Can Industrial Policy Overcome Coordination Failures? Theory and Evidence*

Updated Regularly. Link to the Latest Version

Tishara Garg†

July 8, 2025

Abstract

This paper introduces a method to study the impact of policy events on equilibrium selection in settings where strong complementarities may lead to multiple equilibria and coordination failures. Many industrial policies are rooted in coordination failures and 'big-push' theories, yet empirical evidence on their effectiveness remains limited since distinguishing equilibrium shifts from direct changes in fundamentals is challenging. Leveraging tools from industrial organization and algebraic geometry, I develop an approach to study the coordination effects of policy without imposing strong assumptions on the distribution or responsiveness of economic fundamentals. The method identifies the 'types' of factual and counterfactual equilibria before and after the policy event through a three-step procedure: model estimation and inversion, equilibrium enumeration, and type assignment. Types of factual equilibria may be used to examine how policies like urban infrastructure investment, subsidies, or trade liberalization, affect equilibrium selection. Types of counterfactual equilibria help identify the extent to which observed effects of policy are driven by fundamentals or coordination. I apply this method to study industrial zones in India. Using a newly constructed data on 4,000 industrial zones, I find that municipalities receiving a zone see a 60% increase in non-farm employment over 15 years, with significant spillovers to sectors and municipalities not targeted by the zones. Combining the type assignment methodology with event study designs, I find that industrial zones increase the probability of escaping a low-industrialization equilibrium by 38%, with coordination effects explaining roughly one-third of the observed change in outcomes.

*I am grateful to my advisors — David Atkin, Abhijit Banerjee, Dave Donaldson, and Tobias Salz — for their continuous support and guidance. I thank Treb Allen, Isaiah Andrews, William Banner, Isabel Di Tella, Esther Duflo, Glenn Ellison, Teresa Fort, Apoorv Gupta, Ananya Kotia, Vishan Nigam, Aureo de Paula, Ashesh Rambachan, Satoru Takahashi, Jose Vasquez, Anna Vitali, Ivan Werning, Nathan Zorzi, participants at the MIT Trade, Development and IO lunch, and the participants at the Globalization Seminar at Dartmouth College for their helpful comments and feedback. I extend my gratitude to numerous Indian bureaucrats, most notably Krishna Bhaskar, Amit Govil, Jayesh Ranjan, Bipin Kumar, Arindam Misra, Arbind Modi, Ritwik Pandey and Balasubramanian Krishnamurthy, for generously sharing their institutional insights on Indian industrial policy. Disha Chauhan, Asad Malik, and Qirui Wang provided excellent research assistance. I thank Treb Allen and David Atkin for sharing the historical maps of India's road network. I also acknowledge support from the George and Obie Shultz Fund at MIT.

†Department of Economics, MIT, Email: tishara@mit.edu

1 Introduction

"No coordination, no—or very little—economic development."

Ministry of Production, Peru, in *Mesas Ejecutivas,* July 2016

Industrial policy has long been used to promote industrialization and growth, with recent examples including the CHIPS Act in the United States and the Production Linked Incentive (PLI) Scheme in India. Two efficiency-based rationales support such policies. First, there is a rationale for the government to direct resources toward sectors subject to external economies of scale. Second, in the presence of multiple equilibria, the government can coordinate actions to select a more favorable equilibrium.[1] This coordination rationale, rooted in the seminal work of Rosenstein-Rodan (1943), is particularly relevant for developing economies that may be trapped in a low-industrialization equilibrium. This paper focuses on the second rationale.

Two key challenges have limited empirical study of the coordination rationale. First, equilibrium selection is not directly observable, leading to the possible misattribution of the policy effects. For instance, an increase in industrial activity could result from either an improvement in fundamentals—such as market access—or a movement toward a high-industrialization equilibrium with no change in fundamentals. Second, we lack a framework to compare the equilibrium outcomes of policies that, in addition to switching equilibria, can shift fundamentals —e.g., by building new infrastructure. In this paper, I address these challenges by developing a novel methodology that can be broadly applied to study contexts with multiple equilibria and coordination failures. In particular, my work tests the equilibrium-switching rationale for 'big push' industrial policies. To my knowledge, this is the first empirical study of this kind.

I build on the literature on estimating games and the field of algebraic geometry to address the two methodological challenges posed by multiple equilibria.[2] First, I develop identification arguments and an estimation algorithm to map observed equilibrium outcomes to (a) unobserved fundamentals and (b) the set of possible equilibria. Establishing the set of conditions under which this is possible, while leaving the equilibrium selection rule entirely unspecified, is a valuable contribution because deriving observable implications in a model with multiple equilibria is difficult, as famously noted by Jovanovic (1989). Second, building on Aguirregabiria

[1] Coordination can be explicit, such as Peru's "Mesas Ejecutivas," where the government organizes roundtables with industry leaders to address coordination failures (Ghezzi, 2017). It can also be implicit, where firms invest expecting that industrial policy will attract investment from others. Such implicit coordination was especially noted in my interviews conducted as part of this research. One senior bureaucrat remarked, "Firms are attracted to industrial parks for the same reason home seekers are drawn to residential developments—expectations of others like them nearby" (Ritvik Pandey, Secretary, Finance Commission, Government of India, May 31, 2023).

[2] See De Paula (2013) for a review of the literature on estimating games under multiplicity. See Sommese et al. (2005) for an introduction to the methods in algebraic geometry.

and Mira (2019), I introduce a classification system to systematically assign labels, or 'types' (e.g., high or low industrialization), to every equilibrium in the set of possible equilibria. These labels retain their meaning across different economies with different fundamentals. The equilibrium type selected in the data can then be regressed on policy variables to identify how policy affects equilibrium selection. The classification of counterfactual equilibria enables the decomposition of policy-induced changes into coordination- versus fundamentals-driven.

I apply this methodology to study the coordination effects of industrial zones, a popular yet understudied tool of industrial policy. Announcements of these zones may generate convergent expectations around high investment, potentially leading to self-fulfilling prophecies through economies of scale and density. However, distinguishing these coordination effects from other factors, such as land subsidies and infrastructure improvements, remains challenging. The methodology developed in this paper is well-suited to address this challenge. I construct a novel dataset of over 4,000 industrial zones in India, allowing me to exploit the regional and intertemporal variation induced by numerous localized policy experiments. I find that these zones increase industrial employment by around 60% in 15 years, with about a third of the shift attributable to equilibrium switching.

The analysis is structured into three sections. Section 2 develops the methodology within a general framework, presenting a three-step estimation procedure guided by each step's respective theoretical results. Step 1 uses variation across regions or markets to estimate the underlying complementarities and invert unobserved fundamentals. This step is based on a novel identification result that formalizes the insight from Ahlfeldt et al. (2015) that a structural relationship between equilibrium outcomes can be identified under appropriate exclusion restrictions, even when the reduced-form mapping from primitives to equilibrium is non-unique. For example, an exogenous increase in labor demand in location A that raises employment in location B implies positive spillovers from A to B. If A initially has high employment and B has low despite positive spillovers from A to B, B has weak fundamentals in the baseline. Whether or not the spillovers are strong enough to generate multiple equilibria does not have a bearing on this argument—moment conditions alone can identify spillovers and fundamentals.[3]

Step 2 enumerates all the equilibria in all the markets. It is often challenging to find all possible solutions of a non-linear system of equations. In my application, I find that standard methods like Newton's method or fixed-point iteration are computationally infeasible. In order to solve the system, I establish the conditions under which equilibria can be approximated by

[3]This generalizes the results in Ahlfeldt et al. (2015), who estimate spillovers and recover fundamentals as structural residuals despite the possibility of multiple equilibria by leveraging (a) the log-additive structure of fundamentals (their Proposition 2) and (b) the Berlin Wall shock (their Section 7.2). My general result draws on the insight that the orthogonality and completeness conditions for non-parametric identification of structural equations, as in Newey and Powell (2003), can hold despite equilibrium multiplicity.

the roots of a polynomial system. While enumerating all equilibria in non-linear systems is difficult, the problem of finding all the roots of polynomial systems is well-studied in mathematics. Continuation methods from algebraic geometry enable the enumeration of all equilibria in every market, with the empirically observed equilibrium belonging to the equilibrium set.[4]

Step 3 compares equilibria across different markets with varying policy treatments. I formalize the concept of equilibrium types to enable such comparison, establish the conditions for recovering these types, and propose a practical algorithm for type assignment. Assigning equilibrium types is necessary for answering how policy affects coordination in settings where policy can change fundamentals and, hence, the game itself. Two equilibria are defined as belonging to the same type if a continuous path in the fundamentals space connects them—i.e. if one market can be continuously deformed to another without discontinuous jumps in the equilibrium outcomes.[5] For example, in a simple economy with a traditional and a modern sector, equilibria can be classified into 'high,' 'medium,' and 'low' industrialization types. I show that the type assignment function is (a) well-defined, providing a unique mapping from equilibria to types, and (b) invertible, enabling counterfactual analysis. Assigning types to the entire set of equilibria—both factual and counterfactual—enables the researcher to (a) study how policy events affect equilibrium selection, and (b) quantify the extent to which policy-induced changes in equilibrium outcomes are due to coordination effects or shifts in fundamentals.

Section 3 then turns to the application. Industrial zones have been used widely as tools of industrial policy, both in India (Sanghvi, 1979) and in other countries around the world (UNIDO, 2020). The decentralized nature of zones in India provides independent policy variation but also presents data challenges. There is virtually no centralized database of industrial zones in India, as they are developed independently by local development corporations. My first empirical contribution is constructing a novel dataset that addresses this challenge. I collect information on geocodes for 4,000 industrial zones, along with dates of establishment for approximately 1,500 of them.[6] I combine this dataset with economic and population censuses at the village and

[4]The use of continuation methods in economics is rare but not new. A few papers in Industrial Organization have employed homotopy continuation methods for discrete games (Eaves and Schmedders, 1999; Bajari et al., 2010b). Their application in social interaction models is novel, with the exception of contemporary work by Ouazad (2024). While an excellent review of the continuation method, Ouzad's central contribution—using a symmetric city as a start system—has limitations for empirical studies of equilibrium selection, as it is not guaranteed to find *all* equilibria. The algorithm may miss some equilibria because the total degree of the polynomial system characterizing the symmetric city is lower than that of the target city.

[5]The concept of equilibrium types as a relation was introduced in Aguirregabiria (2012) and Aguirregabiria and Mira (2019). However, the extant notion of equilibrium types is not sufficient for empirical analysis. To take this notion to data, we must have that the types (a) are recoverable from data and (b) allow for a well-defined counterfactual analysis of how equilibrium allocation changes when types change. I therefore further develop this notion by defining the conditions under which the 'equilibrium type' relation becomes a function—(a) well-defined, in that each equilibrium maps to one type, and (b) invertible, in that each type maps to one equilibrium given fundamentals. The idea of recovering these types as a latent variable from the data is, to my knowledge, novel.

[6]While the event studies in this paper are based on a subset of 1,500 parks established during my sample

town (henceforth municipality) levels to construct measures of industrial activity. To account for observed fundamentals, I complement the dataset with novel spatial data on historical road networks and natural attributes. The primary event of interest is the first time a municipality receives an industrial zone.

