work days in a season, including games, practices, and travel. The latter allows states and municipalities to levy taxes on athletes for income earned from performing services within their jurisdictions. Thus, the players potentially owe taxes to multiple locations based on their play and travel schedule throughout the season. However, since the number of home games equals that of away games, the fraction of games played at each away team's location is relatively small.

we extract data on players' positions, ages, salaries, past awards, and team affiliations, as well as on teams' annual performance metrics, home locations, and stadium capacities.[55] For more granular player performance data, we employ wins above replacement (WAR), which is a metric that captures each player's total contribution to his team.[56] To identify free agent declarations, we use the Retrosheet's transactions database, which provides detailed records of trades, contracts, and status changes.[57]

While salary data are available from 1985 onwards, they are not exhaustive (since salary data are available for 62.43% of player-year observations in MLB from 1985 to 2015). To address this limitation, we employ a machine learning approach to predict player salaries for this period. Specifically, we use a random forest algorithm to estimate salaries based on various player characteristics and performance.[58] The random forest model achieves strong predictive performance, with R-squared values of 0.975 for the training set and 0.813 for the test set. Figure H1 presents a scatter plot of actual versus predicted salaries, showing close alignment along the diagonal.

Figure H1: Actual versus predicted salaries.



Notes: The vertical and horizontal axes are predicted salaries and actual salaries in 10 million dollars, deflated by CPI using 2000 as the base year.

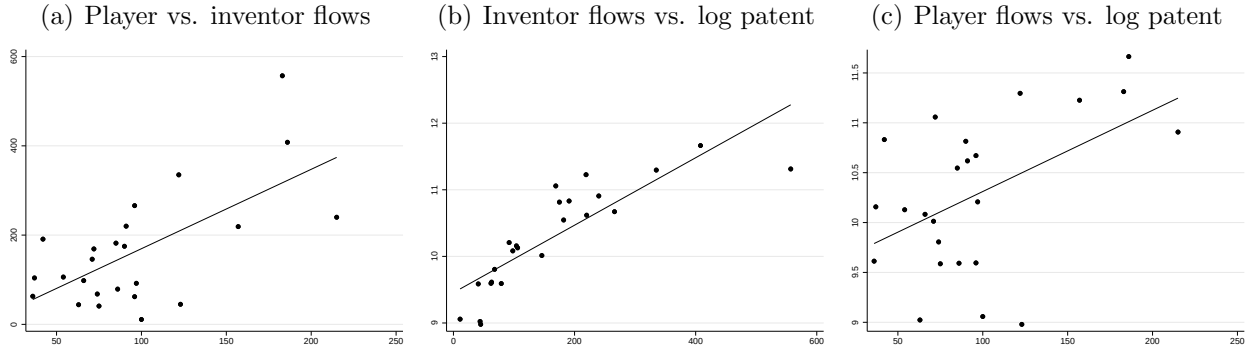
[55]The Sean Lahman Baseball Database is a collection of baseball statistics covering MLB from 1871 to the present, which is accessible at `https://cran.r-project.org/web/packages/Lahman/index.html`

[56]Neil Paine provides the WAR from 1901, which is accessible at `https://github.com/Neil-Paine-1/MLB-WAR-data-historical`.

[57]Tom Ruane provides Retrosheet's transactions database, which covers player, manager, coach, and umpire IDs from 1873 to 2020, at `https://retrosheet.org/transactions/index.html`.

[58]Salaries used in the prediction model are adjusted to 2000 constant dollars to account for inflation. The model includes demographic variables, performance metrics, career milestones, as well as their lagged variables and interaction terms to capture potential non-linear relationships.

Figure H2: Relationship between top inventor flows, top player flows, and local patent productivity.



Notes: The vertical and horizontal axes in Panel (a) are top inventor inflows and top player inflows. The vertical and horizontal axes in Panel (b) are the log number of patents and top inventor inflows. The vertical and horizontal axes in Panel (c) are the log number of patents and top player inflows.

