Institute for Economic Studies, Keio University

Keio-IES Discussion Paper Series

Marshall meets Bartik: Revisiting the mysteries of the trade

Yasusada Murata, Ryo Nakajima

28 September, 2023
DP2023-015
https://ies.keio.ac.jp/en/publications/23108/

Institute for Economic Studies, Keio University
2-15-45 Mita, Minato-ku, Tokyo 108-8345, Japan
ies-office@adst.keio.ac.jp
28 September, 2023

Marshall meets Bartik: Revisiting the mysteries of the trade

Yasusada Murata, Ryo Nakajima

Keio-IES DP2023-015
28 September, 2023
JEL Classification: R12, O31, J61, C26
Keywords: knowledge spillovers, knowledge sharing, Bartik instruments, mysteries of the trade, idea-generating process

Abstract

We identify a causal effect of top inventor inflows on patent productivity of local inventors by combining the idea-generating process described by Marshall (1890) with the Bartik (1991) instruments involving state taxes and commuting zone characteristics of the United States. We find that the 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[†]

September 28, 2023

Abstract

We identify a causal effect of top inventor inflows on patent productivity of local inventors by combining the idea-generating process described by Marshall (1890) with the Bartik (1991) instruments involving state taxes and commuting zone characteristics of the United States. We find that the 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: knowledge spillovers; knowledge sharing; Bartik instruments; mysteries of the trade; idea-generating process

JEL codes: R12; O31; J61; C26

[*]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 spillovers have been central to various fields of economics such as trade, growth, and geography. However, little is known about the idea-generating process through knowledge spillovers despite Marshall's (1890) simple explanation:

> *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, testing for this channel has been challenging. The main difficulty lies in possible endogeneity: those who generate new ideas tend to cluster together.

To uncover this idea-generating process, we identify a causal effect of top inventor inflows on the patent productivity of local inventors at the commuting-zone level in the United States. In doing so, we use 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 inventors in the origin correspond to a shift.

We consider two types of outcomes: the productivity gains of all local inventors and those of external inventors who are not directly connected to incoming inventors through organizations or co-inventor relationships. 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 productivity gains from both internal knowledge sharing and external knowledge spillovers. The latter gains go beyond organizational boundaries and co-inventor relationships and thus 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."

Our identification strategy consists of main three steps. We first analyze the impact of spatial and temporal variation in individual 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 estimated bilateral location choice probabilities across origin commuting zones to construct a Bartik instrument, which depends on the *tax differences* between that destination and all other commuting zones. We finally employ an instrumental variable (IV) approach, where we use the Bartik instrument in the first-stage regression to proxy for top inventor inflows and estimate a structural equation with the outcome being local patent productivity. The identifying assumption is that the patent productivity in the destination does not directly depend on the *tax differences* between that destination and all other commuting zones.[1]

Our novelty lies in the construction of the Bartik instrument. Since we derive the predicted migration probability from a location choice problem of a top inventor who faces spatial and temporal differences in individual income tax rates, 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: the direct gains from tax changes; and the indirect gains from tax changes through top inventor migration. We find that the former effect is 26.5% while the latter is 73.5%.

The contribution of our paper to the literature is threefold. First, we shed new light on the idea-generating process described in Marshall (1890) using Bartik (1991) instruments. Our work thus contributes to the agglomeration and innovation literature, where according to Carlino and Kerr (2015) "there is very little insight into how knowledge is transmitted among individuals living in close geographic proximity." Our paper addresses this important yet unexplored issue in an environment where tax-induced migration of top inventors brings about new knowledge to local inventors in their destination.

[1] For example, the assumption holds if local patent productivity in commuting zone 37500 (Santa Clara–Monterey–Santa Cruz, CA) does not directly depend on the sum of the tax difference between commuting zones 37500 and 19600 (Bergen–Essex–Middlesex, NJ), the tax difference between commuting zones 37500 and 24300 (Cook–DuPage–Lake, IL), and so forth.

Second, we disentangle productivity gains due to external knowledge spillovers from those due to internal knowledge sharing, which allows us revisit the mysteries of the trade in the air. The theoretical background of this distinction between pure knowledge spillovers and market-mediated knowledge transfer dates back at least to Griliches (1979) and Romer (1990). To the best of our knowledge, however, this partially nonexcludable good nature of knowledge has not been studied in a spatial framework using modern causal inference methods. Since the existence of the gains from external knowledge spillovers constitutes a rationale for spatial agglomeration of inventive activity, our analysis, which leverages spatial and temporal variation in tax rates, contributes to the innovation policy literature that considers both the benefits and costs of entrepreneurial clusters (e.g., Chatterji, Glaeser, and Kerr, 2014).

Last, we derive the Bartik instruments from a location choice model involving policy variables à la Moretti and Wilson (2017). The theory-based Bartik instruments can be readily used for counterfactual experiments and applied to various contexts where origin-destination flows are affected by the difference in policy variables across locations. Our counterfactual analysis illustrates this type of experiment and bridges 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). Since our analysis allows not only for the direct productivity gains from tax changes but also for the indirect gains through tax-induced migration of top inventors, it complements these two strands of literature.

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 and construct the Bartik instruments. Section 4 presents our baseline results by employing the IV approach. In Section 5, we check the robustness of the main results in terms of time, space, and aggregation. We estimate an event study model, consider the spatial extent of local productivity gains, and disaggregate local inventors according to their productivity. Section 6 discusses the underlying mechanisms through which local productivity gains materialize by focusing on two key channels: patent citations and trade secrets. 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 $3,011,262$ patent applications by $1,280,374$ unique inventors (see Appendix A for a more detailed description of the main and secondary data sources).[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 determine (i) who qualify as top inventors; (ii) under what condition we detect the migration of a top inventor; and (iii) who in the destination potentially gain 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 weighted by the number of co-inventors. We then identify, for each year, the top five inventors by productivity over the last ten years and call them the *top inventors* for short.[3] We detect the migration of a top inventor if the commuting zone of residence of an inventor recorded in the patent application data differs in two consecutive years *and* if that inventor was a top inventor in the first year.[4] 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 $6,182$. Thus, on average, each top inventor moved 1.485 times, conditional on moving at least once.[5]

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

[3]This definition of top inventors is similar to Moretti and Wilson (2017).

[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.

[5]While the disambiguation algorithm adopted in PatentsView is known to be highly accurate, it is not error-free (see Toole et al., 2021). Thus, 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 more than ninety-nine percent of the inventors who applied for patents in two consecutive years *and* who qualified as

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]

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 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 (weighted by the number of co-inventors) 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.

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 a high correlation does not always imply causation, we take an IV approach to examine the causal effect of the top inventor inflows on the productivity of local inventors. As explained below, we use the variation in the individual income average tax rates (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.

---

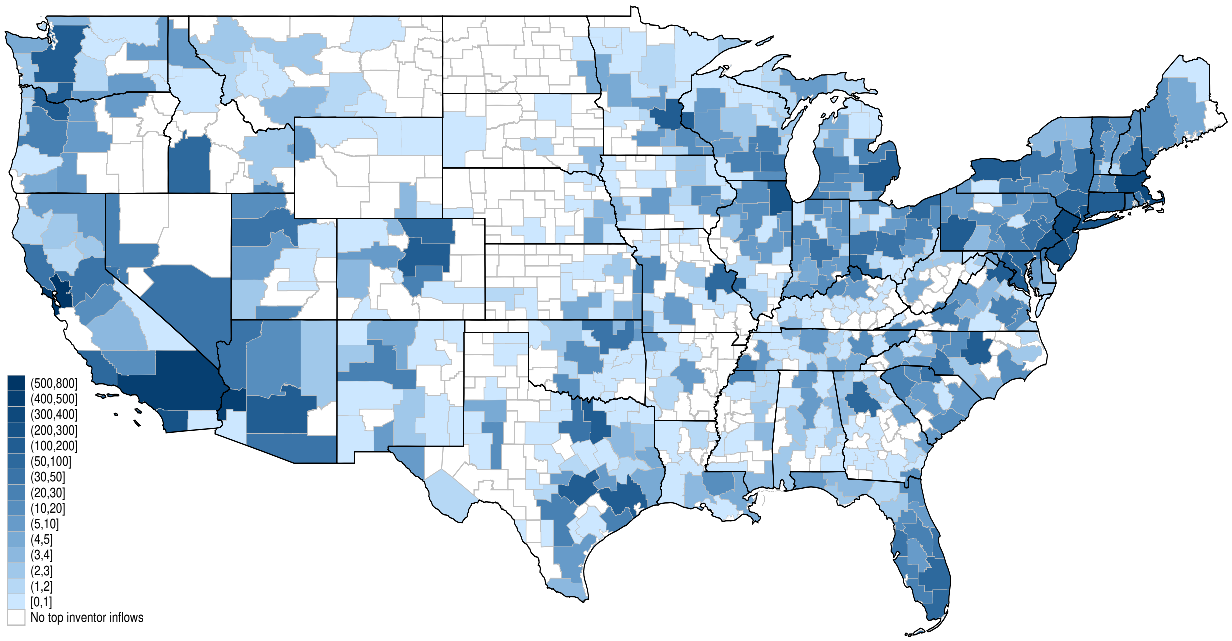
top inventors at least once.

[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 Murata, Nakajima, Okamoto, and Tamura, 2014). We check the robustness of the result regarding geographic space in Section 5.

[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).

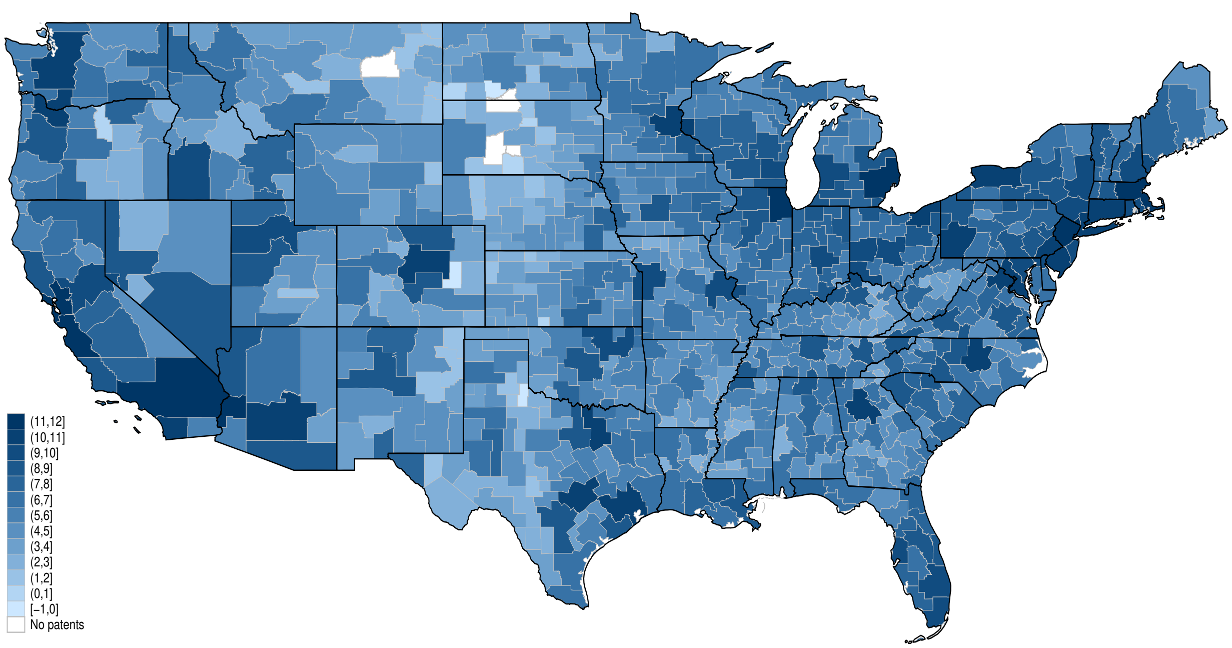
[8]All data on state taxes and tax credits for the years 1977 to 2009 are provided by Moretti and Wilson (2017).

Figure 1: Geographic distribution of top inventor inflows.



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

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



Notes: The productivity in each commuting zone is defined as the number of patents by local inventors from 1977 to 2009, weighted by the number of co-inventors. It excludes the number of patents by top inventors who moved in.

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

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 number of patents by local inventors from 1977 to 2009, weighted by the number of co-inventors. It excludes the number of patents by top inventors who moved in.