I present three stylized facts on industrial zones. Fact 1 shows that receiving an industrial zone leads to a sizable increase in industrial employment, which, in the absence of coordination failures, would require large changes in fundamentals. I demonstrate this with event studies that compare the share of industrial workers in municipalities that receive an industrial zone with control municipalities that (a) never receive an industrial zone, (b) are comparable to treated municipalities in baseline characteristics (operationalized using propensity score matching), and (c) are sufficiently distant (>25km away) from treated municipalities. The three restrictions are aimed at allaying concerns regarding negative weights, endogenous selection, and SUTVA, respectively.[7] Baseline event studies reveal that establishing an industrial zone increases the number of industrial jobs by approximately 100 industrial workers per square kilometer within 15 years—a 60% increase from the baseline. Fact 2 turns to spatial and sectoral spillovers, a critical component of the analysis, as they inform the extent of scale economies in the structural estimation. I compare the catchment areas of treated and control municipalities and find significant positive spatial spillovers extending up to 10-15 kilometers. Additionally, I observe substantial positive spillovers of sector-specific zones to the non-targeted sectors. Facts 1 and 2 together establish the average treatment effect and spillover effects of zones. Fact 3 examines the distribution of these effects. I find that the conditional and unconditional distributions of the treatment effect exhibit bimodality, further suggesting the possibility of equilibrium switching and setting the stage for the empirical analysis that follows in Section 4.

Section 4 implements the three-step estimation procedure outlined in Section 2 in the context of industrial zones described in Section 3. The goal is to recover each region's entire set of equilibria and assign them types. First, I lay out a parsimonious model of the allocation of economic activity across sectors and space. I estimate the model's spillover parameters using zones as instruments, and then I invert the model to recover unobserved fundamentals. The empirical findings from Section 3 on the direct and spillover effects of policy (Fact 1 and 2)

period (1980–2019), information on the broader cross-section helps discipline the set of controls.

[7]Correct inference relies on the Stable Unit Treatment Value Assumption (SUTVA), potentially violated if treated municipalities systematically drew workers from control municipalities. Such a violation is unlikely in my analysis for three reasons. First, I find that any spatial spillovers of zones, positive or negative, dissipate beyond 15km, while control municipalities are chosen to be at least 25km from any zone. Second, by excluding Special Economic Zones (SEZs) and Export Promotion Zones (EPZs), I focus on highly localized industrial parks, likely to attract workers only from nearby villages. Lastly, with around 600,000 municipalities in India and only under 2% in my sample, any SUTVA violation would require treated municipalities to systematically draw workers from a small set of matched controls—a scenario ruled out by a series of robustness checks on the matching algorithm that I report in the appendices.

serve as critical moments in this step. The estimates are consistent under two identification assumptions. The first is the standard parallel trend assumption: conditional on a host of natural attributes, the decision to establish a zone in a particular location is orthogonal to trends in the non-targeted sectors and locations. This assumption is validated by the absence of preexisting differential trends in an event study framework. The second assumption is that, aside from the construction of new infrastructure, industrial parks do not directly benefit untargeted sectors in untargeted locations. For instance, this assumption would be violated if zones provided direct subsidies to the service-sector firms outside the zone—an unlikely scenario in this context.

With the model estimated and inverted, I proceed to compute all the possible equilibria for each region using homotopy continuation methods of solving polynomial systems. In regions consisting of L locations, I find a maximum of $2*L+1$ possible equilibria: one corresponds to low industrialization everywhere, and the others represent varying degrees of industrialization with economic activity agglomerating in one of the L locations. Finally, I apply the type assignment algorithm developed in Section 2 to classify equilibria. The results indicate that around 80% of the regions are in a low-industrialization equilibrium, but for 17% of those, a high-industrialization type is possible, underscoring the potential role of coordination.

After assigning types to all equilibria, I conclude Section 4 with two final analyses. First, I treat the selected equilibrium types as outcomes and examine the impact of industrial zones on these outcomes in an event study framework. I focus on regions that were stuck in a low-industrialization equilibrium in the baseline—those where the low equilibrium was realized, even though a high equilibrium remained possible. I find that industrial zones do act as coordination devices as hypothesized—receiving a zone increased the likelihood of switching to a high-industrialization equilibrium over the next 10 years by approximately 38%. In other words, municipalities that receive industrial zones are 38% more likely to transition to a high-industrialization equilibrium compared to those that do not. Second, I use the type assignments to decompose the observed effects into coordination and fundamentals channels. The equilibrium-switching mechanism explains a significant portion of the reduced-form effect of zones on industrial activity, accounting for about a third. These results confirm the role of industrial zones as a coordination device.

Contribution to the Literature: The topic of industrial policy is fraught with controversy, and, as the survey article by Juhász et al. (2023) emphasizes, disagreements revolve not around theoretical rationales but practical limitations. The theoretical idea of big push—that industrial policy can lift economies out of Pareto-inferior equilibria—is well-established and rooted in the seminal works of Rosenstein-Rodan (1943), Myrdal (1957), and Hirschman (1958).[8] However,

[8]The arguments were further formalized by Murphy et al. (1989); Krugman (1991); Matsuyama (1991); Rodrik (1995, 1996); Adsera and Ray (1998); Frankel and Pauzner (2000); Hoff and Stiglitz (2000); Ciccone (2002), among others. Rodrik (1995) argues for the role of coordination in explaining the East Asian Growth Miracle.

barring a few quantification studies using calibrated models applied to a single policy event (Buera et al., 2021; Becko, 2023; Choi and Shim, 2024), empirical evidence of this idea remains scarce. The primary contributions of this paper are (a) the development of novel tools to evaluate whether real-world industrial policies have successfully facilitated equilibrium shifts, and (b) the application of these tools to study an important industrial policy instrument, industrial zones.[9]

The three-step methodological approach proposed in the paper, although motivated by the need to study the coordination role of industrial policy, is applicable to studying the determinants of equilibrium selection in other contexts with multiple equilibria as well. For instance, the methodology can be applied to study the role of multiplicity in shaping the structure of cities (Owens III et al., 2020; Monte et al., 2023; Redding et al., 2011) and patterns of technology adoption (Crouzet et al., 2023; Higgins, 2024; Hornbeck et al., 2024). Although there is a long tradition of papers within econometrics and industrial organization focusing on the identification of games with multiple equilibria (Jovanovic, 1989; Manski, 1993; Tamer, 2003; Ciliberto and Tamer, 2009), my work makes a distinct contribution. While effective for analyzing games between a small number of firms within a narrow sector, the existing approaches have limitations when evaluating general equilibrium interactions on a macroeconomic scale. The literature either abstracts from unobserved heterogeneity across markets (Seim, 2006; Sweeting, 2009; De Paula and Tang, 2012) or imposes strong restrictions on it.[10] Consequently, the identification challenge—distinguishing equilibrium switching from improvements in fundamentals—remains unaddressed. I develop a flexible, moment-based approach that addresses these limitations.[11]

Another contribution of this paper is its focus on the determinants of equilibrium selection. While multiple equilibria are often treated as a nuisance in the estimation of games, this paper argues that multiplicity represents an empirical reality with both positive and normative implications. The closest related work in its empirical focus on equilibrium selection is Bajari et al. (2010a), who study the frequency with which an efficient equilibrium is selected in a discrete auction game. My paper departs from theirs in both the underlying research question and the

[9]The empirical part of this paper contributes to the growing empirical literature on Industrial Policy across diverse contexts, including Chinese shipbuilding (Kalouptsidi, 2018; Hanlon, 2020; Barwick et al., 2023), the French cotton industry (Juhász, 2018), Indian small-scale firms (Rotemberg, 2019), Romania's IT sector (Manelici and Pantea, 2021), South Korea's HCI drive (Choi and Levchenko, 2021; Lane, 2022; Kim et al., 2021), and China's Great Firewall (Zhou, 2024).

[10]Aguirregabiria and Mira (2019), for example, develop identification arguments for games with multiplicity under the assumption that unobserved heterogeneity has a discrete, finite, and low-dimensional support.

[11]The method proposed in this paper imposes minimal assumptions on equilibrium selection or unobserved heterogeneity. However, it is primarily applicable to games of social interaction—the kind studied by Manski (1993), Brock and Durlauf (2001), or more recently by Allen et al. (2020). The approach depends on the researcher's ability to observe equilibrium outcomes, such as players' choice probabilities or, in my case, labor shares. In games with a small number of players, where players' choice probabilities must first be estimated by aggregating data across markets (e.g., entry or auction games), strong assumptions about unobserved heterogeneity or equilibrium selection are often necessary. For a detailed review of estimation methods for discrete games, see Aguirregabiria et al. (2021) and Bajari et al. (2024).

methodological approach. Whereas Bajari et al. focus on a descriptive analysis of equilibrium selection, my work addresses a causal question: how do policy events *affect* equilibrium selection? Methodologically, Bajari et al. specify a selection rule and jointly estimate it with the rest of the game using a likelihood-based approach. Although appropriate for a narrow and well-understood sector, such a method is not suitable for studying a macroeconomic setting that encompasses various sectors. In contrast, I take a sequential approach. First, I recover the underlying equilibrium type as a latent variable without making any assumptions about the selection rule. Only in the final step do I introduce the equilibrium selection function, allowing for its non-parametric identification.

Finally, this paper contributes to the study of place-based industrial policies (Kline and Moretti, 2014; Neumark and Simpson, 2015; Criscuolo et al., 2019) in general and industrial zones in particular. Although there is a small literature examining industrial zones in various contexts—such as SEZs in the U.S. by Grant (2020), Chinese SEZs by Lu et al. (2019), industrial areas in Karnataka by Blakeslee et al. (2021), Indian SEZs by Gallé et al. (2022), and 110 parks in 8 Chinese cities by Zheng et al. (2017)—this study offers the first comprehensive analysis of industrial zones.[12] The richness of the data enables me to employ rigorous event study designs, where I construct disciplined sets of comparison municipalities to address concerns related to endogenous selection and SUTVA effectively. The novel data collected as part of this research opens new avenues for future empirical analysis of industrial zones, providing valuable insights into their impacts, effectiveness, and optimal design.

2 Identification Results and Estimation Procedure

As highlighted in the introduction, identification of the underlying complementarities, enumeration of equilibria, and comparison of equilibrium outcomes across different markets are all complicated by the presence of multiple equilibria. The goal of this section is to develop a method that addresses these challenges. Importantly, I develop a formal equilibrium classification system and a method for type assignment. This method recovers a) the type of the equilibrium selected in the data, enabling the study of the equilibrium selection rule, and b) the types of the counterfactual equilibria, enabling decomposition of the observed effects into fundamentals- and coordination-driven. In what follows, I first introduce the class of models to which the proposed estimation procedure is applicable. I then provide the three-step estimation procedure for equilibrium types.

[12]For both conceptual clarity and identification purposes, I exclude the few SEZs and EPZs from the analysis, providing insights into a policy tool more widely utilized by local governments worldwide.