Summary statistics. To maintain consistency with our analysis on top inventor migrations, we restrict ourselves to the period from 1977 to 2009 and the teams in the U.S. by excluding Canadian teams.[59] The resulting dataset consists of 28 teams located in 24 commuting zones, as well as 6,436 unique "top players" defined as players whose (predicted) salaries exceeded the ninety-fifth percentile of the U.S. income distribution in the year they declared free agency after six years of MLB service.[60] We examine the team choices of these top players. Of these top players, 42.6% declared free agency at least once during their careers. To facilitate comparison with the top inventor analysis, we focus on the ATR at the ninety-fifth percentile of the U.S. income distribution. In our sample, 62.54% of free agents meet the top player criterion. Panel (a) of Figure H2 illustrates the relationship between top player flows and top inventor flows to destination commuting zones. Panel (b) (resp., Panel (c)) depicts the relationship between top inventor (resp., player) inflows and local patent productivity in destination commuting zones. The correlation coefficients for Panels (a), (b), and (c) are 0.646, 0.847, and 0.495, respectively.

Table H1 summarizes top 10 commuting zones by top player inflows. Generally, commuting zones attracting more top players also tend to have greater top inventor inflows and more local innovations. However, this correlation does not imply causality between top player inflows and

[59]Specifically, we exclude the Montreal Expos and their successor, the Washington Nationals, as well as the Toronto Blue Jays from our analysis.

[60]The data on the U.S. income distribution are obtained from the World Inequality Database: `https://wid.world/data/`.

Table H1: Top 10 commuting zones by top player inflows.

rank	cz number	counties	state	team(s)	inflows
1	19400	Kings–Queens–New York	NY	NYA, NYN	215
2	38300	Los Angeles–Orange–San Bernardino	CA	CAL, LAN	186
3	37800	Alameda–Contra Costa–San Francisco	CA	OAK, SFN	183
4	24300	Cook–DuPage–Lake	IL	CHA, CHN	157
5	33000	Tarrant–Johnson–Parker	TX	TEX	123
6	20500	Middlesex–Worcester–Essex	MA	BOS	122
7	29502	Jackson–Johnson–Wyandotte	KS	KCA	100
8	15200	Cuyahoga–Summit–Lake	OH	CLE	97
9	24701	St. Louis–St. Clair	MO	SLN	96
10	38000	San Diego	CA	SDN	96

Notes: Inflows are defined as the number of top players who migrated into each commuting zone from 1977 to 2009. We adopt the abbreviations for the teams in the Sean Lahman Baseball Database, which are: BOS for Boston Red Sox, CAL for California Angels, CHA for Chicago White Sox, CHN for Chicago Cubs, CLE for Cleveland Guardians, KCA for Kansas City Royals, LAN for Los Angeles Dodgers, NYA for New York Yankees, NYN for New York Mets, OAK for Oakland Athletics, SDN for San Diego Padres, SFN for San Francisco Giants, SLN for St. Louis Cardinals, and TEX for Texas Rangers.

local patent productivity gains. Rather, it may reflect the tendency of larger commuting zones (such as Kings–Queens–New York and Los Angeles–Orange–San Bernardino) to attract high-income, high-skill individuals in various sectors, including both top inventors and top players.