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	10205.625
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. 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 it 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

We first analyze how tax differences across states affect the migration of top inventors from origin commuting zone o to destination commuting zone d in Sections 3.1-3.4, which reproduce the results in Moretti and Wilson (2017). We then present in Section 3.5 a new method to construct a Bartik instrument by using the predicted flows of top inventors.

Let $\sigma(o)$ and $\sigma(d)$ denote the states to which origin and destination commuting zones belong, respectively. In the beginning of period t, the inventors in o, whose number is denoted by I_{ot}, observe individual income tax rates in origin and destination states, $\tau_{\sigma(o)t}$ 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 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 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, inventors choose the location that gives them the highest utility. The utility of an 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 w_{dt} is the wage in d; 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 location. The utility of an 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 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 inventors who move from o to d relative to the share of inventors who stay in o. The log odds ratio is then given by

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

3.2 Firms

In each period, firms choose the location that yields the maximum profit. The profit of firm j, which is 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 CITRs, ITCs, and RTCs in $\sigma(d)$; Z'_d represents production amenities in d; C'_{od} is the cost of migration for a firm; and ε'_{jodt} represents the time-varying idiosyncratic productivity match between a firm and a commuting zone. As in the case with inventors, assume 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

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

3.3 Equilibrium

In equilibrium, labor demand equals labor supply in each commuting zone in each year. To derive the equilibrium, we first solve (2) for $\ln w_{dt} - \ln w_{ot}$. We then plug the resulting expression into (1) and use the equilibrium condition $\ln(P'_{odt}/P'_{oot}) = \ln(P_{odt}/P_{oot})$ to obtain the equation that we estimate as follows (see Appendix B for the derivation):

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

where $\eta = \frac{\alpha}{1+\alpha}$ and $\eta' = \frac{\alpha\beta}{1+\alpha}$ are parameters governing inventor (and firm) mobility; $\gamma_d = \frac{1}{1+\alpha}[Z_d + \alpha Z'_d]$ and $\gamma_o = -\frac{1}{1+\alpha}[Z_o + \alpha Z'_o]$ are origin and destination fixed effects that account for consumption and production amenities; $\gamma_{od} = -\frac{1}{1+\alpha}[C_{od} + \alpha C'_{od}]$ stands for 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.

3.4 Estimation

When estimating (3), we proxy for $\tau_{\sigma(d)t}$ with 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.[9] We regard $\tau'_{\sigma(d)t}$ as consisting of the CITR, ITC, and RTC. We use 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 shows that the interstate migration result 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 the baseline case in Moretti and Wilson (2017).

[9]We report the result at the ninety-fifth percentile in the main body and the result at the ninety-ninth percentile in Appendix C.1 as a robustness check.

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	4866	4866	7226	7225
$\overline{R}^2$ (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 (3). ATR, CITR, ITC, and RTC stand for individual income average tax rate at the ninety-fifth percentile, corporate income tax rate, investment tax credit, and R&D tax credit, respectively. Cluster-robust standard errors are in parentheses.

3.5 Bartik instrument

Our aim here is not to establish a causal relationship between inventor mobility and income taxes. Instead, we construct a Bartik instrument from the estimated parameters in Column (2) of Table 4 as follows:

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

which is used in the first-stage regression that we explain in the next section. Observe that B_{dt} is the prediction of the top inventor inflows defined as $M_{dt} = \sum_{o \neq d} M_{odt} = \sum_{o \neq d} P_{odt} I_{ot}$. We derive the predicted probability $\widehat{P}_{odt}$ that top inventors moved from o to d in year t by the following three steps. First, equation (3) implies that for any pair of commuting zones c

and d, P_{oct} and P_{odt} must satisfy

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

Second, let $\mathcal{C}$ denote the set of commuting zones. Since $\sum_{c\in\mathcal{C}} P_{oct} = 1$ holds, we have

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

Last, using the estimated parameters $\widehat{\eta}$, $\widehat{\eta}'$, $\{\widehat{\gamma}_c\}_{c\in\mathcal{C}}$, and $\{\widehat{\gamma}_{od}\}_{o,d\in\mathcal{C}}$, we obtain

$$\widehat{P}_{odt} = \frac{\exp\left\{\widehat{\eta} \ln(1-\tau_{\sigma(d)t}) + \widehat{\eta}' \ln(1-\tau'_{\sigma(d)t}) + \widehat{\gamma}_d + \widehat{\gamma}_{od}\right\}}{\sum_{c\in\mathcal{C}} \exp\left\{\widehat{\eta} \ln(1-\tau_{\sigma(c)t}) + \widehat{\eta}' \ln(1-\tau'_{\sigma(c)t}) + \widehat{\gamma}_c + \widehat{\gamma}_{oc}\right\}}, \tag{4}$$

where $\widehat{\eta}$, $\widehat{\eta}'$, $\{\widehat{\gamma}_c\}_{c\in\mathcal{C}}$, and $\{\widehat{\gamma}_{od}\}_{o,d\in\mathcal{C}}$ are the estimated parameters.

4 The migration of top inventors and local patent productivity

We analyze the impact of top inventor inflows on local patent productivity. In this section, we present a baseline case using a static framework. We then extend it to a dynamic setting in the next section to check the robustness of our results. In both cases, we consider two types of outcomes: (a) productivity gains of all local inventors and (b) those of external inventors. The former will be interpreted as aggregate productivity gains from both internal knowledge sharing and external knowledge spillovers. The latter will be viewed as the gains from external knowledge spillovers because the gains go beyond organizational boundaries and co-inventor relationships. Our main focus is on the latter since they pertain to what Marshall (1920) referred to as the mysteries of trade in the air.

4.1 Empirical specifications

We start with the fixed effect (FE) model:

$$\ln Y_{dt} = \phi M_{dt} + \xi X_{dt} + \varepsilon_{dt}, \tag{5}$$

where Y_{dt} is the productivity of inventors in destination commuting zone d in period t (measured by the number of applied for patents weighted by the number of 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 measures the productivity effect, X_{dt} captures time-varying factors in destination commuting zone d and taxes in state $\sigma(d)$, as well as commuting zone and time fixed effects,[10] and ε_{dt} is an i.i.d. shock. We assume that $E(\varepsilon_{dt}|X_{dt}) = 0$, i.e., ε_{dt} is mean zero conditional on X_{dt}.[11] When estimating (5), we cluster standard errors at the commuting zone level.[12]

However, the top inventor inflows M_{dt} may be correlated with ε_{dt} due to reverse causality or the existence of omitted variables that have direct impacts on both the 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 employ an IV regression, which consists of the structural equation

$$\ln Y_{dt} = \phi^s M_{dt} + \xi^s X_{dt} + \varepsilon^s_{dt}, \tag{6}$$

[10] In the baseline case, we use the ATR at the ninety-fifth percentile of the U.S. income distribution. As robustness checks, we consider the ATR at the ninety-ninth percentile in Appendix C.1 and manufacturing employment in commuting zone d and other taxes and tax credits in state $\sigma(d)$ in Appendix C.2.

[11] One may worry that the conditional mean assumption $E(\varepsilon_{dt}|X_{dt}) = 0$ fails due to omitted variables. To address this concern, we also consider a specification that allows X_{dt} to include state-year fixed effects.

[12] When estimating (5), we replace $\ln Y_{dt}$ with $\ln(1 + Y_{dt})$ in the baseline case to accommodate commuting zone $\times$ year observations with no patents. As a robustness check, we drop such observations and estimate (5) while retaining $\ln Y_{dt}$. As reported in Appendix C.3, the results are quite similar to those in the baseline case.

and the first-stage regression

$$M_{dt} = \psi^f B_{dt} + \xi^f X_{dt} + \varepsilon^f_{dt} \quad \text{with} \quad B_{dt} = \sum_{o \neq d} \widehat{P}_{odt} I_{ot}, \tag{7}$$

where I_{ot} is the number of top inventors in commuting zone o in the beginning of period t. Since $\widehat{P}_{odt}$ is the fitted value from the regression in (3), it captures the estimated probability that an inventor, located in o at the beginning of period t, moves to d by the end of period t. When estimating (6), we cluster standard errors at the commuting zone level.[13]

Recall that B_{dt} is a Bartik instrument, which consists of the shares $\widehat{P}_{odt}$ and the shifts I_{ot}. It is derived from a simple theory of migration, namely that the flow of inventors to destination d is the sum of the products of these two elements. We follow the shares perspective (Goldsmith-Pinkham et al., 2020), i.e., it is the shares $\widehat{P}_{odt}$ that provide the exogenous variation satisfying $E(\varepsilon_{dt}\widehat{P}_{odt}|X_{dt}) = 0$. Indeed, as seen from (4), the shares $\widehat{P}_{odt}$ depend not only on the tax rates in the destination state, $\{\tau_{\sigma(d)t}, \tau'_{\sigma(d)t}\}$, but also on the distribution of tax rates across states, $\{\tau_{\sigma(c)t}, \tau'_{\sigma(c)t}\}_{c\in\mathcal{C}}$. Thus, the state taxes, $\{\tau_{\sigma(c)t}, \tau'_{\sigma(c)t}\}_{\sigma(c)\neq\sigma(d)}$, other than those in the destination, $\{\tau_{\sigma(d)t}, \tau'_{\sigma(d)t}\}$, have an indirect effect on the destination productivity Y_{dt} only via $\widehat{P}_{odt}$. This indirect effect, rather than the "pull" effect captured by the destination taxes $\{\tau_{\sigma(d)t}, \tau'_{\sigma(d)t}\}$, works as an exclusion restriction for identifying the productivity effect ϕ^s.[14] We describe the conditions to ensure the share exogeneity $E(\varepsilon_{dt}\widehat{P}_{odt}|X_{dt}) = 0$ in Appendix E.

We further consider two variants of the Bartik instruments to assess the sensitivity of our results. One is the prediction of the between-state top inventor inflows, $B^{\sigma}_{dt} = \sum_{o\notin\sigma(d)} \widehat{P}_{odt} I_{ot}$, and the other is the predicted top inventor inflows into commuting zone $\nu(d)$ in the nearest neighborhood of d, $B^{\nu}_{dt} = \sum_{o\neq d,\nu(d)} \widehat{P}_{o\nu(d)t} I_{ot}$.

[13] When estimating (6), we replace $\ln Y_{dt}$ with $\ln(1 + Y_{dt})$ in the baseline case to accommodate commuting zone × year observations with no patents. As a robustness check, we drop such observations and estimate (6) while retaining $\ln Y_{dt}$. As reported in Appendix C.3, the results are quite similar to those in the baseline case.

[14] Although our identification relies on the exclusion restriction that the tax in one state does not directly affect the local productivity in other states, state tax competition may induce their interstate correlation. To alleviate this concern, we examine the possibility of strategic interactions among state governments by estimating a reaction function such that the income tax in one state responds to the income taxes in its neighboring states (see, e.g., Brueckner, 2003). As shown in Appendix D, we find no evidence of state tax competition, which is consistent with the exclusion restriction.

4.2 Main results

Table 5 reports the estimation results for the FE and IV regressions. Column 1 reports the FE case. Columns 2-6 are the results for different IV regressions. Column 2 considers $B_{dt} = \sum_{o \neq d} \widehat{P}_{odt} I_{ot}$ in (7). Columns 3 adds to Column 2 an instrument, $B^{\sigma}_{dt} = \sum_{o \notin \sigma(d)} \widehat{P}_{odt} I_{ot}$, which captures top inventor flows only from other states. Column 4 further adds to Column 3 an additional instrument, $B^{\nu}_{dt} = \sum_{o \neq d, \nu(d)} \widehat{P}_{o\nu(d)t} I_{ot}$, which involves top inventor flows into commuting zone $\nu(d)$ in the nearest neighborhood of d. Columns 5 and 6 report the results when controlling for time-varying state-specific unobservables by year $\times$ destination state FE. 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 $\ln(1 - \text{ATR})$, are virtually identical for all IV regressions within each panel.[15] Panel (a) in Table 5 shows that an inflow of a top inventor raises the patent productivity of all local inventors by approximately 6%. Panel (b) shows that an inflow of a top inventor raises local patent productivity by approximately 4% when 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. Our results obtained from the IV regressions thus differ from Zacchia (2018), who finds no city-wide spillover effect of inventor inflows. Our 4 percent patent productivity gains for local inventors due to an additional inventor inflow could be compared with the 12 percent gains for incumbent plants' TFP due to a new plant opening studied in Greenstone, Hornbeck, and Moretti (2010). However, given the difference between inventor arrival and firm entry, it is not surprising that the former effect is smaller.[16]

[15]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).