Set-up

I focus on the class of models whose equilibrium can be described by a set of equations of the form:

$$y_i = F_i(\mathbf{y},\mathbf{s}) \qquad \forall i = 1 \cdots n \tag{1}$$

where $\mathbf{y} \equiv \{y_i\} \in Y \subseteq \mathbb{R}^n_{++}$ are the equilibrium outcomes. Depending on the application, these outcomes may represent entry probabilities of firms of n sectors, the allocation of labor across n locations, public good provision by n demographic groups, or migration and trade flows between n region- or country-pairs. Primitives of the economy, such as the costs of entry, the fundamental productivity of locations, the preferences of households, migration frictions, and tariffs, are summarized by a vector of fundamentals $\mathbf{s} \in S \subset \mathbb{R}^m$. $F : Y \times S \to \mathbb{R}^n$ is the system of structural equations governing interactions between endogenous outcomes $\mathbf{y}$.

The defining feature of the model is that the outcomes corresponding to the ith element functionally depend on the entire vector of outcomes. However, it might not be possible to reduce the system to express outcomes $\mathbf{y}$ as *functions* of only the fundamentals due to the presence of multiple equilibria. This encompasses models with general equilibrium interactions and external economies of scale, strategic complementarities in games, or peer effects in social interactions. I focus on the models that satisfy the following properties.

Denote $G(\mathbf{y},\mathbf{s}) = \mathbf{y} - F(\mathbf{y},\mathbf{s})$. Denote $A_F = \{(\mathbf{y},\mathbf{s}) \in Y \times S,\, G(\mathbf{y},\mathbf{s}) = 0)\}$

Assumption 1. *The set Y is compact and convex in $\mathbb{R}^n$.*

1. *F is twice continuously differentiable.*
2. ***Transversality:** The Jacobian of G with respect to $(\boldsymbol{y}, \boldsymbol{s})$, denoted with $\nabla G(\boldsymbol{y},\boldsymbol{s})$, is full rank for all $\boldsymbol{y}, \boldsymbol{s} \in A_F$*
3. ***Interior solutions:** $(\boldsymbol{y},\boldsymbol{s}) \in A_F \implies \boldsymbol{y} \notin \partial Y$*

The outcomes will naturally belong to a bounded simplex if the outcomes reflect choice probabilities or equilibrium allocations. The transversality condition ensures that the graph $\{\mathbf{s},\mathbf{y}; I - F(\mathbf{y}) = 0\}$ is a smooth manifold. Prices can be made to lie in a bounded simplex by normalizing, for example, the sum of all prices to unity. Inada conditions typically ensure that the equilibria lie in the interior of the simplices.[13]

An equilibrium of the system is a vector of outcomes $\mathbf{y}^* \equiv \{y_1^* \cdots y_n^*\}$ such that $y_i^* = F_i(\mathbf{y}^*,\mathbf{s})$ for all i. Let's denote the set of equilibria for a given state $\mathbf{s}$ by $\phi(\mathbf{s})$. Brower's Fixed Point Theorem

[13] Importantly, this class of models does *not* cover complete information games with a finite number of players and a discrete action space, unless the attention is restricted to mixed strategy equilibria. The pure strategy equilibria in such games lie in a discrete space, violating the convexity assumption.

ensures that $\phi(\mathbf{s})$ is not empty for any $\mathbf{s} \in S$, but there is a possibility of equilibrium multiplicity. It is possible that for some $\mathbf{s}$, the set $\phi(\mathbf{s})$ contains more than one element.

The goal of the section is to provide a methodology that assigns equilibrium types to all elements of $\cup_{\mathbf{s}\in S}\phi(\mathbf{s})$ under the assumption that the observed equilibrium belongs to this set. Before I introduce the methodology, I provide an example from economic geography that fit the class of models outlined above. Another example from the field of Industrial Organization is provided in Appendix A.

Example: Spatial Allocation of Labor

I use a simplified version of the structural model used in Section 4 as a running example throughout this Section. Consider a model of spatial allocation of labor where the economy is populated by a continuum of workers of unit mass, who simultaneously decide where to work. Each worker chooses between different locations indexed by $i \in I$, as well as an outside option. Further assume that the consumption good is perfectly tradable such that labor supply decisions depend solely on wage differentials. There is a worker-specific amenity component that follows a Type 1 Extreme Value distribution, such that the aggregate labor supply in location i, denoted with y_i, is given by the function

$$y_i = \frac{w_i^\rho}{1+\sum_{i'} w_{i'}^\rho} \qquad \forall i \in I,$$

where w_i is the wage in location i, and ρ parametrizes the labor supply elasticity. The return from the outside option is normalized to unity. Each location is populated by a representative firm that produces output using labor as the only input, taking wages as given. The output of the firm in location $i \in I$ is given by $A_i y_i$, where A_i denotes the productivity in that location. The price for goods produced by firm i is denoted by p_i and is assumed to be set exogenously in the national goods markets. In equilibrium, workers are paid their marginal product: the wage in location i is given by $w_i = p_i A_i$.

A_i features external economies of scale, such that $A_i = g_i(\boldsymbol{y}, s_i)$, where s_i captures the fundamental productivity of location i. The existence of spillovers creates the possibility of an upward-sloping labor demand curve, given by

$$w_i = p_i g_i(\boldsymbol{y}, s_i).$$

The intersection of the labor supply and labor demand curves gives the equilibrium of the system. Substituting the wages implied by the labor demand curve into the labor supply curve, we obtain a system of equations that determines the equilibrium labor allocation:

$$y_i = \frac{(p_i g_i(\boldsymbol{y},s_i))^\rho}{1+\sum_{i'}(p_{i'} g_{i'}(\boldsymbol{y},s_{i'}))^\rho} \equiv F_i(\boldsymbol{y},\boldsymbol{s}) \qquad \forall i \in I.$$

It is straightforward to verify that mild assumptions on g ensure that this system satisfies Assumption 1. Assuming two locations, A and B, and $g_i(\boldsymbol{y},s_i) = \exp\left\{\left(\frac{\delta_i}{\rho}\boldsymbol{y} + \frac{s_i}{\rho}\right)\right\} p_i^{-1}$, we obtain the following equilibrium system of equations:

$$y_i = \frac{\exp(\boldsymbol{\delta}_i \boldsymbol{y} + s_i)}{1+\sum_{i' \in \{A,B\}} \exp(\boldsymbol{\delta}_{i'} \boldsymbol{y} + s_{i'})} \equiv F_i(\boldsymbol{y},\boldsymbol{s};\delta) \qquad i \in \{A,B\}, \tag{2}$$

where $\boldsymbol{\delta}_i = (\delta_{ij})_{j \in \{A,B\}}$ is the vector of within-location ($i = j$) and cross-location ($i \neq j$) spillovers. Figure 1a illustrates the possibility of multiple equilibria for a fixed $\boldsymbol{\delta}$ and $\mathbf{s}$. It shows two interaction curves - y_A as a function of y_B and vice versa - and shows that given the curves intersect at multiple points.

For some illustrations, I will focus on the symmetric case where $s_i = s_{-i} \equiv s$ so that the equilibrium allocations are also symmetric. In the symmetric case, equilibria can be defined as the root of the equation $y = \exp(\delta y + s)(1 + 2 \cdot \exp(\delta y + s))^{-1}$ where $y = y_A = y_B$. This allows $y \in \phi(s)$ to be conveniently depicted in a 2D space. Figure 1b illustrates the set of equilibria for the symmetric case, while Figures 1c and 1d illustrate the set of equilibria for a general case.

2.1 Identifying the Equilibrium System F and Inverting Fundamentals s

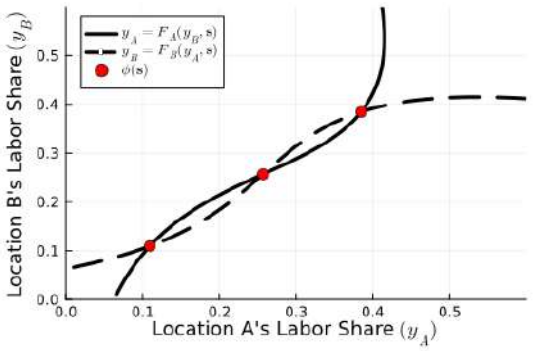
Consider the model represented by the system of Equations 1. An essential first step is the identification of the function F. In this section, I present identification arguments for F that rely on moment conditions alone.[14] Observed and unobserved fundamentals are denoted by variables $\mathbf{x}$ and $\boldsymbol{\xi}$ respectively, such that $\mathbf{s} = (\mathbf{x},\boldsymbol{\xi})$. Markets may vary on fundamentals $\mathbf{s}$ in flexible ways. With regional economies or markets as the unit of analysis, the researcher is assumed to observe the joint distribution of $(\mathbf{y},\mathbf{x})$.[15]

[14] The arguments from the literature on the estimation of games (De Paula and Tang, 2012) cannot be directly imported to the setting here. That literature studies discrete games with a finite set of players—entry decisions of large airlines, for example. In such contexts, a key challenge that multiplicity raises is the identification of the equilibrium choice probabilities y_i itself. Identification requires aggregation of entry decisions across markets but such aggregation runs into problems if markets are on different equilibria. This challenge is typically addressed by imposing strong assumptions on fundamentals and/or the equilibrium selection rule. The contexts of interest in this paper are those studied in the fields of macroeconomics, spatial/urban economics, international trade, etc. For example. While estimating a model of allocation of economic activity across space, it is more conventional to assume that actions of all firms within a narrow sector-location pair can be aggregated than it is to assume that different regions are identical on unobserved fundamentals.

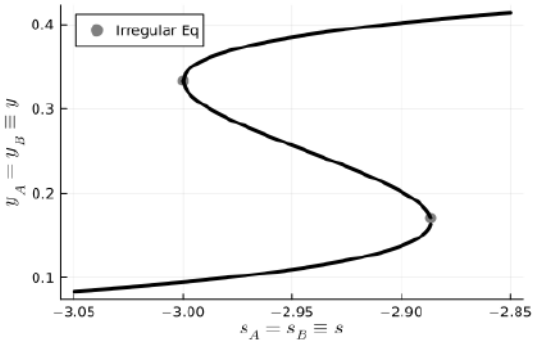
[15] In practice, the researcher observes M regional economies or markets where M is sufficiently large. In each market m, the researcher observes the equilibrium allocation $\mathbf{y}_m \equiv \{y_{1m}, y_{2m} \cdots y_{nm}\}$ and market-level characteristics $\mathbf{x}_m$. Fundamentals include unobserved characteristics $\boldsymbol{\xi}_m$ such that $\mathbf{s}_m = (\mathbf{x}_m,\boldsymbol{\xi}_m)$. The goal is to recover the structural equations F given the observed data $(\mathbf{y}_m,\mathbf{x}_m)_{m=1\cdots M}$.