Counterfactual salaries. Following Kleven et al. (2013), we consider three formulations for counterfactual salaries in the context of professional sports labor markets. We first adopt the "perfect substitution technology" assumption, i.e., perfect competition among players implies that pre-tax salary equals ability, i.e., $w_{ikt} = X_{it}^a$, where X_{it}^a is the ability vector. In this case, the deterministic part of the utility can be rewritten as:

$$V_{ijkt} = \alpha \ln[(1 - \tau_{\sigma(k)t})X_{it}^a] + \gamma^h \text{Home}_{ijkt-1} + \gamma_k^x X_{it} - \gamma^c C_{jk} + Z_k. \tag{28}$$

The second formulation is a variant of the former, given by $w_{ikt} = X_{it}^a \cdot \overline{w}_{kt}$, where $\overline{w}_{kt}$ is the overall salary level at team k in year t approximated by the estimated average salary from a random forest model. We further allow for the possibility that the coefficients on $\ln(1 - \tau_{\sigma(k)t})$, $\ln X_{it}^a$, and $\ln \overline{w}_{kt}$ may differ. In this case, V_{ijkt} can be rewritten as:

$$V_{ijkt} = \alpha^t \ln(1 - \tau_{\sigma(k)t}) + \alpha^w \ln \overline{w}_{kt} + \gamma^h \text{Home}_{ijkt-1} + \alpha^a \ln X_{it}^a + \gamma_k^x X_{it} - \gamma^c C_{jk} + Z_k. \tag{29}$$

The last specification replaces counterfactual salaries with the random forest estimates $\widehat{w}_{ikt}^{rf}$, resulting in the deterministic utility function as follows:

$$V_{ijkt} = \alpha^t \ln(1 - \tau_{\sigma(k)t}) + \alpha^w \ln \widehat{w}_{ikt}^{rf} + \gamma^h \text{Home}_{ijkt-1} + \gamma_k^x X_{it} - \gamma^c C_{jk} + Z_k. \tag{30}$$

Table H2: Multinomial logit regression results.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
ln(1 − ATR)	9.975	9.959	5.485	5.489	5.489	5.489	7.903	7.894	7.894	7.894
	(2.124)	(2.122)	(2.349)	(2.347)	(2.347)	(2.347)	(2.212)	(2.210)	(2.210)	(2.210)
ln(Salary)			0.486	0.484	0.484	0.484	0.379	0.378	0.378	0.378
			(0.100)	(0.100)	(0.100)	(0.100)	(0.117)	(0.117)	(0.117)	(0.117)
Home (team)	1.899	1.748	1.907	1.805	1.805	1.805	1.899	1.751	1.751	1.751
	(0.056)	(0.188)	(0.056)	(0.188)	(0.188)	(0.188)	(0.056)	(0.187)	(0.187)	(0.187)
Home (league)	0.212	0.216	0.207	0.209	0.209	0.209	0.212	0.216	0.216	0.216
	(0.043)	(0.043)	(0.043)	(0.043)	(0.043)	(0.043)	(0.043)	(0.043)	(0.043)	(0.043)
Home (division)	0.117	0.080	0.119	0.093	0.093	0.093	0.118	0.082	0.082	0.082
	(0.045)	(0.052)	(0.045)	(0.053)	(0.053)	(0.053)	(0.045)	(0.052)	(0.052)	(0.052)
Player char.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Player perf.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Team location	No	Yes	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Observations	95,028	95,028	94,503	94,503	94,503	94,503	95,028	95,028	95,028	95,028

Notes: Coefficients are from multinomial logit regressions of top player i's migration from team j to team k in year t. The baseline choice is retirement ($V_{ijkt} = 0$ for $k = 1$). ATR denotes the individual income average tax rate at the ninety-fifth percentile of the U.S. income distribution. Salary represents top players' pre-tax annual earnings. Columns 1-2 assume the perfect substitution technology in salary determination as in equation (28). Columns 3-6 use the estimated average pre-tax salaries for each team as in equation (29). Columns 7-10 adopt the counterfactual salary for each player-team-year combination based on the random forest model as in equation (30). Columns 5 and 9 (resp., Columns 6 and 10) allow α^t (resp., α^w) to vary across players. Home (team), Home (league), and Home (division) are dummy variables indicating whether players' choices match those of the previous year's team, league, or division. Additional controls include player characteristics (age, experience, their squares, and position) and player performance (log of 3-year WAR), as well as team location variables (same state, same census region, and log distance between team $j's$ and $k's$ locations). Team-specific coefficients γ_k^x for player variables are estimated but not reported due to space constraints. All specifications include team fixed effects Z_k. Cluster-robust standard errors are in parentheses.