[16]Our paper also differs from De la Roca and Puga (2017) and Moretti (2021) in that these papers focus

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

	(1)	(2)	(3)	(4)	(5)	(6)
(a) All local inventors						
Top inventor inflows	0.043	0.062	0.060	0.059	0.059	0.060
	(0.006)	(0.013)	(0.012)	(0.011)	(0.013)	(0.012)
ln(1 − 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	35.130	34.997
$\tau = 5\%$		37.418	31.930	34.734	31.214	32.989
$\tau = 10\%$		23.109	19.892	21.389	19.473	20.364
$\tau = 20\%$		15.062	13.094	13.901	12.839	13.272
$\tau = 30\%$		12.039	10.531	11.093	10.336	10.610
(b) External inventors						
Top inventor inflows	0.027	0.042	0.040	0.041	0.036	0.038
	(0.004)	(0.010)	(0.009)	(0.009)	(0.010)	(0.010)
ln(1 − 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	35.130	34.997
$\tau = 5\%$		37.418	31.921	34.738	31.203	32.988
$\tau = 10\%$		23.109	19.887	21.392	19.467	20.364
$\tau = 20\%$		15.062	13.091	13.902	12.835	13.272
$\tau = 30\%$		12.039	10.529	11.094	10.333	10.610
CZ FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	No	No
Year × state FE	No	No	No	No	Yes	Yes
Observations	23,628	23,628	23,628	23,463	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 − ATR) is converted to elasticity. Cluster-robust standard errors are in parentheses. Column 1 does not control for 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 and 6 replace ln(1 − ATR) in Columns 3 and 4 with Year × state FE.

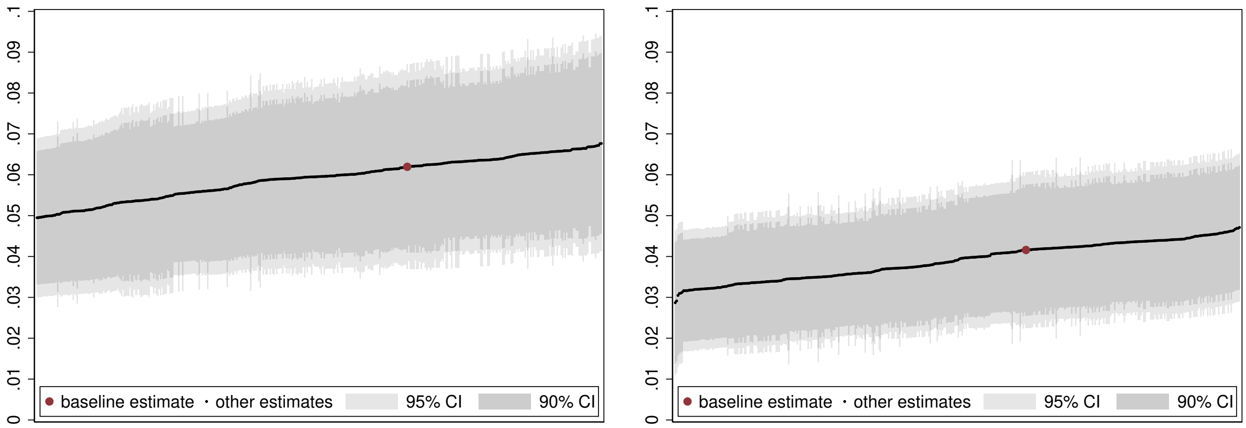
We check the robustness of these results using the specification curve analysis as in Simonsohn, Simmons, and Nelson (2020). We consider different specifications of the IV regressions by focusing on five dimensions.[17] Figure 3 plots the specification curve for ϕ^s with 90% and

on the impacts on those who migrate.

[17] 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 $\{B_{dt}, B^{\sigma}_{dt}, B^{\nu}_{dt}\}$; (iv) whether to include state-year FE; and (v) whether to include each of the other controls. 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 3: Specification curve analysis.

(a) all local inventors (b) external inventors



Notes: Panel (a) (Panel (b)) illustrates the impacts of inventor inflows on local patent productivity without (by) excluding co-inventors and inventors in the same assignee. The specification curve is depicted using 409 alternative specifications explained in footnote 17. The vertical axis is the value of ϕ^s.

95% confidence intervals. As seen from Panels (a) and (b), the productivity gains of approximately 6% and 4% are fairly robust for 409 alternative specifications, thus verifying that the baseline estimates do not come from data mining.[18] In what follows, we use the specification in Column 3 of Table 5 as a baseline unless otherwise stated.

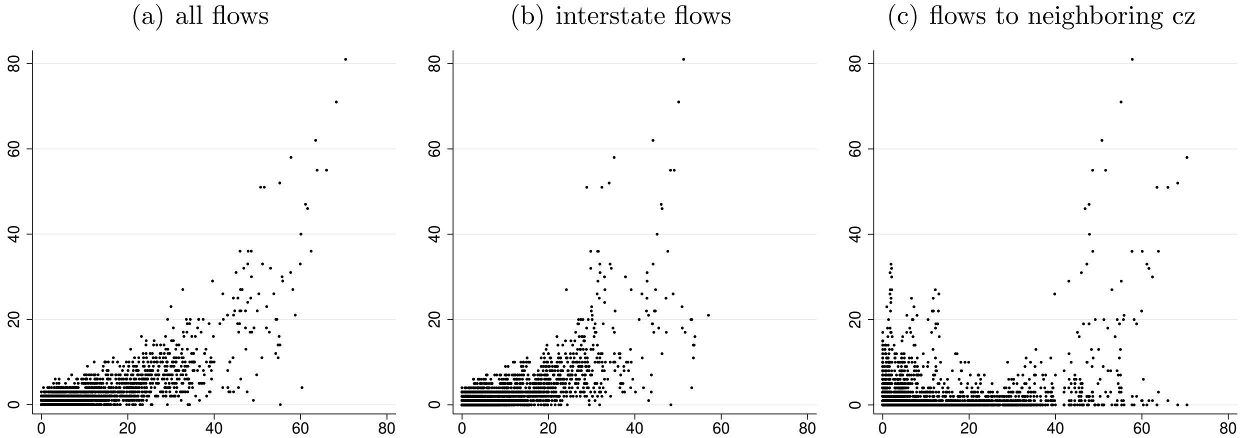
4.3 Relevance and validity of empirical strategy

Our empirical strategy relies on relevance and validity of the Bartik instruments. We now discuss these two assumptions.

To assess the relevance of the Bartik instruments, we first plot in Figure 4 the relationship between the actual migration flows $M_{dt} = \sum_{o \neq d} M_{odt}$ and the Bartik instruments constructed from the fitted migration 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.

[18]We summarize the results for some representative cases of these alternative specifications in Appendix C.1-C.3. In particular, we check the robustness of our results by using an alternative ATRs in Appendix C.1, by including other controls in Appendix C.2, and by dropping commuting zone × year observations with no patents in Appendix C.3. As seen from Tables C2, C3, and C4, the results are fairly robust.

Figure 4: Actual versus fitted flows of top inventors.



Notes: In each panel, the vertical and horizontal axes are the actual migration flows and the Bartik instruments constructed from the fitted migration 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.

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, Stock, and Sun, 2019). The bottom panel of 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 $\widehat{P}_{odt}$ conditional on the characteristics of the commuting zones X_{dt}. We assess the plausibility of the exogeneity assumption 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 $\widehat{\phi}^s = \sum_{o \in C} \widehat{\omega}^s_o \widehat{\phi}^s_o$, which consists of the origin-specific weight $\widehat{\omega}^s_o$ 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. We then assess the exogeneity of the shares $\{\widehat{P}_{odt}\}_{d\in\mathcal{C}}$ for commuting zones o with the top five highest Rotemberg weights $\widehat{\omega}_o^s$ by considering the correlation between the predicted shares $\{\widehat{P}_{odt}\}_{d\in\mathcal{C}}$ and the characteristics of the commuting zones $\{X_{dt}\}_{d\in\mathcal{C}}$.

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.[19] Table F2 reports the destination commuting zones to which top inventors moved from these top five origin commuting zones. The result that most migrations are interstate is in line with the assumption that the main source of identifying variation comes from inventor mobility caused by personal income tax differences between states. Table F3 reports, for each top five origin commuting zone, the correlation between the share and the log of manufacturing employment conditional on the controls and fixed effects.[20] Reassuringly, the correlations are low, and the regression coefficients on the log of manufacturing employment are not statistically significant at the five percent level in all specifications.

5 Robustness

In this section, we examine the robustness of our main results in terms of time, space, and aggregation. 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 of our main results in terms of the geographic extent of produc-

[19]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.

[20]Table C3 shows that manufacturing employment is related to local patent productivity. We thus use manufacturing employment as an observable confounder. In principle, we could examine the correlation of the share with other commuting zone characteristics that affect local patent productivity. However, in practice, such time-varying data on commuting zone characteristics are not readily available for the time period that we consider.

tivity gains. We finally consider a disaggregated case where local inventors are classified by their patent productivity. In Appendix C, we further conduct additional robustness checks by considering: the case where the ATR for a hypothetical taxpayer is assumed to be at the ninety-ninth percentile of the U.S. income distribution; the case where we include other controls in our baseline specification; and the case where we drop commuting zone × year observations with no patents in our main analysis.

5.1 Dynamic analysis

As a first robustness check, we conduct event study analysis to examine whether the baseline results are sensitive to the inclusion of lead and lag effects of top inventor inflows. To this end, we extend (6) as follows:

$$\ln Y_{dt} = \sum_{j=\underline{j}+1}^{\overline{j}} \phi_j^{\ell} M_{dt-j} + \xi^{\ell} X_{dt} + \varepsilon_{dt}^{\ell}, \tag{8}$$

which is a distributed lag model in levels with a binning window $[\underline{j}+1, \overline{j}]$. Thus, when $\underline{j} = -1$ and $\overline{j} = 0$, the model degenerates into the static model (6). Schmidheiny and Siegloch (2023) show that equation (8) is equivalent to the event study model given by[21]

$$\ln Y_{dt} = \sum_{j=\underline{j}}^{\overline{j}} \mu_j^{es} \Delta M_{dt}^{(j)} + \xi^{es} X_{dt} + \varepsilon_{dt}^{es}, \tag{9}$$

where

$$\Delta M_{dt}^{(j)} = \begin{cases} \sum_{k=-\infty}^{\underline{j}} (M_{dt-k} - M_{dt-k-1}) & \text{if} \quad j = \underline{j} < 0 \\ M_{dt-j} - M_{dt-j-1} & \text{if} \quad \underline{j} < j < \overline{j} \\ \sum_{k=\overline{j}}^{\infty} (M_{dt-k} - M_{dt-k-1}) & \text{if} \quad j = \overline{j} > 0 \end{cases}. \tag{10}$$

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., $\mu_j^{es} = \mu_{j-1}^{es} +$

[21]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.