Figure 1: System F and Equilibrium Set $\phi(\mathbf{s})$ for Example 1 'Spatial Allocation of Labor'

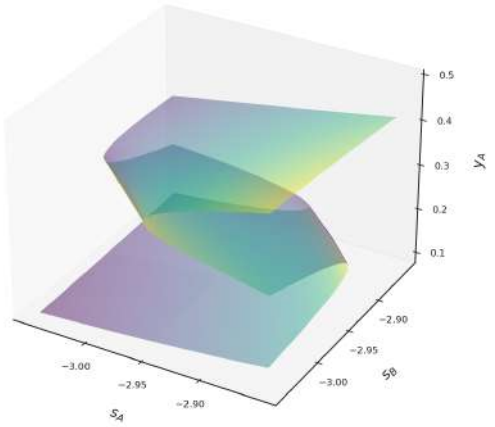
(a) Interaction Curves



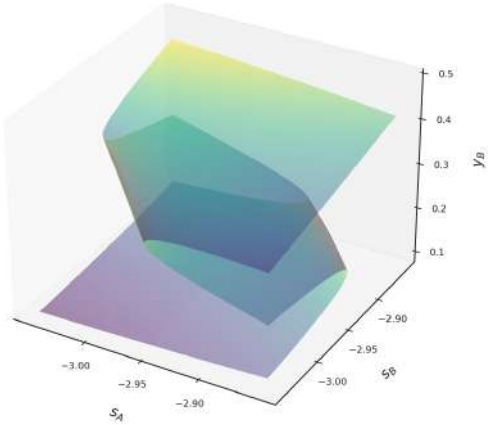
(b) Equilibrium Shares in Symmetric Case



(c) Equilibrium Share of Location A



(d) Equilibrium Share of Location B



Notes: The top left panel shows the interaction curves defined in Equation 2 for parameter values $s_A = s_B = -2.95$, where the red markers indicate equilibrium allocations. The top right panel depicts the equilibrium set $\phi(\mathbf{s})$ in the symmetric case $s_A = s_B$, with irregular equilibria highlighted by the gray points. The bottom two panels illustrate the equilibrium set $\phi(\mathbf{s})$ over a range of $\mathbf{s}$ values: the bottom left panel presents y_A on the z-axis, and the bottom right panel presents y_B, both showing the equilibrium surfaces as functions of s_A and s_B on the x and y axis respectively. In all panels, the spillover parameter δ_{ij} is set to $5 \times 1_{i=j} + 4 \times 1_{i \neq j}$. The equilibrium set $\phi(\mathbf{s})$ is computed using a combination of total degree homotopy continuation and coefficient parameter homotopy continuation, as described later in the section. The exponential function in the system F is approximated by a 17th-degree Maclaurin series expansion.

I establish identification under a conditional determinacy condition - that, even though, given a set of fundamentals $\mathbf{s}$, the system of equations F admits multiple solutions, each structural equation defines the equilibrium object y_i as a well-defined function of $(\mathbf{y}_{-i}, \mathbf{s})$. I further assume that the unobserved component of the structural equations, $\boldsymbol{\xi}$, is separable from $\mathbf{y}_{-i}$ in F.[16]

Proposition 1. *If*

[16] An advanced econometric literature discusses identification under non-separable errors. Similarly, there are methods of mixture model estimation that can employed when each structural equation admits multiple solutions of y_i given y_{-i}. These methods may be applicable to the arguments below, but doing so is beyond the scope of this paper.

1. ***Conditional Determinacy*** *and* ***Separability.*** F *is of the form*

$$y_i = F_i(\boldsymbol{y}_{-i}, \boldsymbol{x}_i) + \xi_i \quad \forall i = 1 \cdots n$$

where $\boldsymbol{x}_i$ *is the vector of observed fundamentals relevant for the outcome variable* y_i*.*

2. ***Exclusion****: For every* i*, there exists a known variable* $\boldsymbol{z}_i$ *such that* $E(\xi_i \,|\, \boldsymbol{z}_i = z) = 0$ *for all* z *in the support of* $\boldsymbol{z}_i$*.*

3. ***Completeness****:* $E[\, \zeta(y_{-i}, \boldsymbol{x}_i \,|\, \boldsymbol{z}_i = z)] = 0 \;\; \forall z \in Supp(\boldsymbol{z}_i) \Longrightarrow \zeta(y_{-i}, \boldsymbol{x}_i) = 0$

then F *is identified.*

Proof. We can restrict attention to bounded functions ζ since F is bounded. Since bounded functions have finite expectation, the result follows from the direct application of Newey and Powell (2003) Proposition 1. □

In the statement of the proposition, the expectation is taken across markets that vary in fundamentals $\mathbf{s}$. The first assumption states that even if the system of equations admits multiple solutions — i.e., $\phi(\mathbf{s})$ is a *correspondence* — each structural equation defines the equilibrium object y_i (e.g., the labor share of location i) as a well-defined *function* of $(\mathbf{y}_{-i}, \mathbf{s})$. In other words, while the entire vector $\mathbf{y}$ may exhibit indeterminacy due to complementarities among its components, this indeterminacy disappears once we fix the other $n-1$ components of $\mathbf{y}$.

Once it is established that the model at hand admits structural relationships of the form outlined in the first assumption, identification follows from the direct application of non-parametric instrumental variables (IV) arguments. The second and third assumptions outline the exclusion and completeness conditions that characterize identification under the IV approach. The exclusion restriction states that, for each structural equation i, the researcher needs a set of excluded variables, $\mathbf{z}_i$, mean independent of the error term ξ_i, that serve as the instruments in that equation. The third assumption establishes the first stage for these instruments, stating that the instruments complete $(\mathbf{y}_{-i}, \mathbf{x}_i)$. In other words, conditional on the included regressors $\mathbf{x}_i$, the instruments $\mathbf{z}_i$ induce sufficient variation in $\mathbf{y}_{-i}$ to identify F_i.[17] Once F is estimated, unobserved fundamentals can be recovered as structural residuals of the model. The whole vector of fundamentals $\mathbf{s}$ is recovered as $\mathbf{s} = (\mathbf{x}, \mathbf{y} -$

[17] In practice, one can assume that F is known up to some parameters θ, isolate the residuals as $\xi_i(\theta) = y_i - F_i(\mathbf{y}_{-i}, \mathbf{x}_i;\, \theta)$, contruct moments $g_i(\theta) = E(\xi_i(\theta)\; h(\mathbf{z}_i))$, and estimate θ by using generalised method of moments (GMM). One benefit of such a GMM-based approach is that it can be applied without assuming conditional determinacy. Separability alone allows the researcher to construct the moments and hence identify the model under appropriate exclusion and completeness conditions. The disadvantage of giving up conditional determinacy is that functional form restrictions might drive the identification.

$F(\mathbf{y},\mathbf{x})$).

Example continued. In the example economic framework, the conditional determinacy assumption requires that while scale economies may be present within a location, the definition of a location is narrow enough that conditioning on the rest of the economy eliminates the indeterminacy. If that is the case, then shifters that affect only certain locations can be used as instruments to estimate spillovers to other locations. The system of Equations 2 from the example can be rewritten as

$$\ln y_i - \ln(1 - y_A - y_B) = \boldsymbol{\delta}_i \mathbf{y} + \boldsymbol{\beta} \mathbf{x}_i + \xi_i \qquad \forall i \in A,B$$

where $\mathbf{x}_i$, say, would consist of the government subsidy support to location i. If the government support is pay-off relevant for i ($\beta \neq 0$), but is excluded from the pay-off equation of $-i$ ($E(\xi_{-i} x_i) = 0$), then x_i can be used as an instrument to estimate spillovers from y_i to y_{-i}. The key assumption here is that the government support to location i affects the labor demand in location $-i$ only through the productivity spillovers acting through y_i. □

2.2 Enumerating Equilibria $\phi(\mathbf{s})$ given F and $\mathbf{s}$

For empirical analysis of equilibrium selection, it is essential to compute all equilibria of the system across different parameter values or markets. Therefore, once the structural relationships summarized by F and the fundamentals $\mathbf{s}$ are recovered, the second step involves enumerating the entire set of equilibria $\phi(\mathbf{s})$. In the rest of this section, we do not need conditional determinacy or separability restrictions. All we need is that $\mathbf{s}$ is identified either directly from data or recovered through inversion of the model once F is identified.

Computing all equilibria of a system of non-linear equations is a difficult problem. Most conventional methods such as the fixed point iteration or Newton's method suffer from the problem that they are local in nature. Employing such a method to compute *all* the equilibria would require us to divide the whole domain Y into fine grid points so that they can be used as starting points in an iteration-based algorithm. This can be computationally expensive, especially if Y is high dimensional.[18] Moreover, such a gridding method does not help with equilibrium classification.

I address both equilibria enumeration and type assignment simultaneously by employing tools from algebraic geometry. There is a long-standing literature in mathematics that focuses on numerically feasible methods for finding all the roots of polynomial systems. Such methods, called homotopy continuation methods, although applicable to polynomial systems only, can

[18] Suppose we have I locations and the outcome corresponding to each location is gridded into P initial points. Then, a iteration-based algorithm would necessarily compute P^I solutions.

find solutions much faster than the conventional methods. While economic models typically feature non-polynomial F, if the set of all equilibria can be approximated arbitrarily well by the roots of a polynomial system, then homotopy continuation can be applied.

I first provide conditions under which such an approximation is possible. Then, I provide a continuation algorithm that uses the approximation to enumerate all equilibria. In the next subsection, I will introduce the notion of equilibrium types and show how a similar continuation algorithm can be used to assign types to the observed equilibria without additional computation time.

Under mild conditions on F specified in Assumption 1, the fixed points of F can be approximated arbitrarily well by a series of polynomial systems. I begin by reviewing the definitions of regular and isolated equilibria, which are essential for understanding the result on approximation.[19]

Definition 1. ***Regular and Isolated Equilibria.***

1. *For a given value of $\mathbf{s} \in S$, $\mathbf{y} \in Y$ is a regular equilibrium if $\mathbf{y} = F(\mathbf{y},\mathbf{s})$ and $I - F_y(\mathbf{y},\mathbf{s})$ is full rank.*

2. *Given $\mathbf{s}$, an equilibrium $\mathbf{y}$ is isolated if there exists a neighborhood around $\mathbf{y}$ such that for all $\mathbf{y}'$ in that neighborhood, $\mathbf{y}' \neq F(\mathbf{y}',\mathbf{s})$. That is, $\mathbf{y}$ is an isolated equilibrium if there exists an ϵ such that for all $\mathbf{y}' \in B_\epsilon(\mathbf{y})/\mathbf{y}$, $\mathbf{y}' \neq F(\mathbf{y}',\mathbf{s})$*

Irregular equilibria, intuitively, are those for which the interaction curves are tangent to each other, therefore arising only under knife-edge cases. Under the transversality condition that ∇F is full rank for all equilibria, the set of $\mathbf{s}$ with irregular equilibria is discrete and thus of measure zero.[20] This characterization is useful because the approximation property holds for the regular equilibria.

Denote the set $\bar{S} \subset S$ as the set such that for all $s \in \bar{S}$, all equilibria are regular.

Proposition 2. ***Polynomial Approximation.***

1. *All equilibria are regular for almost all $\mathbf{s} \in S$. That is, the set $S/\bar{S}$ is measure-zero. All regular equilibria are isolated.*

2. *The set of all equilibria is finite for all $\mathbf{s} \in \bar{S}$. They can be approximated arbitrarily well by the real roots of a polynomial system.*

Proof. See Appendix A.2.1 □

[19] See Definition 17.D.1 in Mas-Colell et al. (1995).