Estimation results. Table H2 presents the estimation results for α, α^t, α^w, and γ^h from the team choice model for top players. Columns 1-2 assume the perfect substitution technology in salary determination as in (28). Columns 3-6 use the estimated average pre-tax salaries for each team as in (29). Columns 7-10 adopt the counterfactual salary for each player-team-year combination based on the random forest model as in (30). To account for heterogeneity in team selection among top players, we also implement random coefficient specifications. Columns 5 and 9 (resp., Columns 6 and 10) allow α^t (resp., α^w) to vary across top players. All models include team fixed effects Z_k to capture unobservable characteristics of each team. The estimation results consistently show a significantly positive α^w, indicating that top players tend to choose teams offering higher pre-tax salaries. The positive and significant α or α^t suggests that top players prefer teams in locations with lower income tax rates. Furthermore, the positive and significant γ^h confirms the presence of top players' home preferences in team

selection as in Kleven et al. (2013). We use Column 8 of Table H2 to construct the Bartik instrument for the top player inflows M_{dt}^{ply}.

The main IV regression results are reported in Table 6. Given that the IV regressions involve two endogenous variables, we employ the Sanderson-Windmeijer tests to assess the relevance of our instruments (Sanderson and Windmeijer, 2016). The test results are presented in Table H3. In all specifications, the test statistic for each endogenous variable exceeds 10, which is the commonly used rule-of-thumb value (Staiger and Stock, 1997).

Table H3: First-stage statistics for the falsification analysis.

	(1)	(2)	(3)	(4)	(5)	(6)
(a) Sanderson-Windmeijer (under identification)						
Top inventor inflows	23.272	59.351	62.768	27.105	53.047	56.442
Top player inflows	72.869	429.338	425.385	79.685	415.145	422.672
(b) Sanderson-Windmeijer (weak instruments)						
Top inventor inflows	23.206	29.590	20.861	25.323	24.779	17.567
Top player inflows	72.663	214.052	141.378	74.447	193.918	131.552

Notes: The Sanderson-Windmeijer tests are used for assessing our instruments since the IV regression involves two endogenous variables (Sanderson and Windmeijer, 2016). Panel (a) presents the chi-squared test statistics for under-identification, while Panel (b) presents the F-test statistics for weak instruments. Columns 1 and 4 use B_{dt} as an instrument for top inventor inflows M_{dt}. Columns 2 and 5 use B_{dt} and B_{dt}^{σ} as instruments for M_{dt}. Columns 3 and 6 use B_{dt}, B_{dt}^{σ}, and B_{dt}^{ν} as instruments for M_{dt}. In all cases, we use B_{dt}^{ply} as an instrument for top player inflows M_{dt}^{ply}. In all specifications, the test statistic for each endogenous variable exceeds 10, which is the commonly used rule-of-thumb value (Staiger and Stock, 1997).

Appendix I Shift exogeneity

Following Borusyak et al. (2022), we exploit quasi-experimental variation in origin shocks by rewriting the destination-level structural equation (8) in terms of an origin-level equation as follows:

$$\ln \overline{Y}_{ot}^{\perp} = \phi^{s} \overline{M}_{ot}^{\perp} + \delta^{s} q_{ot} + \overline{\varepsilon}_{ot}^{\perp}, \tag{31}$$

where $\overline{v}_{ot}^{\perp} = \sum_d P_{odt} v_d^{\perp} / \sum_d P_{odt}$ denotes the share-weighted average of a variable $v_{dt}^{\perp}$ across all destinations d at time t, and $v_{dt}^{\perp}$ represents the residual from a sample projection of v_{dt} on the control variable vector X_{dt} in the destination-level equation (8).[61] The vector q_{ot} includes appropriate origin-level control variables, which we detail below.