$\phi_j^\ell = \sum_{h=0}^{j} \phi_h^\ell$ for $j = 0, 1, ..., \bar{j}$ 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 $\boldsymbol{\Delta B}_{dt} = [\Delta B_{dt}^{(\underline{j})} \cdots \Delta B_{dt}^{(\bar{j})}]'$ denote a $(\underline{j}+\bar{j}+1) \times 1$ vector of the first time difference of the IVs, where $\Delta B_{dt}^{(j)}$ is defined in a similar way as in (10). The IV event study model consists of the structural equation (9) and the first-stage regression analogous to (7) as follows[22]

$$\Delta M_{dt}^{(j)} = \boldsymbol{\psi}^{ef(j)} \boldsymbol{\Delta B}_{dt} + \xi^{ef(j)} X_{dt} + \varepsilon_{dt}^{ef(j)}, \tag{11}$$

where $\boldsymbol{\psi}^{ef(j)} = [\psi^{ef(j,\underline{j})} \cdots \psi^{ef(j,\bar{j})}]$ is a vector of coefficients.[23]

Figure 5 shows the results for the IV event study regressions. Panel (a) illustrates the dynamic impact of top inventor inflows on the patent productivity of all local inventors, which includes not only internal knowledge sharing within the same assignee and between co-inventors but also 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 baseline results in Section 4.2. In contrast, the pre-event semi-elasticities are close to zero and are not statistically significant in any pre-event year in Panel (b), thus suggesting no productivity gains prior to the event of top inventor migration.[24]

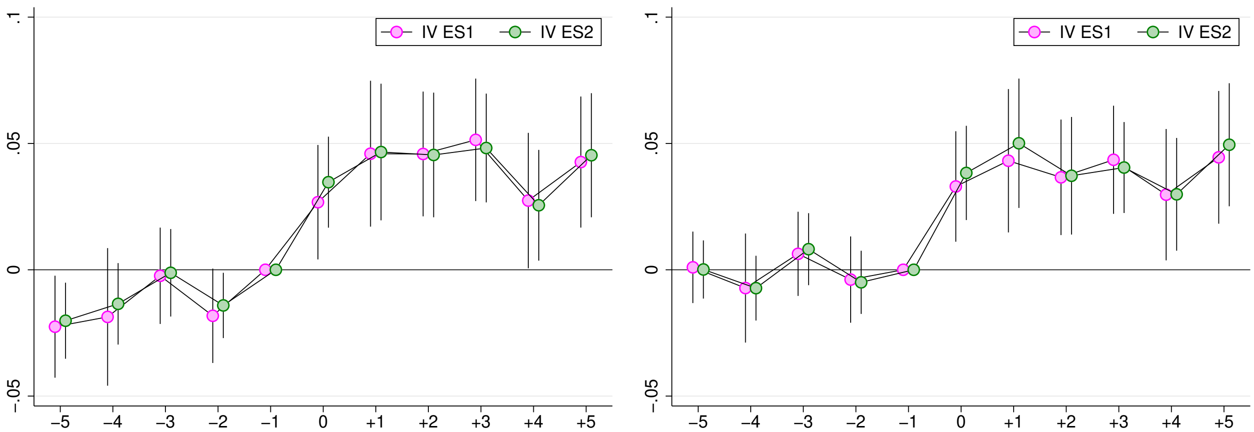
[22]In this robustness check we abstract from the possibility that treatment effects can be heterogeneous. Although several recent papers have explored under what conditions event study models provide valid average treatment effects in the presence of heterogeneous treatment effects (e.g., de Chaisemartin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

[23]When estimating the event study models with multiple instruments, we set $\boldsymbol{\Delta B}_{dt}^{\sigma} = [\Delta B_{dt}^{\sigma(\underline{j})} \cdots \Delta B_{dt}^{\sigma(\bar{j})}]'$ and $\boldsymbol{\Delta B}_{dt}^{\nu} = [\Delta B_{dt}^{\nu(\underline{j})} \cdots \Delta B_{dt}^{\nu(\bar{j})}]'$ and use $[\boldsymbol{\Delta B}_{dt}{}' \ \boldsymbol{\Delta B}_{dt}^{\sigma}{}']'$ or $[\boldsymbol{\Delta B}_{dt}{}' \ \boldsymbol{\Delta B}_{dt}^{\sigma}{}' \ \boldsymbol{\Delta B}_{dt}^{\nu}{}']'$ as instruments in (11).

[24]We report the numbers used in Figure 5 and the associated first-stage statistics in Tables C5 and C6, respectively. In Appendix G, we further assess robustness to possible violations of the parallel trends assump-

Figure 5: IV event study regressions.

(a) all local inventors (b) external inventors



Notes: Panel (a) (Panel (b)) illustrates the dynamic impact of inventor inflows on local patent productivity without (by) excluding co-inventors and inventors in the same assignee. 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.

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 in (6) with

$$\ln Y_{dt} = \sum_{r(d)=1}^{6} \phi^{sr}_{r(d)} M_{r(d)t} + \xi^{sr} X_{dt} + \varepsilon^{sr}_{dt}, \tag{12}$$

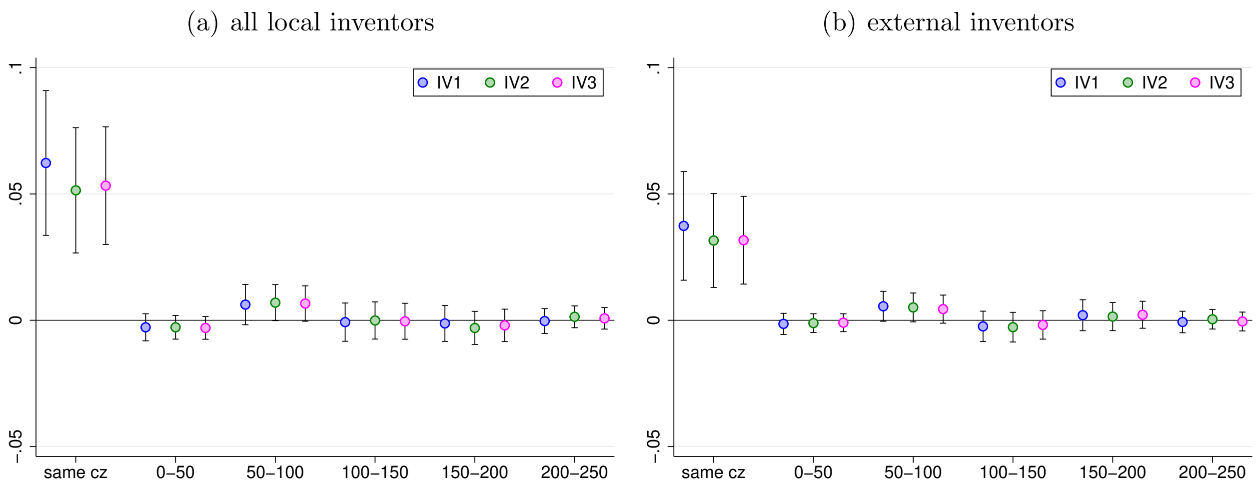
where $r(d)$ is the distance ring defined for each destination commuting zone d and $M_{r(d)t}$ is the predicted inflows of top inventors in the $r(d)$-th ring. The first ring $r(d) = 1$ stands for destination d itself. $r(d) \neq 2, ..., 6$ corresponds to commuting zones that are 0-50, 50-100, 100-150, 150-200, and 200-250 miles away from commuting zone d.[25]

Figure 6 illustrates the estimated coefficients $\{\widehat{\phi}^{sr}_{r(d)}\}^{6}_{r(d)=1}$. Panel (a) (Panel (b)) illustrates the impact of top inventor flows in $r(d)$ on patent productivity in d without (by) excluding co-inventors and inventors in the same assignee. In each panel, the impact is significant only

tion, following the method developed by Rambachan and Roth (2023).

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

Figure 6: Distance-ring regression.



Notes: Panel (a) (Panel (b)) illustrates the impact of inventor inflows on local patent productivity without (by) excluding co-inventors and inventors in the same assignee. 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.

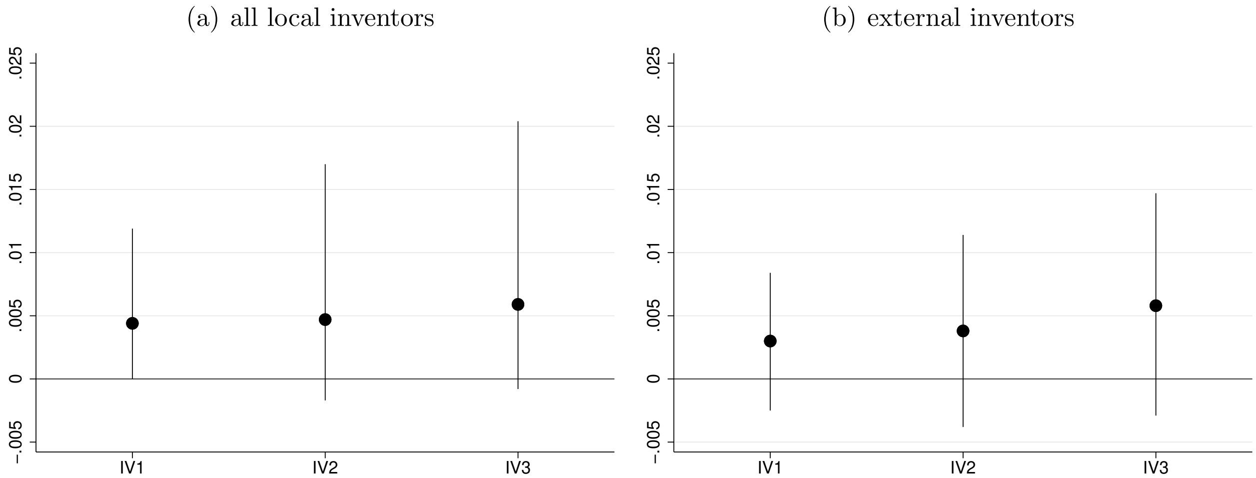
in the first ring for all three different IVs, which implies that top inventor inflows affect patent productivity only in the commuting zone where they enter. Such localized productivity gains are reminiscent of localized knowledge spillovers in Jaffe et al. (1993) and Murata et al. (2014). We will discuss the mechanism of localized productivity gains in terms of localized knowledge spillovers as evidenced by patent citations in Section 6.1.

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 hypothetical commuting zone $R(d) \neq d$ in state $\sigma(d)$. We thus replace the structural equation in (6) with

$$\ln Y_{R(d)t} = \phi^{sR} M_{dt} + \xi^{sR} X_{dt} + \varepsilon^{sR}_{dt} \tag{13}$$

and estimate (13) for each IV specification 1000 times with replacement to obtain the distribution of $\{\phi^{sR}_i\}_{i=1}^{1000}$. We then check if the null hypothesis of no productivity gains, $\phi^{sR} = 0$,

Figure 7: Placebo.



Notes: Panel (a) (Panel (b)) illustrates the impact of inventor inflows on local patent productivity without (by) excluding co-inventors and inventors in the same assignee when $\ln Y_{dt}$ in the structural equation is replaced with $\ln Y_{R(d)t}$, where $R(d) \neq d$ 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.

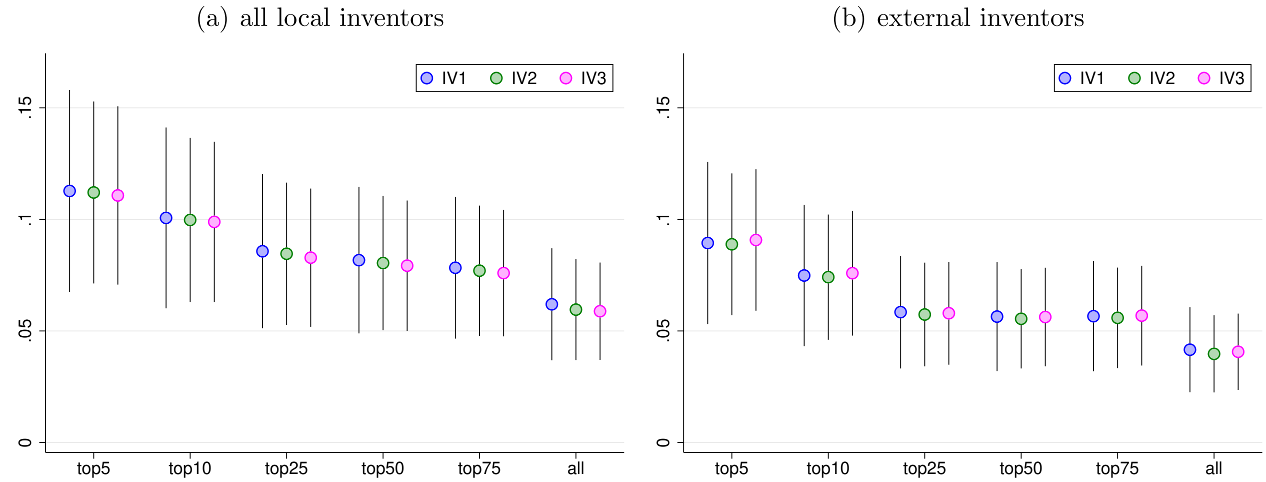
is rejected. Figure 7 depicts the 95% confidence interval and the mean of the distribution for each IV specification. The results in both Panels (a) and (b) show that top inventor flows into commuting zone d do not significantly change patent productivity in randomly drawn commuting zone $R(d) \neq d$ from state $\sigma(d)$, thus implying that the extent of productivity gains is geographically limited within each commuting zone.

5.3 Productivity gains by inventor productivity

We have thus far shown that top inventor inflows enhance patent productivity only for local inventors. Since local inventors differ in their patent productivity, we further delve into the foregoing result by addressing who gain more from top inventor inflows. To this end, we disaggregate local inventors by their patent productivity and estimate the causal effect for each productivity level.

Panel (a) of Figure 8 illustrates the case with all local inventors. We observe that there

Figure 8: Productivity gains by inventor productivity.