[20] This set will almost never be encountered in practice if the empirical distribution of $\mathbf{s}$ does not have atoms at the knife edge cases.

The proposition says that almost all equilibria are regular and, hence, isolated. Since the domain of outcomes is bounded, that means that there is a finite set of equilibria. The continuity properties of F and regularity properties of equilibria ensure that there exists a sequence of polynomial systems whose roots converge uniformly to the roots of the original system. This implies that we can focus on polynomial approximation of the system while enumerating equilibria.

Example continued. Recall that the parametrized system of equations from the example (Equations 2) is given by

$$y_i - \frac{\exp\{\boldsymbol{\delta}_i y_{-i} + \mathbf{s}_i\}}{1 + \sum_{i'} \exp\{\boldsymbol{\delta}_{i'} y_{-i'} + \mathbf{s}'_i\}} = 0 \qquad \forall i \in \{A, B\}.$$

Consider an alternate system of equations

$$y_i(1 + \sum_{i'} \widetilde{\exp}(\boldsymbol{\delta}_{i'} y_{-i'} + \mathbf{s}'_i, d)) - \widetilde{\exp}(\boldsymbol{\delta}_i y_{-i} + \mathbf{s}_i, d) = 0 \qquad \forall i \in \{A, B\}$$

where $\widetilde{\exp}(x, d)$ is the d-th degree Taylor approximation of $\exp(x)$ around zero. The roots of the approximate system can be made to lie arbitrarily close to the roots of the original system by choosing a large enough d. □

We can appeal to Proposition 2 to construct a polynomial approximation to the original system:

$$\mathbf{y} - F(\mathbf{y}, \mathbf{s}) \approx \mathbf{y} - \widetilde{F}(\mathbf{y}, \mathbf{s}) \equiv \mathcal{P}_{\mathbf{d}}(\mathbf{a}(\boldsymbol{s}), \mathbf{y}); \quad \mathbf{d} = (d_i)_{i=1 \cdots n}, \mathbf{a} = [a_{ij}]_{i=1 \cdots n, j=1 \cdots C_d^{n+d}}$$

where $\widetilde{F}$ denotes the polynomial approximation of F. $\mathbf{d}$ is the vector of degrees, each element corresponding to each equation in the polynomial. $\mathbf{a}$ is the matrix of coefficients, and $\mathcal{P}_{\mathbf{d}}(\mathbf{a}, \mathbf{y})$ denotes the system of polynomial in $\mathbf{y}$ of total degree $\Pi\mathbf{d}$. With the polynomial approximation, it is possible to extend the domain of $\mathbf{y}$ as well as the coefficients $\mathbf{a}$ to complex numbers, as is necessary because homotopy continuation methods work in the complex space. Given this, all the roots that satisfy $\mathbf{y} - \widetilde{F}(\mathbf{y}, \mathbf{s}) = 0$ can be computed using a total degree homotopy continuation algorithm.

Total degree homotopy continuation is a foundational approach in numerical algebraic geometry for solving systems of polynomial equations. By starting with a simple system whose solutions are known and continuously deforming it into the target system, we can enumerate all the possible equilibria. The process begins by constructing a simpler polynomial system (commonly referred to as the start system) whose solutions are explicitly known. The essential property of the start system is that it shares the same total degree as the target system. By continuously deforming the start system into the target system, one can trace the known solutions of the start system as they evolve along the solution paths. Starting with the same total degree

ensures that the number of solution paths corresponds to the maximum number of solutions to the target system (Bézout's theorem). The deformation is guided by a homotopy parameter and tracked numerically using methods such as differential equation solvers or predictor-corrector algorithms. The procedure is outlined in Algorithm 1.

Algorithm 1: Equilibrium Enumeration for Arbitrary Market

1: Determine the total degreeof the system as $D=\Pi_{i\in 1\cdots n}d_i$ and construct a start system of equations, denoted with $S(\mathbf{y})$, such that it has a) total degree D i.e the same total degree as $\widetilde{F}$, and b) known solutions. [a]

2: Pick a random scalar λ from the space of complex numbers.

3: Construct a homotopy that deforms the start system $S(\mathbf{y})$ to the target system $\mathbf{y}-\widetilde{F}(\mathbf{y},\mathbf{s})$:

$$H(\mathbf{y},\mathbf{s},h)=\lambda h S(\mathbf{y})+(1-h)[\mathbf{y}-\widetilde{F}(\mathbf{y},\mathbf{s})]=0$$

where h is the continuation parameter that varies from 0 to 1.

4: Solve the Ordinary Differential Equation (ODE) for $\mathbf{y}(h)$ that tracks the solutions of the homotopy from the start system ($h=0$) to the target system ($h=1$).

$$\frac{dH(\mathbf{y}(h),\mathbf{s},h)}{dh}=H_y\frac{d\mathbf{y}^\tau}{dh}+H_h=0;\quad \mathbf{y}^\tau(0)=\mathbf{y}_S^\tau$$

where τ indexes the solutions of the start system and hence, each solution path.

[a] A common choice for the start system is the set of monomial equations $S(\mathbf{y})=\{y_i^{d_i}=1\}_{i=1\cdots n}$. Since all the solutions of each equation in this system lie on the complex unit circle, their cartesian product can be used as the set of starting points for the continuation algorithm.

It is a well-known result in the field of algebraic geometry that this algorithm is guaranteed to find all the solutions of the target system, with each solution given by $y^\tau(1)$ if it exists. See Appendix A.1 for a formal statement of the result. The complex scalar λ ensures that the solution paths are well behaved from $h=0$ to $h=1$.

2.3 Assigning Types to $\mathbf{y}\in\phi(\mathbf{s})$

Once all the potential equilibria have been identified, the next task is to determine which equilibrium is actually selected in the data. For instance, suppose that in a given market, we find three possible equilibria: $(0.2,0.1)$, $(0.4,0.5)$, and $(0.8,0.1)$. If the empirically observed outcome is $(0.4,0.5)$, we can conclude that the second equilibrium is the one realized in this market. However, simply identifying the realized equilibrium is insufficient for empirical analysis without assigning meaningful labels to these equilibria.

Labeling equilibria allows for consistent comparison across markets with different funda-

mentals $\boldsymbol{s}$. For example, if an industrial zone is introduced and the observed equilibrium shifts from $(0.4,0.5)$ to $(0.6,0.6)$, the label assignment enables us to interpret this as a shift from one specific equilibrium type to another. This labeling step is essential for studying the impact of policy changes that may simultaneously alter both $\boldsymbol{s}$ and the selected equilibrium, thus facilitating analysis of equilibrium switching in response to policy events.

One such way of assigning labels is to use the notion of equilibrium types of Aguirregabiria and Mira (2019). According to this notion, two equilibria belong to the same type if they can be connected by a continuous path without encountering irregular equilibria along the way. This notion is appealing because it does not rely on the microfoundation of F which is necessary, for example, to Pareto-rank the equilibria. I further prove that the notion of equilibrium types is well-defined even if the number of equilibria changes as $\mathbf{s}$ changes. I also show that it admits counterfactual analysis since per this notion, for a given set of fundamentals only one equilibrium belongs to a given type.[21]

Denote the set of regular equilibria (Definition 1) for a given set of fundamentals $\mathbf{s}$ with the correspondence $\psi : S \to Y;\ \mathbf{s} \to \psi(\mathbf{s})$ where

$$\psi(\mathbf{s}) = \{\mathbf{y} \in Y;\ \mathbf{y} = F(\mathbf{y},\mathbf{s});\ \det\big(I - F_{\mathbf{y}}(\mathbf{y},\mathbf{s})\big) \neq 0\}.$$

Denote the graph of the correspondence by $gr(\psi)$ defined as $gr(\psi) = \{(\mathbf{s},\mathbf{y}) \in S \times Y;\ \mathbf{y} \in \psi(\mathbf{s})\}$.

For rest of the discussion, let's also assume that S is convex, so that we can consider locus of equilibria along the S-space.

Definition 2. ***Equilibrium Type.*** *Two equilibria $(\boldsymbol{s}',\boldsymbol{y}'), (\boldsymbol{s}'',\boldsymbol{y}'') \in gr(\psi)$ belong to the same equilibrium type if there exists a continuous path $\gamma : [0,1] \to Y;\ t \to \gamma(t)$ such that $\gamma(0) = \boldsymbol{y}', \gamma(1) = \boldsymbol{y}''$ and $\gamma(t) \in \psi(\boldsymbol{s}'(1-t) + \boldsymbol{s}''t)$.*

Intuitively, two equilibria in two different markets belong to the same type if we can continuously deform one market to the other without any abrupt changes in the equilibrium allocation. However, this definition alone doesn't guarantee that any equilibrium observed in the data can be uniquely assigned to a type, which is crucial for empirical analysis.

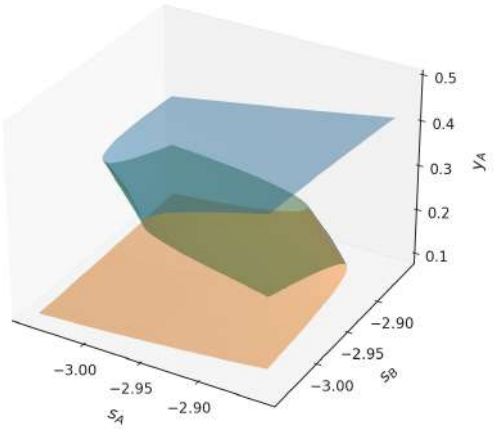
Example continued. Equilibria corresponding to different types are shown with different colors in Figure 2. The specific types shown correspond to cases where the total labor share of location A and B is 'low', 'medium', or 'high', all of which are possible within a particular range of the fundamentals (s_A, s_B). Outside of this range, only one of the three is possible. Figures 2a

[21] When possible, one could use Pareto rankings as an alternate way to assign types, as in Bajari et al. (2010a). Since Pareto rankings are typically not complete, the type assignment will often not be admissible to counterfactual analysis.