To establish shift-share instrumental variable consistency as in Borusyak et al. (2022), we impose key assumptions as follows. First, we assume that, conditional on some controls, the

[61] We will modify the control variable vector X_{dt} to analyze the case with incomplete shares below.

shifts are quasi-randomly assigned, i.e., $\mathrm{E}\left(I_{ot} \mid \bar{\varepsilon}, q, P\right) = \mu q_{ot}$, where $\bar{\varepsilon} = \left\{\bar{\varepsilon}_{ot}\right\}_{ot}, q = \left\{q_{ot}\right\}_{ot}$, and $P = \left\{P_{odt}\right\}_{odt}$ are vectors of origin-level shocks, observable commuting zone characteristics, and observable migration shares, respectively. We include origin commuting zone fixed effects and year fixed effects in q_{ot} to control for time-invariant commuting-zone-level confounders and common time trends. Our identification strategy is that, after accounting for these factors, the residual variation in the number of top inventors in origin commuting zone o is as-good-as-randomly assigned over time. We examine the validity of this assumption using Table I3 below.

Second, we impose two conditions namely, $\mathrm{E}\big(\sum_{\text{cluster}}\big(\sum_{o\in\text{cluster}}\sum_{d\neq o}P_{odt}\big)^2\big) \to 0$ for all t, and $\mathrm{Cov}\big(\widetilde{I}_{ot}, \widetilde{I}_{o't}|\bar{\varepsilon}, q, P\big) = 0$ for all t and for all o and o' such that $\text{cluster}(o) \neq \text{cluster}(o')$, where cluster stands for regions or states and $\widetilde{I}_{ot} = I_{ot} - \mu q_{ot}$ denotes the residual of the number of top inventors after controlling for origin commuting zone fixed effects and year fixed effects. The former condition implies that the effective sample size, determined by the inverse of the Herfindahl-Hirschman index of the shares, increases asymptotically. The latter ensures that the residuals are mutually uncorrelated between commuting zones such that $\text{cluster}(o) \neq \text{cluster}(o')$. We examine the validity of these assumptions using Tables I1 and I2 below.

Since we consider incomplete shares, i.e., $\sum_{o\neq d}P_{odt} \neq 1$ for each destination d and year t, we include two additional controls in our destination-level specification: the sum of the shares $\sum_{o\neq d}P_{odt}$; and the exposure-weighted sum of the origin controls $\sum_{o\neq d}P_{odt}q_{ot}$ (see Section 4.2 in Borusyak et al., 2022). Thus, to rewrite the shift-level equation as the origin-level equation, we control for the destination-level characteristics $\{X_{dt}, \sum_{o\neq d}P_{odt}, \sum_{o\neq d}P_{odt}q_{ot}\}$. This comprehensive set of controls allows us to address both the incomplete shares and the shift exogeneity.

Table I1 presents summary statistics for the shift-related variables. The first panel shows the distribution of the number of top inventors I_{ot} and the distribution of the residual $\widetilde{I}_{ot}$ after controlling for origin commuting zone fixed effects and year fixed effects. Despite the difference in dispersion—the distribution of I_{ot} exhibiting a larger spread than that of $\widetilde{I}_{ot}$—both demonstrate substantial variability.

Table I1: Summary statistics of shift variables.

statistics	I_{ot}	$\widetilde{I}_{ot}$
Shift distribution		
Mean	47.509	0.000
Standard deviation	95.921	57.757
Interquartile range	43.000	37.256
Effective sample size (1/HHI of P_{ot} weights)		
Across CZs and years	1180.018	
Across CZs	298.514	
Largest P_{ot} weight		
Across CZs and years	0.001	
Across CZs	0.010	
Observations		
Number of CZ-year pairs	2,859	
Number of CZs	391	

Notes: The first panel summarizes the distribution of the number of top inventors I_{ot} and the distribution of the residual $\widetilde{I}_{ot}$ after controlling for origin commuting zone fixed effects and year fixed effects. The second and third panels report the effective sample size, calculated as the inverse of the Herfindahl-Hirschman index of $P_{ot} = \sum_{d \neq o} P_{odt}$ weights, and the largest P_{ot} weight, respectively. The effective sample size and the largest P_{ot} weight are computed for the full panel (across commuting zones and years) and cross-section (across commuting zones). The number of observations is provided for CZ-year pairs and CZs.