Notes: Panel (a) (Panel (b)) illustrates the impact of inventor inflows on local productivity gains by inventor productivity without (by) excluding co-inventors and inventors in the same assignee. 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.

are productivity gains at each productivity level 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 we focus on external inventors who are not directly connected to the migrating top inventors, our result suggests that more-productive local inventors benefit more from their inflows.

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 they were in the air.

6.1 Patent citations

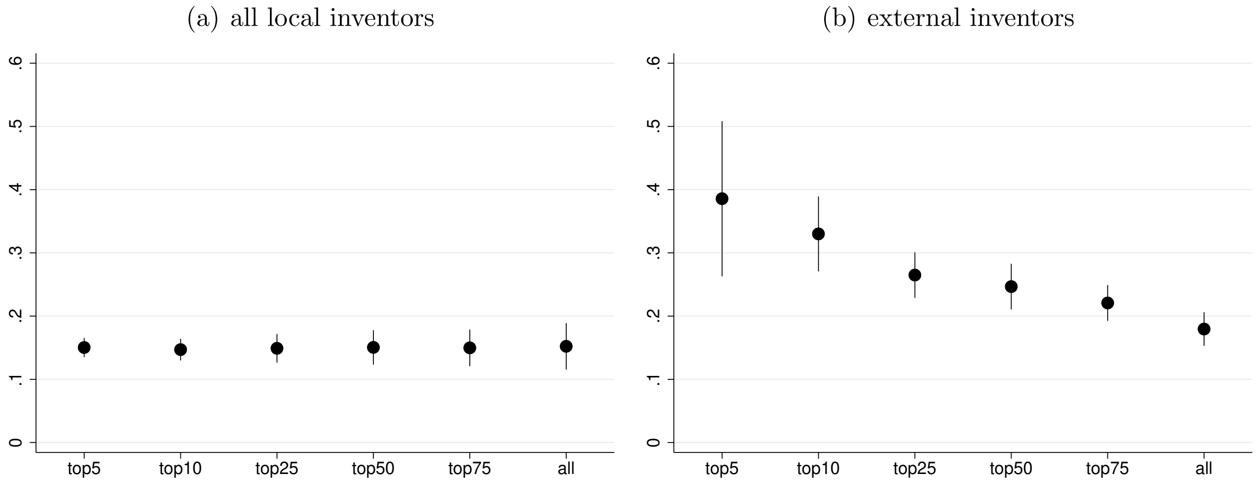
To see the impact of top inventor inflows on local patent citations, we count how many times the patents of the top inventors who migrated into commuting zone d in year t were cited by the local inventors in commuting zone d in year t and denote it by C_{dt}. When constructing C_{dt}, we focus on the patents that had been applied over the last ten years. Replacing patent productivity Y_{dt} in (6) 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 (7). The coefficient ϕ^{sc} gauges the magnitude of knowledge flows from migrating top inventors to local inventors.

Panels (a) and (b) of Figure 9 illustrate the estimated coefficients for the case of all local inventors and that of external inventors, respectively. In both panels, local inventors are classified by their productivity. In Panel (a), an additional top inventor inflow raises the number of local inventors' citations to incoming top inventors 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 cite substantially more patents of the top inventors who moved in the same commuting zone. These results suggest the existence of knowledge flows from the migrating top inventors to the local inventors, even when we focus on the

Figure 9: Citations.



Notes: Panel (a) (Panel (b)) illustrates the impact of an additional top inventor inflow on the number of all local inventors' (external inventors') citations to incoming inventors (in percentage). In each panel, we use B_{dt} and B^{σ}_{dt} as instruments.

external inventors who are not directly connected to the migrating top inventors.[26]

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, inventors who migrate to a

[26]It is perhaps puzzling that the impact is smaller in Panel (a). One possible explanation is 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 less additional percentage changes in the number of internal inventors' citations after the arrival of the top inventors.

state with weaker protection would exchange knowledge more with other inventors outside their organizations, thereby bringing about additional productivity gains to local external inventors. In contrast, the change in legal protection of trade secrets would not affect knowledge sharing within an organization, thus leaving the productivity gains of local internal inventors unaffected.

We examine those differential impacts of top inventor migrations on local patent productivity. To this end, we use the state-level index of trade secrets in Png (2017a, 2017b), which captures both the legal protection under common law and the enactment of the UTSA.[27] The index ranges between 0 and 1, and 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 due to a greater amount of knowledge brought about by inventor migration.

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

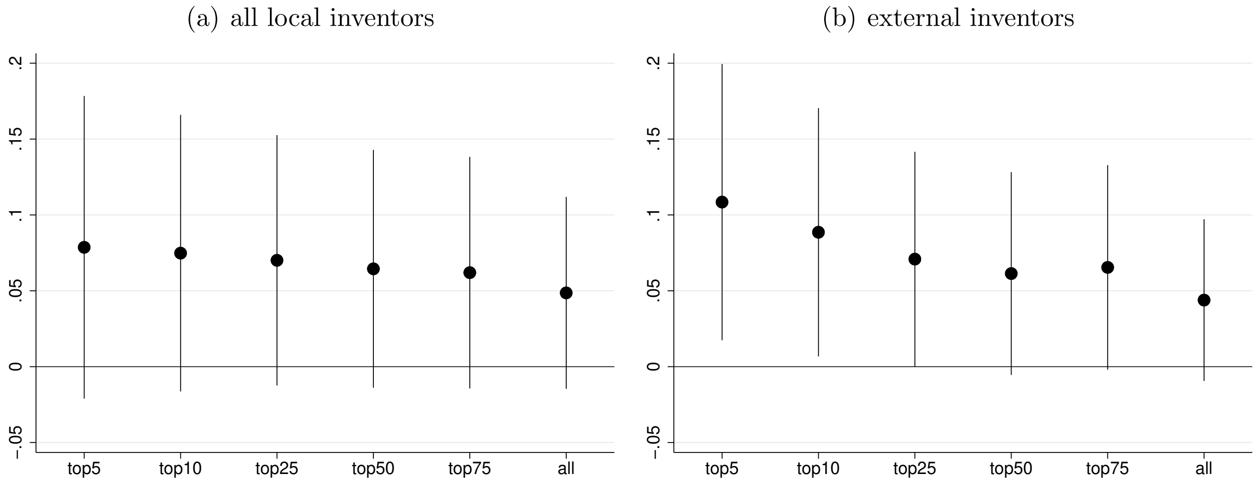
$$\ln Y_{dt} = \boldsymbol{\phi}^s E_{dt} + \xi^s X_{dt} + \varepsilon_{dt}^s. \tag{14}$$

Our interest is in the coefficient ϕ_2^s on $M_{dt}S_{\sigma(d)t}$. If $\phi_2^s > 0$, top inventor inflows lead to higher local patent productivity in commuting zones with lower trade secrets protection, which suggests that knowledge brought about by top inventor migration is more likely to spill over to local inventors in commuting zones with lower trade secrets protection.

As before, we use the Bartik instruments B_{dt} and B_{dt}^{σ} for the top inventor flows 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

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

Figure 10: Trade secrets.



Notes: Panel (a) (Panel (b)) illustrates the coefficients ϕ_2^s on the interaction term $M_{dt}S_{\sigma(d)t}$ in (14) without (by) excluding co-inventors and inventors in the same assignee. In each panel, we use B_{dt} and B_{dt}^{σ} as instruments.

Transfer Act (UFTA) because these laws were introduced to harmonize state laws. Since the three laws are unlikely to be associated with local shocks, we use them as instruments for the trade secrets indicator. The first-stage equation that accompanies (14) is thus given by

$$E_{dt} = \boldsymbol{\psi}^f Z_{dt} + \xi^f X_{dt} + \varepsilon_{dt}^f, \tag{15}$$

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.[28]

Panel (a) in Figure 10 illustrates the result for all local inventors. 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 is insignificant regardless of the productivity of local inventors. Panel (b) in Figure 10 presents the result for external inventors, where the impacts for top 5, top 10, and top 25 local inventors (top

[28]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, $[\mathfrak{S}_1\#\#\mathfrak{S}_2]' = [\{A, B\}\#\#\{C, D\}]' = [A\ B\ C\ D\ AC\ AD\ BC\ BD]'$.

50 and top 75) are significant at the five percent (ten percent) level. Hence, top inventor migration tends to enhance the patent productivity of external inventors in commuting zones with lower trade secrets protection.

7 Counterfactuals

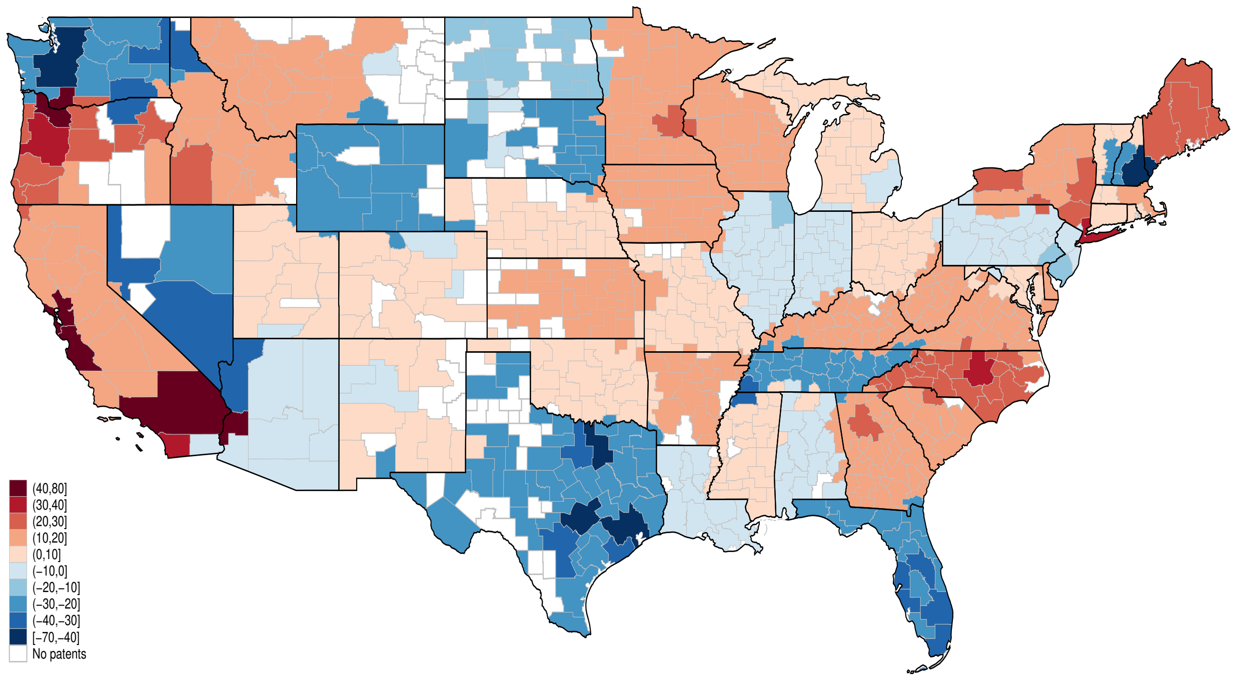
We now conduct a counterfactual experiment. Using the baseline specification in Section 4, we consider what happens to the geographic distribution of patent productivity if all state individual income taxes are set to their average. This experiment is useful for assessing to what extent state tax differences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

Recalling that the changes in state taxes affect the choice probabilities $\widehat{P}_{odt}$ in (4) and the Bartik instruments B_{dt} in (7), as well as B^{σ}_{dt}, the procedure of the counterfactual experiment can be summarized as follows. We first derive the counterfactual probabilities $\widetilde{P}_{odt}$ to construct the counterfactual Bartik instruments $\{\widetilde{B}_{dt}, \widetilde{B}^{\sigma}_{dt}\}$, which allow us to estimate the counterfactual migration flows $\widetilde{M}_{dt}$ via the first-stage regression (7). We then define the counterfactual changes in the migration 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 migrations flows, respectively.[29] We finally apply the counterfactual changes in migration flows $\widetilde{\Delta} M_{dt}$ to the structural equation (6) to construct the counterfactual changes in the log patent productivity $\widetilde{\Delta} \ln Y_{dt} = \widehat{\phi}^s \widetilde{\Delta} M_{dt} + \widehat{\xi}^s \widetilde{\Delta} \ln(1 - \text{ATR}_{\sigma(d)t})$, where $\widetilde{\Delta} \ln(1 - \text{ATR}_{\sigma(d)t}) = \ln(1 - \widetilde{\widehat{\text{ATR}}}_{\sigma(d)t}) - \ln(1 - \text{ATR}_{\sigma(d)t})$ captures the counterfactual tax changes.[30] Thus, the overall impact of tax changes on $\widetilde{\Delta} \ln Y_{dt}$ can be decomposed into two: the direct effect $\widehat{\xi}^s \widetilde{\Delta} \ln(1-\text{ATR}_{\sigma(d)t})$ and the indirect effect via the changes in migration flows $\widehat{\phi}^s \widetilde{\Delta} M_{dt}$. The indirect effect can be further decomposed into two: productivity gains due to internal knowledge sharing and those due to external knowledge spillovers.