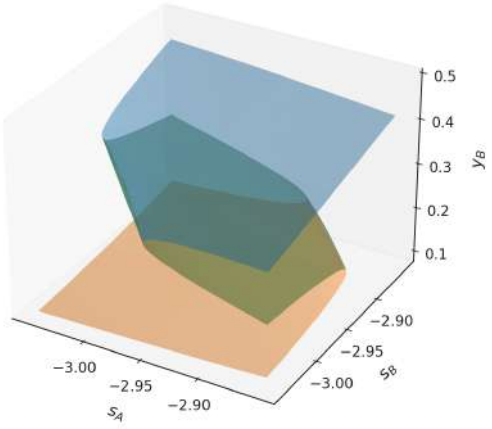
and 2b highlight the type assignment across the entire space of fundamentals while 2c takes a particular slice of this space (symmetric case) with gray markers highlighting the irregular equilibria separating the types. □

Figure 2: Equilibrium Types for Example 1 'Spatial Allocation of Labor'

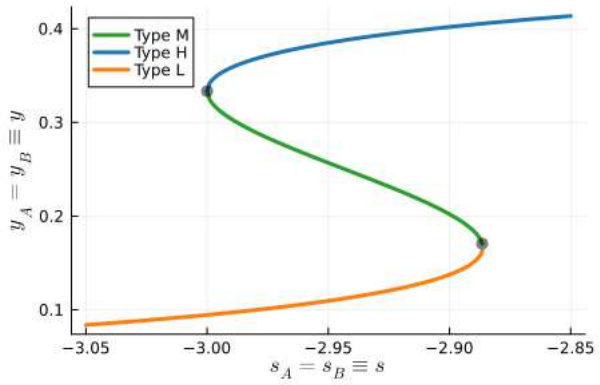
(a) Type Assignment: Location A



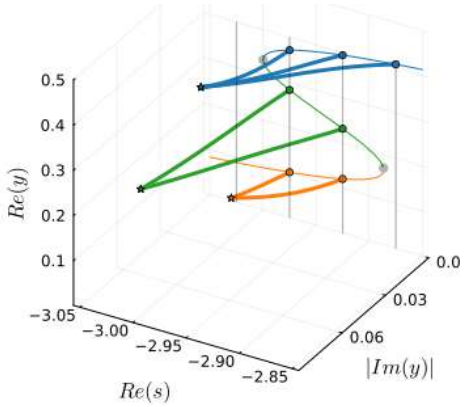
(b) Type Assignment: Location B



(c) Type Assignment in the Symmetric Case



(d) Type Assignment Algorithm



Notes: The top two panels show the equilibrium surfaces $\phi(\mathbf{s})$ for $\mathbf{s} \in [-3.05,-2.85]^2$, where different types are shown with different colors. The bottom left panel shows the equilibrium types in the symmetric case $s_A = s_B$. The bottom right panel shows the coefficient-parameter continuation paths in the complex space for the symmetric case. For the general case, the start point is chosen to be $\mathbf{s}_0 = (-3+0.1\iota, -3+0.11\iota)$. For the symmetric case, the start point is $s_0 = -3+0.1\iota$. In the bottom right panel, the x-axis denotes symmetric fundamentals s. The y and z-axes plot the absolute value of the imaginary part and the real part of the solutions, respectively. The continuation paths are illustrated for $s = (1-h)s_0 + hs^{\text{target}}$, where $h \in [0,1]$ and $s^{\text{target}} \in \{-3.025; -2.975; -2.925; -2.875\}$. For ease of exposition, only the real part of s is shown on the x-axis. The spillover parameter δ_{ij} is set to $5 \times 1_{i=j} + 4 \times 1_{i \neq j}$. The equilibrium set $\phi(\mathbf{s})$, as well as types, are computed using a combination of total degree homotopy continuation and coefficient parameter homotopy continuation. The exponential function in the system F is approximated by a 17th-degree Maclaurin series expansion.

The existing literature only defines equilibrium types as a relation between equilibria, not necessarily as a function. It is possible to *not* be able to define types unambiguously, if for

example an equilibrium is connected to two other equilibria of different types in $gr(\psi)$. Proposition 3 argues that such a case is not possible given the assumptions on F.[22] The first part establishes that the full set of equilibria can be grouped into a finite number of equivalence classes, and the second part shows that each equilibrium can be assigned a label corresponding to its equivalence class. The final part of the proposition shows that, given a set of fundamentals $\mathbf{s}$, only one equilibrium from the set $\phi(\mathbf{s})$ belongs to a given type. This property is essential for counterfactual exercises, such as evaluating how an equilibrium allocation changes when a policy shift leads to a certain equilibrium type.

Proposition 3. ***Equilibrium Type is a well-defined, invertible function.***

If for any $\boldsymbol{s} \in S/\bar{S}$; $\alpha \in R^m/\mathbf{0}$, and $y \in \Phi(\boldsymbol{s})/\Psi(\boldsymbol{s})$, the matrix

$$\begin{bmatrix} -F_{\mathbf{s}}(\boldsymbol{s},\mathbf{y})\alpha & I - F_{\mathbf{y}}(\boldsymbol{s},\mathbf{y}) \\ \frac{\partial \det(I - F_{\mathbf{y}}(\boldsymbol{s},\mathbf{y}))}{\partial \boldsymbol{s}}\alpha & \frac{\partial det(I - F_{\mathbf{y}}(\boldsymbol{s},\mathbf{y})}{\partial \mathbf{y}}) \end{bmatrix}$$

is full rank, then

1. *The 'Equilibrium Type' relation is an equivalence relation.*
2. *$gr(\psi)$ can be partitioned into a countable number of equivalence classes where each class corresponds to an 'Equilibrium Type'. That is, there exists a type assignment rule $T : gr(\psi) \to \mathbb{N}; (\boldsymbol{s},\boldsymbol{y}) \to T(\boldsymbol{s},\boldsymbol{y})$ such that $T(\boldsymbol{s}',\boldsymbol{y}') = T(\boldsymbol{s}'',\boldsymbol{y}'')$ iff $(\boldsymbol{s}',\boldsymbol{y}')$ and $(\boldsymbol{s}'',\boldsymbol{y}'')$ belong to the same equilibrium type.*
3. *The type assignment rule $T(\boldsymbol{s},\boldsymbol{y})$ is invertible in $\boldsymbol{y}$. That is, $T(\boldsymbol{s},\boldsymbol{y}) = T(\boldsymbol{s},\boldsymbol{y}')$ implies $\boldsymbol{y} = \boldsymbol{y}'$*

Proof. See Appendix A.2.2. □

The assumption states that for all irregular equilibria $\mathbf{s},\mathbf{y}$, each vector $[-F_{s_i}(\mathbf{s},\mathbf{y}); \frac{\partial \det(I - F_{\mathbf{y}}(\mathbf{s},\mathbf{y}))}{\partial s_i}]$, indexed with $i \in 1...M$, is outside the span of the n columns of the matrix $[I - F_{\mathbf{y}}(\mathbf{s},\mathbf{y}); \frac{\partial det(I - F_{\mathbf{y}}(\mathbf{s},\mathbf{y})}{\partial \mathbf{y}}]$. Under this assumption, an $m-1$ dimensional manifold of irregular equilibria partitions the entire set of equilibria (represented by an m-dimensional manifold) into equivalence classes. [23] This result not only establishes that types can be recovered in theory but also suggests a practical algorithm for doing so. Since the set of all equilibria $\{\phi(\mathbf{s}),\mathbf{s} \in S\}$ can be divided into a

[22] This might seem trivial at first. If equilibrium 1 is connected by a continuous path to equilibria 2 and 3, then equilibria 2 and 3 should also be connected through the path that passes via equilibrium 1. However, this reasoning is incorrect. For equilibria to belong to the same equilibrium type, they must be connected by a specific type of path—a straight line connecting the fundamentals. Although one could relax the definition to establish the equivalence result more easily, this would invalidate the invertibility arguments and the assignment algorithm, making them harder to establish.

[23] Note that this assumption is stronger that the transversality assumption in Assumption 1. The transversality assumption ensures that the the set of irregular equilibria is a manifold of dimension lower than m. This assumption implies that the manifold is $m-1$ dimensional.

countable number of equivalence classes each corresponding to a type, we can start by choosing an arbitrary set of fundamentals, $\mathbf{s}_0$, solving for all the equilibria in $\psi(\mathbf{s}_0)$, assigning them types arbitrarily. From there, we continuously deform *along the space of fundamentals* the market with fundamentals $\mathbf{s}_0$ and known equilibrium types into a target market with fundamentals $\mathbf{s}$ with unknown equilibrium types. By tracking the solution paths from $\psi(\mathbf{s}_0)$ to $\psi(\mathbf{s})$, we can assign types to the equilibria of the target market as the continuous paths that they they lie on. The steps are summarised in Algorithm 2. The enumeration part of the algorithm, referred to as coefficient-parameter homotopy continuation, is guaranteed to find all the equilibria of the target market (Morgan and Sommese, 1989). The type-assignment part of the algorithm is also guaranteed to correctly assign equilibrium types (Proposition 3).

Algorithm 2: Type Assignment for All Markets

1: Choose a random $\mathbf{s}_0$ from the space of complex vectors

2: Solve for all equilibria that satisfy $\mathbf{y} = \widetilde{F}(\mathbf{y}, \mathbf{s}_0)$ using the total degree homotopy continuation outlined in Algorithm 1.[a]

3: Construct a homotopy function that continuously deforms this market into a target market along the space of fundamentals

$$H(\mathbf{y},\mathbf{s},h) = F(\mathbf{y},(1-h)\mathbf{s}_0 + h\mathbf{s})$$

where $\mathbf{s}$ represents the fundamentals of the target market.

4: Track the solution paths from $\psi(\mathbf{s}_0)$ to $\psi(\mathbf{s})$ by solving for the ODE

$$\frac{dH(\mathbf{y}(h),\mathbf{s},h)}{dh} = H_y \frac{d\mathbf{y}^\tau}{dh} + F_h = 0; \quad \mathbf{y}^\tau(0) = \mathbf{y}_S^\tau.$$

where τ indexes the solutions of the start system and hence, each solution path.

5: Assign types to the solutions of the target market based on the paths they follow.

[a] Note that $\mathbf{y} - \widetilde{F}(\mathbf{y},\mathbf{s}) = \mathcal{P}_D(\mathbf{a}(\mathbf{s}),y)$ admits a complex vector $\mathbf{s}$ only if $\mathbf{a}(\mathbf{s})$ admits a complex vector $\mathbf{s}$ as an argument. If that is not the case, we can add a random complex noise to each element of $\mathbf{a}(\mathbf{s})$ where $\mathbf{s}$ is a randomly picked real vector of fundamentals.

Example continued. Figure 2d illustrates the continuation algorithm for assigning equilibrium types, focusing on a symmetric case for ease of visualization. Here, the x-axis represents the real part of the fundamentals, $\mathrm{Re}(s)$, while the y-axis and z-axis depict the imaginary and real parts of the equilibrium share, respectively. In this figure, three equilibria of a randomly chosen market within the complex fundamentals space are marked with stars. Each of these equilibria corresponds to a distinct equilibrium type for markets with real fundamentals. Starting from this initial market, each equilibrium is continuously tracked as the fundamentals are deformed

toward a target market. This deformation process is visualized through the orange, green, and blue curves, each representing the path of an equilibrium as it transforms. As the target market is approached, the imaginary component of each equilibrium diminishes, converging to a purely real value. The circles on each path indicate the equilibria in the target market, with colors matching the corresponding equilibrium types. □

2.4 Summary

Suppose we have data on equilibrium outcomes $\mathbf{y}_m$ and some observed fundamentals $\mathbf{x}_m$ in every market $m = 1 \cdots M$. In this section I have provided a three-step procedure for assigning types to factual and counterfactual equilibria $\phi(\mathbf{s}_m)$ in every market. Step 1 uses variation across markets and appropriate IVs to estimate F and recover the whole vector $\mathbf{s}_m$. Given an estimate of the function F and given that we have recovered fundamentals $\mathbf{s}_m$ from any observed equilibrium $\mathbf{y}_m \in \phi(\mathbf{s}_m)$, we can assign a type to the factual equilibrium $\mathbf{y}_m$ as well as the counterfactual equilibrium $\mathbf{y}_m \in \phi(\mathbf{s}_m)/\mathbf{y}_m$. Step 2 enumerates the full set of regular equilibria $\psi(\mathbf{s}_0)$ for a generic market $\mathbf{s}_0$ in the complex space using total degree homotopy continuation outlined in Algorithm 1. Step 3 enumerates the equilibria $\psi(\mathbf{s}_m)$ for all $m = .1 \cdots M$ and assigns the equilibria types using coefficient parameter homotopy continuation outlined in Algorithm 2. While Step 1 relies on the conditional determinacy, separability, exclusion and completeness conditions, Step 2 and 3 rely on the continuity and transversality properties of F alone.