As in Borusyak et al. (2022), we further examine the importance weight, $P_{ot} = \sum_{d \neq o} P_{odt}$. The second panel of Table I1 reports the effective sample size, given by the inverse of the Herfindahl-Hirschman index (HHI) of P_{ot} weights, which is sufficiently large across commuting zones and years. The largest P_{ot} weight is small, which is consistent with the assumption that $\mathrm{E}\big(\sum_{\text{cluster}}\big(\sum_{o \in \text{cluster}} \sum_{d \neq o} P_{odt}\big)^2\big) \to 0$. These results support the validity of the large-sample approximation and suggest favorable finite sample performance of the Bartik estimator.

To assess the correlation patterns of shocks and determine appropriate clustering for robust standard errors, we estimate intra-class correlation coefficients (ICCs) within U.S. census regions, states, and commuting zones using the random effects model:

$$\begin{aligned} I_{ot} &= \iota_t + a_{\text{region}(o)t} + b_{\sigma(o)t} + c_o + \varepsilon_{ot} \\ \widetilde{I}_{ot} &= \widetilde{\iota}_t + \widetilde{a}_{\text{region}(o)t} + \widetilde{b}_{\sigma(o)t} + \widetilde{c}_o + \widetilde{\varepsilon}_{ot}, \end{aligned}$$

where ι_t is year fixed effects, $a_{\text{region}(o)t}$ and $b_{\sigma(o)t}$ are time-varying regional and state random effects (to which commuting zone o belongs), and c_o is time-invariant random effects that are specific to commuting zone o.

Table I2: Intra-class correlation coefficients (ICCs).

	region	state	CZ	observations
I_{ot}	0.046	0.065	0.403	2,813
	(0.191)	(0.109)	(0.073)	
	[0.027]	[0.035]	[0.105]	
$\widetilde{I}_{ot}$	0.063	0.090	0.000	2,715
	(0.341)	(0.178)	$(\cdot)^{\dagger}$	
	[0.052]	[0.074]	[0.014]	

Notes: The reported intra-class correlation coefficients (ICCs) are estimated from the hierarchal model imposing an identify covariance structure for region random effects and that for state random effects. Robust standard errors are reported in parentheses, whereas bootstrapped standard errors are reported in square brackets. †The estimated ICC for CZ lies at the lower bound of the possible range for ICC values, which is zero. Thus, we do not report the robust standard error for this estimate.

Table I2 presents the estimated ICCs, which shows limited evidence for clustering at the regional and state levels. It implies that the number of top inventors, which is the shift of our Bartik instrument, is uncorrelated across broader geographic units such as census regions and states. This aligns with our identifying assumptions based on Borusyak et al. (2022).

To examine the validity of the assumption that $\mathrm{E}(I_{ot}|\bar{\varepsilon}, q, P) = \mu q_{ot}$, we further conduct a regression-based falsification test proposed by Borusyak et al. (2022). The test aim to corroborate the plausibility of the quasi-random assignment of the number of top inventors across years within each origin commuting zone. For this falsification exercise, we use the lagged employment in four sectors as a potential confounder at the origin level, as we do in examining the share exogeneity at the destination level in Appendix F.