[29] 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 migration 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}.

[30] 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.

Figure 11: Counterfactual experiment (Setting state taxes to their average).



Notes: The 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.

Figure 11 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.[31] Table 6 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 7 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)

[31]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 all state taxes equal in (4), include fixed effects at the commuting zone level.

Table 6: 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: The 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^{σ}_{dt} as instruments.

Table 7: Bottom 10 commuting zones by patent productivity gains (%).

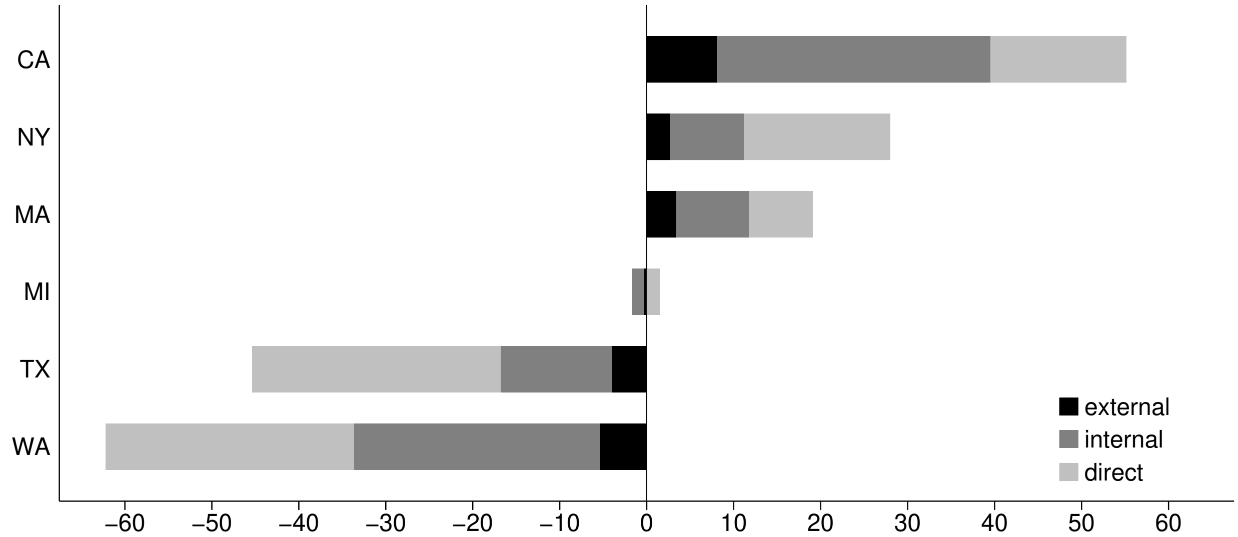
rank	cz number	counties	state	gains (%)
1	39400	King–Pierce–Snohomish	WA	-64.777
2	31201	Travis–Williamson–Hays	TX	-50.710
3	32000	Harris–Fort Bend–Galveston	TX	-49.002
4	33100	Dallas–Denton–Collin	TX	-46.491
5	20600	Hillsborough–Rockingham–York	NH	-41.326
6	7100	Palm Beach–St. Lucie–Martin	FL	-38.194
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: The 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^{σ}_{dt} 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 we define $\widetilde{\Delta} M_{dt}$, where Y_{dt}, $\widehat{Y}_{dt}$, and $\widetilde{Y}_{dt}$ are the actual, fitted, and counterfactual numbers of patents in commuting zone 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

Figure 12: Counterfactual experiment (Setting state taxes to their average)



Notes: The 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 migration. The latter consists of the internal knowledge sharing effect and the external knowledge spillover effect. We use B_{dt} and B^{σ}_{dt} as instruments.

level $\frac{\widetilde{\Delta} Y_{\sigma t}}{Y_{\sigma t}}$.[32]

Figure 12 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 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

[32]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 $\frac{\widetilde{\Delta} Y_{\sigma t}}{Y_{\sigma t}} = \frac{\sum_{d\in\sigma} \widetilde{Y}_{dt} - \sum_{d\in\sigma} Y_{dt}}{\sum_{d\in\sigma} Y_{dt}}$.

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 tax-induced migration can be substantial. To see the relative role of the direct and indirect effects 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 0.735 (and that of the direct effect is 0.265). 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 top inventor inflows on the patent productivity of all local inventors. Even when we focus on local external inventors who are not directly connected to incoming 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.

Since our analysis has disentangled productivity gains due to pure knowledge spillovers from those due to market-mediated knowledge transfer, our findings are consistent with the partially nonexcludable good nature of knowledge, whose implications have been explored theoretically in the technology and growth literature. The existence of the gains from pure knowledge spillovers constitutes a rationale for spatial agglomeration of inventive activity, thus contributing to the innovation policy literature that considers both the benefits and costs of entrepreneurial clusters.

Our counterfactual experiment suggests that the existence 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 from tax changes through top inventor migration matter. Thus, more research is needed to bridge the gap between the tax and innovation literature and the tax and migration literature.

References

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

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

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

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

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

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

[7] de Chaisemartin, Clément and Xavier D'Haultfœuille (2023). Two-Way Fixed Effects and Differences-in-Differences with Heterogeneous Treatment Effects: A Survey. *Econometrics Journal*, forthcoming.

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

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

[10] 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.

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

[12] 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.

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

[14] 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.

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

[16] Kelejian, Harry H. and 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.

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

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

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

[20] Moretti, Enrico and 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.

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

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

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

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

[25] Rambachan, Ashesh, Jonathan Roth (2023). A More Credible Approach to Parallel Trends. *Review of Economic Studies*, forthcoming.

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

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

[28] 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.

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

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

[31] 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.

[32] 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 data construction

Patent data. The main data come from the USPTO PatentsView website (`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 the PatentsView data) 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`).

State taxes and tax credits. Data on U.S. state taxes and tax credits for 1976-2019 are distributed on the Open ICPSR web page (`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 is at the county level, we aggregate it at the commuting zone level. We use 2012 NAICS codes 31-33 to obtain the number of employees in manufacturing.

Trade secrets data. Data on trade secrets are 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 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
EMP	22398.524	61797.154	0.000	1152493.572
# CITES ALL	21.966	347.065	0.000	17344.000
# CITES EXT	2.098	37.218	0.000	2215.000
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, and TSI stand for individual income average tax rate at the ninety-fifth (ninety-ninth, fiftieth) percentile, corporate income tax rate, investment tax credits, R&D tax credits, and trade secrets index, respectively. EMP, # CITES ALL, and # CITES EXT are manufacturing employment, the number of citations by all local inventors, and the number of citations by local 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
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, and TSI stand for individual income average tax rate at the ninety-fifth (ninety-ninth, fiftieth) percentile, corporate income tax rate, investment tax credits, R&D tax credits, and trade secrets index, respectively.

Appendix B Derivation of equation (3)

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

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

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

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

which yields

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

We thus have

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

Setting $\eta=\frac{\alpha}{1+\alpha}$, $\eta'=\frac{\alpha\beta}{1+\alpha}$, $\gamma_d=\frac{1}{1+\alpha}\left[Z_d+\alpha Z'_d\right]$, $\gamma_o=-\frac{1}{1+\alpha}\left[Z_o+\alpha Z'_o\right]$, $\gamma_{od}=-\frac{1}{1+\alpha}\left[C_{od}+\alpha C'_{od}\right]$ and adding u_{odt}, we obtain the expression in (3).

Appendix C Additional robustness checks

C.1 An alternative average tax rate

In the log odds regressions, we proxy for $\tau_{\sigma(d)t}$ with the ATR for a hypothetical taxpayer at the ninety-fifth percentile of the U.S. income distribution. We report an alternative case where we use the ninety-ninth percentile in Table C1. Using specification (2) in Table C1 we construct the three types of Bartik instruments as before and check the robustness of our main results in Table C2. As seen from Table C2, our main semi-elasticities are virtually identical to the baseline cases: 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 an inflow of a top inventor raises local patent productivity by approximately 4% when we focus on external inventors.

Using this alternative ATR, we also conduct the dynamic analysis in Figure C1. Panel (a) illustrates the dynamic impact of inventor inflows on the patent productivity of all local inventors, which includes not only internal knowledge sharing within assignees and between co-inventors but also external knowledge spillovers. Panel (b) corresponds to the case of "the mysteries of the trade in the air", where we exclude co-inventors and inventors in the same

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

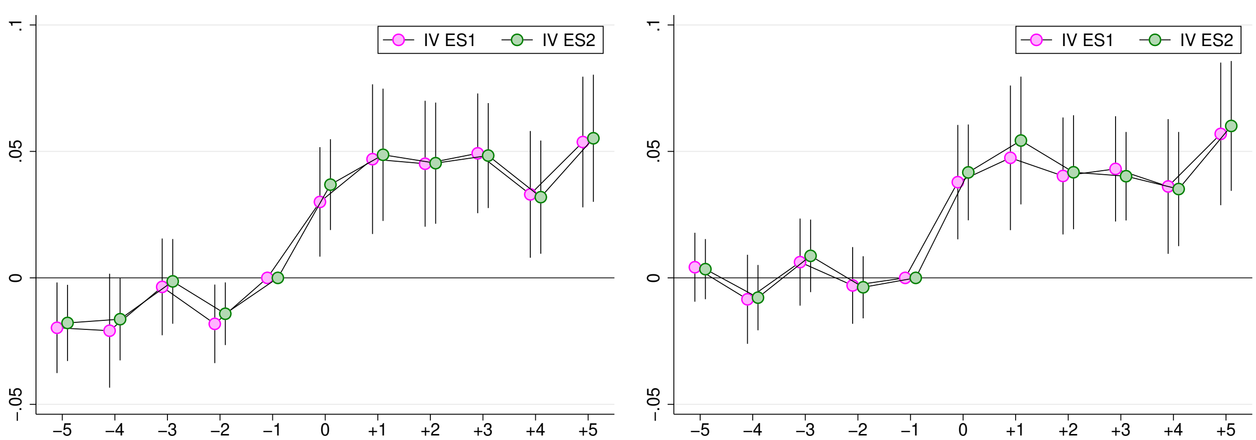
	(1)	(2)	(3)	(4)
$\Delta\ln(1-\text{ATR99})$	2.340	2.504	2.382	2.455
	(1.315)	(1.198)	(1.150)	(0.966)
$\Delta\ln(1-\text{CITR})$	-0.440	-0.251	-0.147	-0.018
	(1.115)	(1.063)	(0.841)	(0.750)
$\Delta\ln(1+\text{ITC})$	-0.700	-0.913	-0.549	-0.711
	(0.719)	(0.672)	(0.948)	(0.666)
$\Delta\ln(1+\text{RTC})$	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
Year FE × Region pair	No	Yes	No	Yes
Observations	4866	4866	7226	7225
$\overline{R}^2$ (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 (3). ATR99, CITR, ITC, and RTC stand for individual income average tax rate at the ninety-ninth percentile, corporate income tax rate, investment tax credits, and R&D tax credits, respectively. Cluster-robust standard errors are in parentheses.

Figure C1: IV event study regressions (robustness).