In the remainder of the paper, I apply this methodology to analyze industrial zones in India. The unit of analysis is a region or market m in period t. Some regions are assigned industrial zones, while others are not. Equilibrium outcomes, denoted by $\mathbf{y}_{mt}$, represent the allocation of labor across sectors and spatial distribution within a region. Observable characteristics, $\mathbf{x}_{mt}$, include natural attributes such as proximity to major towns and highways, treatment variables indicating the presence of an industrial zone, and fixed effects. The primary source of multiplicity—and the key parameter of interest in Step 1—is a measure of aggregate scale economies within the industrial sector. In Step 2, I compute all possible labor allocations consistent with the fundamentals and the estimated scale economies in each region. Step 3 assigns types to both factual and counterfactual equilibria. These types are then regressed on the treatment variable in an event study framework to estimate the effect of industrial zones on equilibrium selection. The assigned types are further used to decompose changes in $\mathbf{y}_{mt}$, relative to the period immediately preceding the establishment of a zone, into two components: changes driven by fundamentals and those attributable to type-switching. Each component is subsequently regressed on the treatment variable to isolate the coordination effect.

3 Context and Three Empirical Facts

Industrial zones—also known as industrial parks, areas, or estates—are designated areas equipped with basic infrastructure to support industrial activities. The United Nations Industrial Development Organization (UNIDO) defines them as "a tract of land developed and subdivided into plots according to a comprehensive plan, with or without built-up factories, sometimes with common facilities for the use of a group of industries"(UNIDO, 1997). Their primary purpose is to provide manufacturing firms with a dedicated location to establish operations, benefiting from shared infrastructure and reduced setup costs.

The theory behind industrial zones as tools of industrial policy is based on three rationales. First, they address market failures in the provision of non-tradable public goods to industries. To support the manufacturing sector, governments must supply essential public goods, such as infrastructure, which require geographic proximity. Therefore, governments concentrate them in specific locations through the development of industrial zones. Second, such concentration promotes agglomeration economies, leveraging other Marshallian forces like labor market pooling and knowledge-sharing (Jordan and Saleman, 2014). Third, as highlighted in the big-push literature, increasing returns in the non-tradable public goods sector or the industrial sector can lead to coordination failures (Rodrik, 1996). Industrial zones may mitigate these failures, for example by providing implicit investment guarantees, encouraging firms to co-locate and invest.[24]

In India, industrial estates have been integral to the country's industrial policy since the inception of the First Five-Year Plans (Sarma, 1958). Initially envisioned to ensure balanced regional development and support small scale enterprises (Alexander, 1963; Roth, 1970), they have evolved into tools for attracting investment and promoting industrial clusters.[25] In Indian context, I use the term 'industrial zones' to encompass various place-based industrial policies, including industrial areas and industrial parks.[26]

Industrial zones in India present an ideal setting for studying the coordination effects of industrial policy. In an environment characterized by fragmented production among numerous small firms, industrial zones facilitate the coordination of firm actions to co-locate while also improving economic fundamentals by alleviating land market frictions and providing basic infrastructure. This dual role makes the relative importance of the coordination channel unclear ex-ante. Moreover, they are well-suited for applying the methodology developed in this paper. First, they provide a large number of localized policy experiments: according to a database

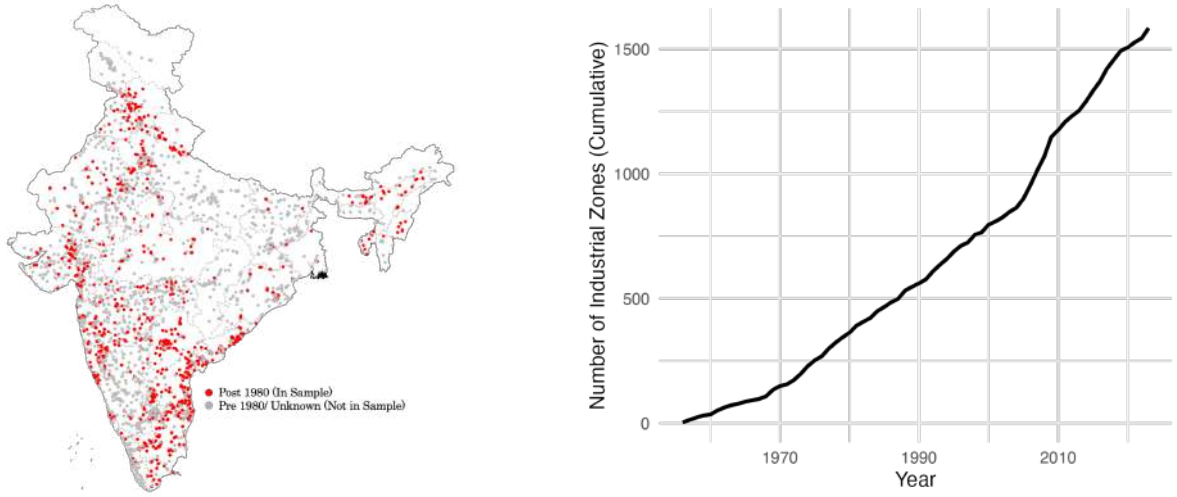
[24] This idea, while implicit in Rodriguez-Clare (2005) in the context of cluster-promoting micro interventions, was explicitly highlighted by several bureaucrats in interviews conducted for this project.

[25] Aggarwal (2011). Interview with Arbind Modi, Indian Revenue Services (December 11, 2024).

[26] Except for SEZs and EPZs, which aim to promote exports in environments with fewer regulatory barriers, the other industrial zones share the common goal of promoting cluster formation. SEZs and EPZs are excluded from this analysis due to their special incentive structures and objectives.

compiled by the Department for Promotion of Industry and Internal Trade (see Figure 3a), there are close to 4,000 industrial zones covering around 600,000 hectares in India. Second, they are promoted independently by State Industrial Development Corporations (SIDCs) and local development authorities, allowing for the exploitation of cross-sectional variation.[27]

Figure 3: Industrial Zones in India



(a) Cross-Section of industrial zones

(b) Number of industrial zones over time.

Notes: Panel (a) displays all the zones in the India Industrial Land Bank database. Zones established after 1980 are colored in red. The zones established before 1980 and the zones with unknown years of establishment are depicted in gray. Panel (b) shows the evolution of industrial zones, with known years of establishment, over time. Source: India Industrial Land Bank and primary data collected by the author.

The location decisions of industrial zones feature both economic and political considerations. Local governments explicitly emphasize factors such as proximity to water sources, major towns, ports, rivers, and the availability of contiguous parcels of land. However, these decisions are also influenced by broader objectives like regional equalization and balanced growth. Anecdotally, political considerations also play a role in determining the location of the zones. Given that the placement of industrial zones is unlikely to be random, even after accounting for observable factors, I employ a staggered difference-in-differences (DiD) approach that leverages variation in the timing of zone establishment. To implement the DiD design, I construct a new dataset of the establishment years of industrial zones. The data were gathered from various sources, including annual reports of State Industrial Development Corporations (SIDCs) and requests filed under the Right to Information Act of 2005 with various state departments, followed by

[27] SIDCs are special purpose vehicles registered under the Companies Act of 1956, established explicitly to develop industrial infrastructure in the form of Industrial Parks and Estates. They also provide microfinance to firms. (Telephone interview with Arindam Mishra, Joint Commissioner at Tax Policy Research Unit, Ministry of Finance, Government of India).

the digitization of hundreds of documents.[28] This data collection effort allows me to determine the establishment years for 1,585 out of 4,000 zones (shown in Figure 3a). There is substantial variation in the timing of establishment of industrial zones, as depicted in Figure 3b. I exclude from the sample all the municipalities that are in the 25km buffer areas of zones without a known establishment dates to avoid contamination of the control group.

Fact I: Industrial Zones Cause Large Increases in Non-Farm Activity

I begin by examining the reduced-form impact of industrial zones on industrial activity in municipalities that received them. This analysis serves two purposes. First, it provides a crucial moment for the first step of the three-step estimation procedure outlined in Section 2. Second, it lays the groundwork for the event study results in Section 4, which use a similar empirical strategy to study the effect of zones on equilibrium selection.

Data. For the baseline event studies, I use the Economic Census data for 1990, 1998, 2005, and 2013 to construct a panel of approximately 600,000 municipalities.[29] A municipality is treated in year YY if (i) the first zone in the municipality was established in year YY, and (ii) there were no other zones established before year YY within a 25 km radius. These conditions ensure that the effects of earlier zones and spillovers from nearby zones do not contaminate the results.[30]

To validate the findings from the Economic Census, I incorporate three additional datasets. First, I use the Population Census data for 1991, 2001, and 2011, which independently reports the number of non-farm workers.[31] Second, I use DMSP-OLS annual nighttime luminosity data, available for 1994–2013 at the municipality level.[32] Third, I use data on new company registrations from the Ministry of Corporate Affairs, available annually since 2001 at the level of 9,000 zip codes.[33]

Empirical Strategy. I estimate the municipality-level effects of the introduction of an industrial zone using a staggered DiD design. The first difference is between municipality-level outcomes in a given census period and outcomes in the period immediately before treatment. The second difference is between contemporaneous outcomes of treated municipalities and

[28] See Appendix B.1 for examples.

[29] I use the consistent municipality identifiers created by the SHRUG Data Development Lab. See Asher et al. (2021) for details.

[30] Municipalities treated in a year between two census rounds are defined to be treated in the latter census round. For instance, a municipality that receives a zone in 2001 is treated in the 2005 census round.

[31] The treatment period for a municipality when studying Population Census variables is defined similarly. For instance, a municipality treated in 1997 is defined to be treated in the 2001 census round.

[32] The original data is gridded at 1/120 degree, but I use the version aggregated to the municipality level by the SHRUG Data Development team.

[33] A zip code is treated in year YY if the first zone in that zip code was established in year YY. At this level, spatial contamination is less of a concern, so the second condition is not imposed—a zip code is treated in year YY even if a nearby zip code has a zone established earlier.

comparison municipalities more than 25 km from any industrial zone. Formally,

$$y_{it} = \alpha_i + \beta_{r(i)t} + \sum_{s \neq -1} \beta_{DiD,s} \mathbf{1}[t - TreatRound_i = s] Zone_i + \epsilon_{it} \tag{3}$$

where i represents the municipality, t represents the census round or period, α_i is the municipality fixed effect, and $\beta_{r(i)t}$ is the period fixed effect capturing common shocks across municipalities within a state r. $TreatRound_i$ is the census round in which municipality i, with $Zone_i = 1$, received its first zone. Identification relies on the assumption that comparison municipalities form a valid counterfactual after accounting for time-invariant differences and common shocks. The absence of differential pre-trends between treated and comparison municipalities is crucial for the validity of this assumption.