Table I3 presents the results of an origin-level falsification test, which provides additional support for the validity of our identification strategy. As in Borusyak et al. (2022), we regress the number of top inventors on the log of lagged sectoral employment, controlling for origin-commuting-zone fixed effects and year fixed effects and weighting by migration shares P_{ot}. The coefficients are not statistically significant at the conventional 5% level.[62] This falsification test complements our previous analysis and reinforces the exogeneity of our Bartik instruments.

[62]Similar to Table F3, the absence of significance at the 5% level does not necessarily imply that the true coefficient is zero. Since our analysis involves multiple regressions, the rejection of the null hypothesis, if any, might be an artifact of multiple hypothesis testing rather than a genuine effect. To address this concern, we report p-values adjusted to control the family-wise error rate (FWER) using the Romano-Wolf multiple hypothesis correction method. We also compute the sharpened false discovery rate (FDR) q-values using the code provided by Anderson (2008).

Table I3: Regression coefficients.

	Manufacturing	Finance and insurance	Professional, scientific, and technical services	Management of companies and enterprises
Regression coefficients	−0.01017	−0.00692	0.00766	0.01184
	(0.00555)	(0.00586)	(0.00576)	(0.01135)
	[0.06661]	[0.23826]	[0.18380]	[0.29673]
	{0.08890}	{0.23480}	{0.22880}	{0.23480}
	<0.36400>	<0.36400>	<0.36400>	<0.36400>

Notes: This table presents the coefficients obtained from regressing the number of top inventors on the lagged manufacturing employment, controlling for origin commuting zone fixed effects and year fixed effects and weighting by migration shares P_{ot}. The parentheses contain the conventional standard errors, while the square brackets show the p-values. The braces present the family-wise error rate (FWER) adjusted p-values, computed by using the Romano-Wolf multiple hypothesis correction method, and the angle brackets present the sharpened false discovery rate (FDR) q-values, calculated by using the code provided by Anderson (2008).

References for Appendices

[1] Anderson, Michael L. (2008). Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects. *Journal of the American Statistical Association* 103(484): 1481–1495.

[2] Chernozhukov, Victor, Christian Hansen (2008). The Reduced Form: A Simple Approach to Inference with Weak Instruments. *Economics Letters* 100(1): 68–71.

[3] Eckert, Fabian, Teresa C. Fort, Peter K. Schott, Natalie J. Yang (2021). Imputing Missing Values in the US Census Bureau's County Business Patterns. *NBER Working Paper* #26632.

[4] Kelejian, Harry H., Ingmar R. Prucha (1998). A Generalized Spatial Two-Stage Least Squares Procedure for Estimating a Spatial Autoregressive Model with Autoregressive Disturbances. *Journal of Real Estate Finance and Economics* 17(1): 99–121.

[5] Klarner, Carl (2015). State Economic and Government Finance Data. *Harvard Dataverse*, V1.

[6] Monath, Nicholas, Christina Jones, Sarvo Madhavan (2021). Disambiguating Inventors, Assignees, and Locations. American Institutes for Research.

[7] Pierson, Kawika, Michael L. Hand, Fred Thompson (2015). The Government Finance Database: A Common Resource for Quantitative Research in Public Financial Analysis. PLoS ONE doi: 10.1371/journal.pone.0130119.

[8] Sanderson, Eleanor, Frank Windmeijer (2016). A Weak Instrument F-test in Linear IV Models with Multiple Endogenous Variables. *Journal of Econometrics* 190: 212–221.

[9] Staiger, Douglas, James H. Stock (1997). Instrumental Variables Regression with Weak Instruments. *Econometrica* 65(3): 557–586.

[10] Toole, Andrew A., Christina Jones, Sarvothaman Madhavan (2021). PatentsView: An Open Data Platform to Advance Science and Technology Policy. *USPTO Economic Working Paper* No. 2021-1.

[11] Trajtenberg, Manuel, Gil Shiff, Ran Melamed (2006). The "Names Game": Harnessing Inventors' Patent Data for Economic Research *NBER Working Paper* #12479.