(a) all local inventors

(b) external inventors



Notes: Panel (a) (Panel (b)) illustrates the dynamic impact of inventor inflows on local patent productivity without (by) excluding co-inventors and inventors in the same assignee. 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 C2: The impact of top inventor inflows on local patent productivity (robustness)

	(1)	(2)	(3)	(4)	(5)	(6)
(a) All local inventors						
Top inventor inflows	0.044	0.064	0.062	0.062	0.059	0.060
	(0.007)	(0.013)	(0.012)	(0.012)	(0.013)	(0.012)
ln(1 − 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	35.170	35.141
$\tau = 5\%$		37.418	31.742	34.735	30.971	33.112
$\tau = 10\%$		23.109	19.781	21.385	19.330	20.431
$\tau = 20\%$		15.062	13.025	13.894	12.751	13.309
$\tau = 30\%$		12.039	10.478	11.086	10.269	10.637
(b) External inventors						
Top inventor inflows	0.027	0.044	0.043	0.044	0.035	0.038
	(0.004)	(0.010)	(0.009)	(0.009)	(0.010)	(0.010)
ln(1 − 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	35.170	35.141
$\tau = 5\%$		37.418	31.713	34.738	30.948	33.112
$\tau = 10\%$		23.109	19.764	21.386	19.318	20.431
$\tau = 20\%$		15.062	13.016	13.895	12.744	13.309
$\tau = 30\%$		12.039	10.471	11.086	10.263	10.636
CZ FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	No	No
Year × state FE	No	No	No	No	Yes	Yes
Observations	23,628	23,628	23,628	23,463	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−ATR99) is converted to elasticity. Cluster-robust standard errors are in parentheses. Column 1 does not control 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 and 6 replace ln(1−ATR99) in Columns 3 and 4 with Year×destination state FE.

assignee. In both cases, the results are quite similar to the baseline cases: the pre-event semi-elasticities are close to zero and post-event semi-elasticities increase to approximately 0.05, which ensures the results of our static model. We report the numbers that we use in Figure C1 and the associated first-stage statistics in Tables C5 and C6, respectively.

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 (EMP) at the commuting zone level. Table C3 shows that our baseline results are virtually identical 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 an inflow of a top inventor raises local patent productivity by approximately 4% when we focus on external inventors.

C.3 Dropping commuting zone × year observations with no patents

We check the robustness of our main results in Table 5 by dropping commuting zone × year observations with no patents, instead of using $\log(1 + Y_{dt})$ as in the baseline case. Table C4 shows that the results are qualitatively similar to our baseline results 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 an inflow of a top inventor raises the patent productivity of external inventors by approximately 3-4%.

Table C3: The impact of top inventor inflows on local patent productivity

	(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)
$\ln(1+\text{EMP})$					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)
$\ln(1+\text{EMP})$					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
Year × state 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. ATR (ATR50) stands for the individual income average tax rate at the ninety-fifth (fiftieth) percentile. The other coefficients are converted to elasticity. Cluster-robust standard errors are in parentheses. Column 1 does not control 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 to 8 repeat the same specifications as Columns 1 to 4 with $\ln(1+\text{EMP})$ at the commuting zone level.

Table C4: The impact of top inventor inflows on local patent productivity

	(1)	(2)	(3)	(4)	(5)	(6)
(a) All local inventors						
Top inventor inflows	0.039	0.054	0.051	0.050	0.051	0.052
	(0.006)	(0.012)	(0.010)	(0.010)	(0.012)	(0.011)
ln(1 − 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	34.258	34.170
$\tau = 5\%$		37.418	31.930	34.704	31.378	33.007
$\tau = 10\%$		23.109	19.892	21.372	19.569	20.377
$\tau = 20\%$		15.062	13.094	13.890	12.897	13.282
$\tau = 30\%$		12.039	10.531	11.085	10.380	10.619
CZ FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	No	No
Year × state FE	No	No	No	No	Yes	Yes
Observations	20,038	20,038	20,038	19,941	19,972	19,875
(b) External inventors						
Top inventor inflows	0.023	0.034	0.032	0.033	0.029	0.031
	(0.004)	(0.009)	(0.008)	(0.008)	(0.009)	(0.009)
ln(1 − ATR)	5.450	5.380	5.357	5.340		
	(1.054)	(1.056)	(1.054)	(1.044)		
CZ FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	No	No
Year × state FE	No	No	No	No	Yes	Yes
Effective F statistic		37.072	32.971	32.651	34.229	34.137
$\tau = 5\%$		37.418	31.917	34.706	31.350	32.995
$\tau = 10\%$		23.109	19.885	21.373	19.553	20.370
$\tau = 20\%$		15.062	13.090	13.891	12.888	13.278
$\tau = 30\%$		12.039	10.528	11.085	10.373	10.615
Observations	19,903	19,903	19,903	19,806	19,837	19,740

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 − ATR) is converted to elasticity. Cluster-robust standard errors are in parentheses. Column 1 does not control 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 and 6 replace ln(1 − ATR) in Columns 3 and 4 with Year × destination state FE.

Table C5: 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	15752	15642	15752	15642	15752	15642	15752	15642

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) to (4) use ATR95, whereas Columns (5) to (8) use ATR99 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 C6: 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
	(0.000)	(0.000)	(0.000)	(0.000)
Top inventor inflows ($j = -4$)	156.024	284.008	160.880	248.067
	(0.000)	(0.000)	(0.000)	(0.000)
Top inventor inflows ($j = -3$)	177.020	308.252	217.101	325.561
	(0.000)	(0.000)	(0.000)	(0.000)
Top inventor inflows ($j = -2$)	163.037	223.759	211.554	254.252
	(0.000)	(0.000)	(0.000)	(0.000)
Top inventor inflows ($j = 0$)	257.894	446.932	204.805	397.033
	(0.000)	(0.000)	(0.000)	(0.000)
Top inventor inflows ($j = +1$)	442.020	477.121	372.415	463.968
	(0.000)	(0.000)	(0.000)	(0.000)
Top inventor inflows ($j = +2$)	208.149	421.991	194.555	381.307
	(0.000)	(0.000)	(0.000)	(0.000)
Top inventor inflows ($j = +3$)	255.615	465.129	277.758	503.246
	(0.000)	(0.000)	(0.000)	(0.000)
Top inventor inflows ($j = +4$)	126.082	266.307	130.618	229.608
	(0.000)	(0.000)	(0.000)	(0.000)
Top inventor inflows ($j = +5$)	271.808	417.144	216.519	338.164
	(0.000)	(0.000)	(0.000)	(0.000)
(b) Sanderson-Windmeijer (weak identification)				
Top inventor inflows ($j = -5$)	30.805	27.280	35.435	31.812
	(0.000)	(0.000)	(0.000)	(0.000)
Top inventor inflows ($j = -4$)	14.127	13.461	14.567	11.758
	(0.000)	(0.000)	(0.000)	(0.000)
Top inventor inflows ($j = -3$)	16.028	14.610	19.658	15.431
	(0.000)	(0.000)	(0.000)	(0.000)
Top inventor inflows ($j = -2$)	14.762	10.606	19.155	12.051
	(0.000)	(0.000)	(0.000)	(0.000)
Top inventor inflows ($j = 0$)	23.351	21.183	18.544	18.818
	(0.000)	(0.000)	(0.000)	(0.000)
Top inventor inflows ($j = +1$)	40.023	22.614	33.721	21.991
	(0.000)	(0.000)	(0.000)	(0.000)
Top inventor inflows ($j = +2$)	18.847	20.001	17.616	18.073
	(0.000)	(0.000)	(0.000)	(0.000)
Top inventor inflows ($j = +3$)	23.145	22.046	25.150	23.852
	(0.000)	(0.000)	(0.000)	(0.000)
Top inventor inflows ($j = +4$)	11.416	12.622	11.827	10.883
	(0.000)	(0.000)	(0.000)	(0.000)
Top inventor inflows ($j = +5$)	24.611	19.771	19.605	16.028
	(0.000)	(0.000)	(0.000)	(0.000)

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) and (2) use ATR95, whereas Columns (3) and (4) use ATR99 as robustness checks.

Appendix D State tax competition

We test the null hypothesis that state governments strategically determine their taxes. Based on the previous studies (e.g., Brueckner, 2003), we estimate a reaction function where a state responds to the choices of its neighboring states as follows:

$$\tau_{\sigma t} = \rho \sum_{\sigma' \neq \sigma} \omega_{\sigma\sigma'} \tau_{\sigma' t} + \boldsymbol{\beta} \mathbf{X}_{\sigma t-1} + \gamma_\sigma + \gamma_t + \varepsilon_{\sigma t}. \tag{16}$$

The dependent variable $\tau_{\sigma t}$ is the tax 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 taxes in the neighboring states with the weight being $\omega_{\sigma\sigma'}$. $\mathbf{X}_{\sigma t-1}$ stands for local patent productivity and a vector of socio-politico-economic characteristics for state σ in year $t-1$ (see Table D1).[33] γ_σ and γ_t denote state and year fixed effects, respectively, and $\varepsilon_{\sigma t}$ is the error term.

Our null hypothesis is that $\rho > 0$. If $\rho > 0$ 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 since the top inventor inflows would not mediate the interstate correlation.

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 $\mathbf{X}_{\sigma t-1} = [x_{1,\sigma t-1}, \cdots, x_{K,\sigma t-1}]'$, 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

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

Table D1: Summary statistics

	mean	sd	min	max	obs
log patent productivity	6.285	1.422	2.881	10.201	1617
log population	15.011	1.009	12.951	17.430	1617
share of black or African American population	0.114	0.121	0.002	0.705	1617
share of population younger than 20	0.292	0.028	0.205	0.414	1617
share of population older than 64	0.124	0.018	0.075	0.184	1617
unemployment rate	5.919	2.046	2.342	17.350	1617
log total income	18.015	1.185	15.010	21.186	1617
log gross state product	18.191	1.170	15.106	21.403	1617
democrat governor (dummy variable)	0.511	0.500	0.000	1.000	1617
log tax revenue	15.291	1.159	12.318	18.581	1584
log government debt	15.302	1.253	10.872	18.819	1584
log government revenue	16.110	1.134	13.301	19.511	1584

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

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 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.[34]

Table D2 reports the regression results for the sample period 1977-2009.[35] Columns 1-2 (3-4) use the first-order contiguity (top-inventor-inflow) weights. Columns 1 and 3 show that the estimated values of ρ are negative and insignificant. Columns 2 and 4 include the own-state productivity and show that it does not significantly affect the estimates of ρ. These results imply that there is neither strategic tax competition nor a direct relationship between tax and productivity across states. 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

[34] In the specification curve analysis below, we consider four other weights: (i) the second-order contiguity weight, i.e., $\omega_{\sigma\sigma'} = 1$ if state σ' is contiguous with 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'; (iii) the inverse-distance weight with a cutoff distance, i.e., the interstate effect is assumed to be zero beyond 1000 miles; and (iv) the top-inventor-inflow weight, i.e., $\omega_{\sigma\sigma'}$ is proportional to the total number of top inventors who migrated from σ' to σ. All weights are normalized such that $\sum_{\sigma'\neq\sigma} \omega_{\sigma\sigma'} = 1$ for any σ.

[35] We proxy for $\tau_{\sigma t}$ with the log of one minus ATR at the ninety-fifth percentile, $\ln(1 - \text{ATR}_{\sigma t})$, in state σ in year t since it significantly affects top inventor migration as shown in Section 3. We drop Washington D.C. from the sample because some state characteristics are unavailable.

Table D2: State tax competition

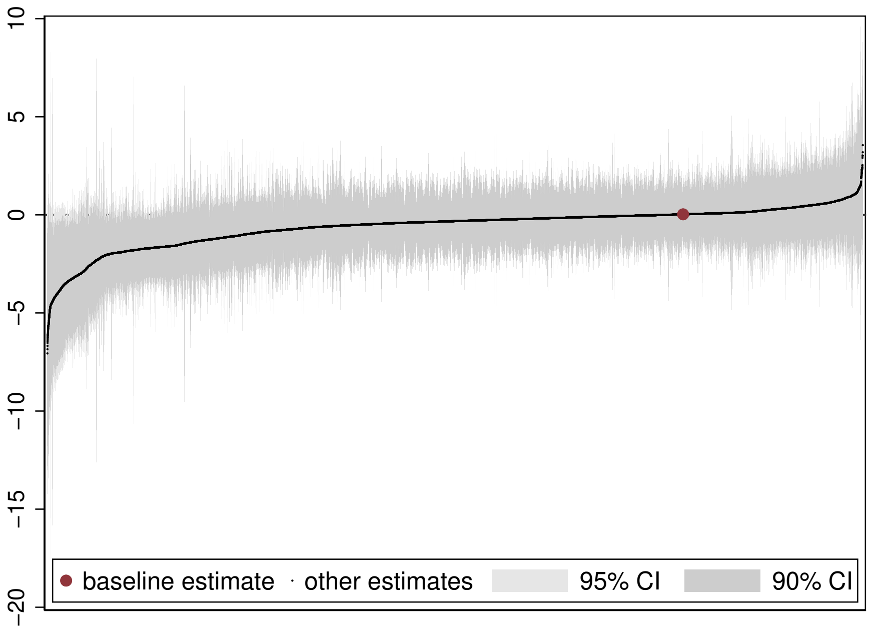
	(1)	(2)	(3)	(4)
$\sum_{\sigma' \neq \sigma} \omega_{\sigma\sigma'} \tau_{\sigma' t}$	-0.163	0.027	-1.364	-1.071
	(0.314)	(0.260)	(0.822)	(0.797)
log patent productivity		0.005		0.005
		(0.003)		(0.003)
log population	0.016	0.016	0.017	0.013
	(0.038)	(0.033)	(0.029)	(0.029)
share of black or African American 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 (3-4) use the first-order contiguity (top-inventor-inflow) weights. Columns 2 and 4 include the own-state productivity $\ln Y_{\sigma t-1}$.

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, Simmons, and Nelson (2020). We consider different specifications of equation (16) by focusing on three dimensions. First, we include four other weights discussed in footnote 34 in the specification curve analysis to examine the sensitivity of the estimates. Second, we estimate the model for every possible combination of the variables listed in Table D1. Last, since

Figure D1: Specification curve



Notes: The specification curve is depicted using 30,843 alternative specifications explained in Appendix D. The vertical axis is the value of ρ.

the model is overidentified, we use different combinations of neighboring states' characteristics as instruments, which allows us to explore the sensitivity of the estimates.[36]

Figure D1 plots the specification curve for ρ with 90% and 95% confidence intervals. The signs of the estimated coefficient are not significantly positive at the 10% level for 30,843 alternative specifications, thus verifying that the baseline estimates do not come from data mining. Hence, we may conclude that there was no strategic tax competition between states during the sample period.

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

Appendix E Assumptions to ensure share exogeneity

We impose the following two assumptions to obtain share exogeneity in Section 4.3. First, ε_{dt} has mean zero conditional on X_{dt}, i.e.,

$$E(\varepsilon_{dt}|X_{dt}) = 0, \tag{17}$$

where X_{dt} consists of the *own state tax* $\tau_{\sigma(d)t}$ and fixed effects $\boldsymbol{\delta} = \{\delta_d, \delta_t\}$. Second, *other state taxes* $\{\tau_{\sigma(c)t}\}_{c \neq d}$ and ε_{dt} are independent conditional on X_{dt}, i.e.,

$$\tau_{\sigma(c)t} \perp \varepsilon_{dt}|X_{dt} \quad \text{for} \quad c \neq d \quad \text{and} \quad c, d \in \mathcal{C}, \tag{18}$$

where $A \perp B|C$ denotes conditional independence of A and B given C. Under these assumptions, we can show that $E(\varepsilon_{dt}\widehat{P}_{odt}|X_{dt}) = E(\varepsilon_{dt}\widehat{P}_{odt}|\tau_{\sigma(d)t}, \boldsymbol{\delta}) = 0$ as follows.

By the property of conditional expectations, we have

$$E(\varepsilon_{dt}\widehat{P}_{odt}|\tau_{\sigma(d)t}, \boldsymbol{\delta}) = E[E(\varepsilon_{dt}\widehat{P}_{odt}|\boldsymbol{\tau}_t, \boldsymbol{\delta})|\tau_{\sigma(d)t}, \boldsymbol{\delta}],$$

where $\boldsymbol{\tau}_t = \{\tau_{\sigma(c)t}\}_{\forall c \in \mathcal{C}}$. The right-hand side of the above equation becomes:

$$\begin{aligned} E[E(\varepsilon_{dt}\widehat{P}_{odt}|\boldsymbol{\tau}_t, \boldsymbol{\delta})|\tau_{\sigma(d)t}, \boldsymbol{\delta}] &= E[\widehat{P}_{odt}E(\varepsilon_{dt}|\boldsymbol{\tau}_t, \boldsymbol{\delta})|\tau_{\sigma(d)t}, \boldsymbol{\delta}] \\ &= E[\widehat{P}_{odt}E(\varepsilon_{dt}|\tau_{\sigma(d)t}, \boldsymbol{\delta}, \{\tau_{\sigma(c)t}\}_{c \neq d})|\tau_{\sigma(d)t}, \boldsymbol{\delta}] \\ &= E[\widehat{P}_{odt}E(\varepsilon_{dt}|\tau_{\sigma(d)t}, \boldsymbol{\delta})|\tau_{\sigma(d)t}, \boldsymbol{\delta}] \\ &= E[\widehat{P}_{odt} \cdot 0|\tau_{\sigma(d)t}, \boldsymbol{\delta}] = 0. \end{aligned}$$

The first equality follows because P_{odt} is a function of $\boldsymbol{\tau}_t$ and $\boldsymbol{\delta}$. The second equality holds because $\boldsymbol{\tau}_t = \{\tau_{\sigma(d)t}, \{\tau_{\sigma(c)t}\}_{c \neq d}\}$. The third equality is due to the conditional independence assumption (18). The last equality comes from the conditional mean assumption (17).

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 IV estimators, $\widehat{\phi}^s = \sum_{o \in \mathcal{C}} \widehat{\omega}_o \widehat{\phi}^s_o$. In our context, the coefficient $\widehat{\phi}^s_o$ is the origin-specific productivity effect using the share $\widehat{P}_{odt}$ as a separate instrument. The Rotemberg weight $\widehat{\omega}_o$ measures how sensitive the estimate $\widehat{\phi}^s$ is to possible misspecification in each instrument.

Table F1 presents the summary statistics.[37] 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 times t. Panel C reports origin commuting zones with the top five highest Rotemberg weights. These commuting zones account for 26.3 percent of the total share.

Panel D provides the heterogeneous effects interpretation of the Bartik instrument. The Bartik estimator can be rewritten as $\widehat{\phi}^s = \sum_d \widehat{\phi}^s_d \sum_o \widehat{\omega}_o \widehat{v}_{od}$, where $\widehat{\phi}^s_d$ is the destination-specific productivity effect and $\widehat{v}_{od} \geq 0$ 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 $\sum_o \widehat{\omega}_o \widehat{v}_{od}$ 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 negative part $\sum_{o|\widehat{\omega}_o<0} \widehat{\omega}_o \widehat{\phi}^s_o$, the negative Rotemberg weights are unlikely to be a problem for the LATE-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 most migrations are predicted to be interstate aligns with the assumption that the main source of identifying variation comes from inventor mobility caused by individual income tax differences across states.

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

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

	$\widehat{\omega}_o$	$\bar{I}_o$	$\widehat{\phi}_o^s$	$\widehat{F}_o$	$\text{var}(\widehat{P}_o)$
$\widehat{\omega}_o$	1.0000				
$\bar{I}_o$	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	
$\text{var}(\widehat{P}_o)$	0.7865	0.7801	-0.0417	-0.0845	1.0000

Panel C: Top five Rotemberg weight commuting zones

	$\widehat{\omega}_o$	$\bar{I}_o$	$\widehat{\phi}_o^s$	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 reports 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 ($\bar{I}_o$), 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 times ($\text{var}(\widehat{P}_o)$). 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 parenthesis, where the representative county is the one with the largest number of inventors. The ninety-five percent 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.

We finally examine the relationship between the top five shares $\widehat{P}_{odt}$ and location-specific characteristics that may be correlated with the outcome Y_{dt} as suggested by Goldsmith-Pinkham et al. (2020). For the shares $\widehat{P}_{odt}$ to satisfy the conditional exogeneity assumption in Section 4.1, they should not be correlated with destination commuting zone characteristics conditional on the controls and fixed effects in the structural equation (6).[38] Based on the

[38]Let W_{dt} be destination commuting zone characteristics that may be correlated with Y_{dt}. No correlation between the shares and destination characteristics, $\text{corr}(\widehat{P}_{odt}, W_{dt}|X_{dt}, \boldsymbol{\delta}) = 0$, is an observable analogue of $\text{corr}(\widehat{P}_{odt}, \varepsilon_{dt}|X_{dt}, \boldsymbol{\delta}) = E(\widehat{P}_{odt} \cdot \varepsilon_{dt}|X_{dt}, \boldsymbol{\delta}) = 0$ under the conditional mean zero assumption $E(\varepsilon_{dt}|X_{dt}, \boldsymbol{\delta}) = 0$.

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

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

Notes: The top five destination commuting zones are defined by the predicted migration probability of the top inventors from each origin commuting zone. The representative county is the one with the largest number of inventors.

Table F3: Relationship between the shares and manufacturing employment size

	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 coefficient	0.0002763	-0.0027881	0.0013622	-0.0018730	0.0026412
Regression coefficient	-0.0000013	-0.0000209	0.0000071	-0.0000080	0.0000201
	(0.0000108)	(0.0000128)	(0.0000143)	(0.0000112)	(0.0000105)

Notes: The first row reports the correlation coefficient between the first-order time difference in the share and the first-order time difference in the log of manufacturing employment. The second row reports the regression coefficient of the log of manufacturing employment on the share conditional on $\ln(1 - \text{ATR})$, commuting zone fixed effects, and year fixed effects. To control for commuting zone fixed effects in the regression, we take the first-order time difference of the variables in the regression. Cluster-robust standard errors are in parentheses.

result in Table C3, we regard the log of manufacturing employment in destination commuting zone d as such a characteristic. Table F3 presents the results of this analysis. We take the time difference of the variables to control for commuting zone fixed effects. Reassuringly, the correlations are low in all cases and the regression coefficients on the log of manufacturing employment are not statistically significant at the five percent level, thus implying the absence of correlation between the shares and possible confounders.

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 $\Delta = \{\delta : |(\delta_{t+1} - \delta_t) - (\delta_t - \delta_{t-1})| \leq M, \forall t\}$ to bound the degree to which the slope of differential trend δ 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 illustrates the robustness test result, where we assess the presence of a differential trend δ_1 for the first-period effect μ_1 under the normalization of $\mu_{-1} = 0$. In Panel (a)

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

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

Notes: Panel (a) (Panel (b)) presents 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 (external inventors). 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.

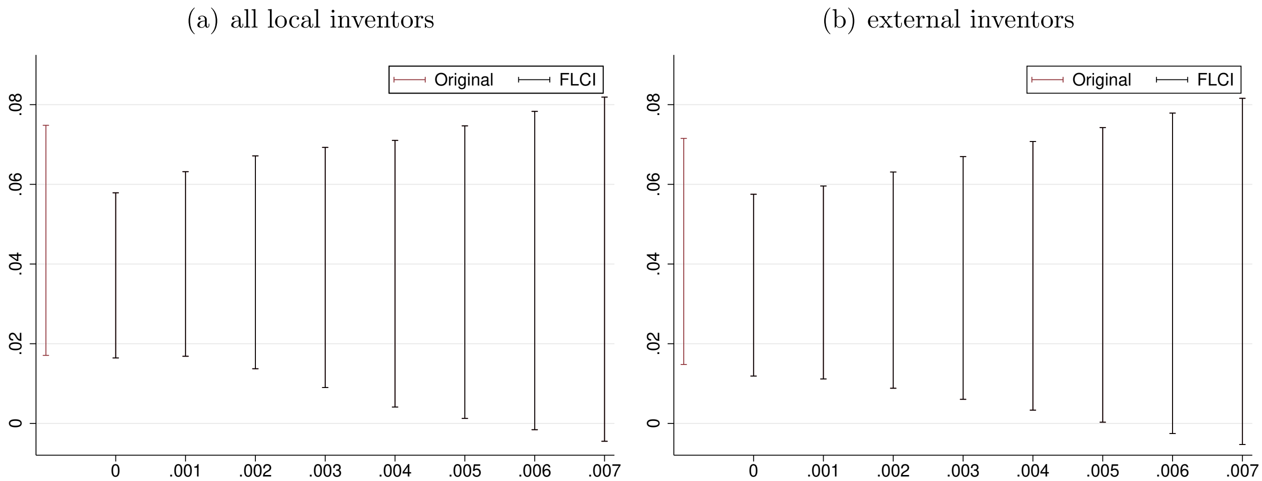
(Panel (b)), the "break-down" value of M, at which the null hypothesis that the first-period effect is zero can no longer be rejected, is 0.006 (0.006). Since the value is 40 (42.90) percent of the standard error of the estimated effect $\hat{\mu}_1$, the parallel trends assumption holds for a reasonable deviation from a linear trend.[39]

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 ninety-five percent 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.

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

Figure G1: Robustness to possible violations of parallel trends assumption



Notes: Panel (a) (Panel (b)) illustrates the sensitivity analysis proposed by Rambachan and Roth (2023) for the event study model, where we consider all local inventors (external inventors). In each panel, the leftmost bar is the ninety-five percent confidence interval of the estimate $\hat{\mu}_1$ and the other bars are fixed length confidence intervals (FLCIs) considered Rambachan and Roth (2023).