To improve comparison, I refine the control group using propensity score (P-Score) matching.[34] I estimate the propensity to be treated through a logit regression on baseline characteristics, including variable attributes such as proximity to roads. The details of the P-Score matching are provided in Appendix B.4. For each treated municipality, I select two control municipalities with the closest P-Score.[35] The covariate balance for both matched and unmatched samples is shown in Figure 21.[36] A 'P-Score group' is defined as the treated municipality and its two matched controls. The event study regressions based on P-Score matching also control for common shocks within a P-Score group.[37]

Baseline Results. Figure 4 plots the DiD estimates from equation 3. Estimates for the full unmatched sample (all municipalities at least 25 km away) are in black, while estimates for the matched sample are in gray. Across all outcomes, I observe no differential pre-trends between treated and comparison municipalities up to three census periods (31 years) before treatment. After treatment, treated municipalities experience significant increases in firm entry and non-farm jobs, leading to 125.5-141.8 more non-farm jobs and 37.48-40.95 more firms per square kilometer within 15 years (see Figures 4a and 4b).

As the Economic Census lacks establishment-level data on sales, inputs, or wages, I cannot directly assess the impact on productivity.[38] Nonetheless, the significant increase in the number of hired jobs— 150.9-162.7 jobs per square kilometer compared to 22.99-31.30 for non-hired jobs (see

[34] Narrowing down the control group also helps reduce computational demands in Section 2.

[35] Two nearest neighbors are selected with replacement. Since the pool is sufficiently large, the same candidate is rarely selected twice. If it happens, the candidate is dropped from the analysis.

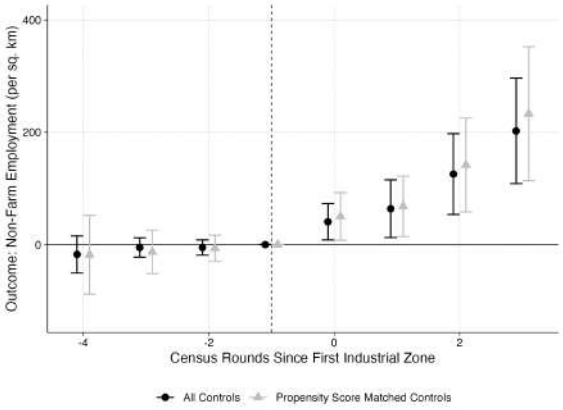
[36] In the unmatched sample, treated municipalities tend to be larger, closer to the coast and harbors, and more likely to have a river within the 25 km radius. Matching significantly improves the comparability of municipalities on these and other baseline characteristics.

[37] Robustness checks include matching only on baseline immutable characteristics, selecting one, three, or four control units instead of two, and using different distance restrictions of 10, 15, or 20 km. Additional checks include restricting controls to different subdistricts, districts, or states.

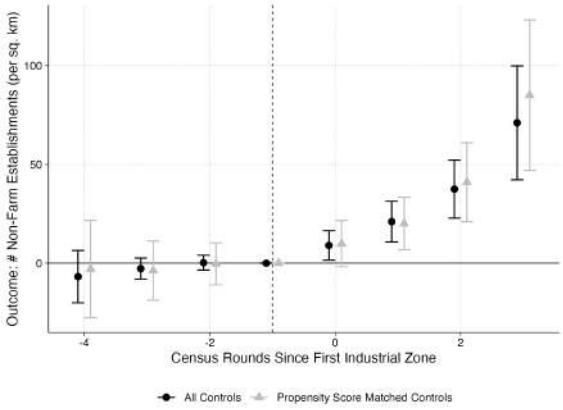
[38] The Annual Survey of Industries provides firm-level data, and Labor Force Surveys capture wage data, but these datasets are only available at the district level—a spatial unit too broad for this analysis. Most districts in India have industrial zones, leaving all but nine districts as pure controls.

Figure 4: Industrial Zones Increase Non-Farm Activity in Targeted Municipalities

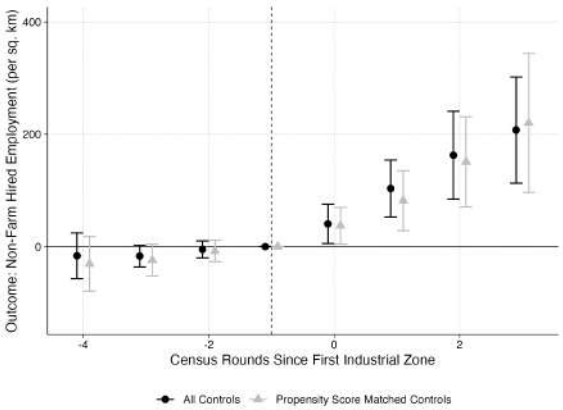
(a) Non-Farm Employment (Total)



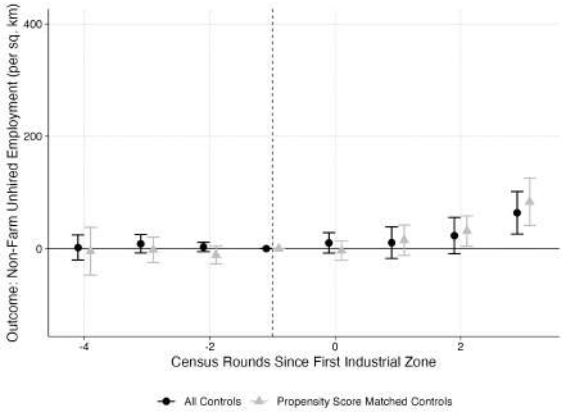
(b) Number of Non-Farm Establishments



(c) Non-Farm Employment (Hired)



(d) Non-Farm Employment (Unhired)



Notes: The figure depicts the event study coefficients from a difference-in-difference regression for different periods, each equivalent to 7-8 years. Outcomes are number of (a) workers, (b) establishments, (c) hired workers, and (d) non-hired workers (including family members, unpaid apprentices, and owners), restricted to the non-farm sector. Outcomes are normalized by the area of the municipality measured in square kilometers. All regressions include municipality and state×period fixed effects. Event studies using P-Score matching also include P-Score group×period fixed effects. Standard errors are clustered at the municipality level. Municipalities with zero non-farm employment in any period are excluded from the sample, but results are robust to their inclusion.

Figures 4c and 4d) suggests these are higher-paying, good jobs (Gindling and Newhouse, 2014).

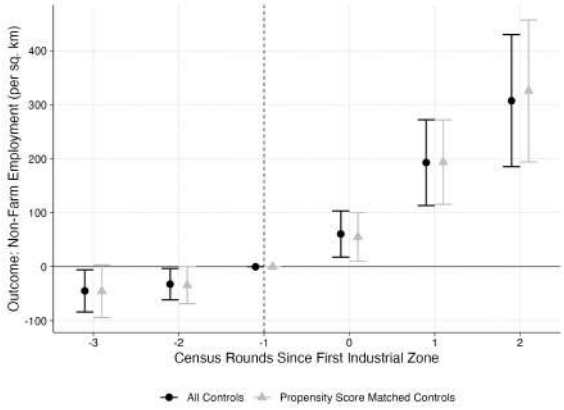
Robustness Checks. The strong and lasting effect of industrial zones holds across different comparison groups, specifications, and alternate measures of non-farm activity. Robustness to additional geographical restrictions is shown in Figure 22 (Appendix B.5.1). Restricting control units to be 10 km away, rather than 25 km, lowers the estimates, consistent with positive spillovers noted in Fact II. However, restricting controls to different subdistricts, districts, or

states has little impact, allaying SUTVA related concerns. The estimates also hold under different fixed effect specifications and adjustments for the 'negative-weights' concern in staggered DiD designs (Appendix B.5.2).

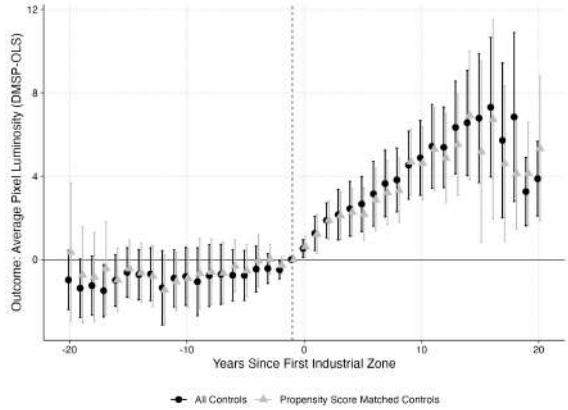
Alternative measures of economic activity show similar results. Non-farm employment, as measured by total number of workers excluding cultivators and agricultural laborers in the Population Census, rises significantly (Figure 5a). Appendix B.5.3 reports an increase in the total population, total employment as well as an average increase in the non-farm employment share, indicating reallocation across locations and sectors. Luminosity as measured by night lights also increases (Figure 5b).[39] It remains unclear if the shift from farm to non-farm employment directly improves welfare without data on productivity or wages. However, given that agriculture employs 54.6% of workers but contributes only 17% to GDP (MOSPI), this reallocation likely leads to higher-paying jobs, supporting structural transformation of the local economy.

Figure 5: Alternate Measures of Non-Farm Activity

(a) Non-Farm Employment, Population Census



(b) Night Light Luminosity



Notes: The figure depicts the event study coefficients from a difference-in-differences regression for different periods, with each period equivalent to ten years in the left panel and one year in the right panel. The outcomes are (a) total workers, excluding cultivators and agricultural laborers, per square kilometer, and (b) Night Time Luminosity (DMSP-OLS) averaged across the cells in the municipality. All regressions include municipality and state×period fixed effects. Event studies using P-Score matching also include P-Score group×period fixed effects. Standard errors are clustered at the municipality level.

Fact II: Industrial Zones Have Sectoral and Spatial Spillovers

In this subsection, I examine the spillover impacts of industrial zones on economic activity in non-targeted sectors and municipalities. The reduced form evidence presented in this section is sug-

[39] Interestingly, the share of females in the total non-farm employment falls, suggesting that women are less mobile across sectors and space.

gestive of the presence of scale and agglomeration economies within the non-farm sector. The results inform both the modeling assumptions and the identification strategy for Step 1 in Section 4.

Empirical Strategy. I expand the sample to include municipalities within 25 km of the treated and P-Score-matched control municipalities, henceforth referred to as treated and control catchment areas. By design, each industrial zone is the first zone in the corresponding treated catchment area. I estimate the spillover effects of introducing an industrial zone on nearby municipalities, comparing outcomes in 5 km distance bins (from 0 to 25 km). The DiD design compares outcomes within municipalities over time and between municipalities at varying distances from the treated and control municipalities. Formally,

$$y_{imt} = \alpha_i + \gamma_{g(m)b(i)t} + \sum_{s \neq -1} \beta_{DiD,b,s} \mathbf{1}[b(i) = b] \mathbf{1}[t - TreatRound_{g(m)} = s] Zone_m + X_{it} + \epsilon_{imt} \quad (4)$$

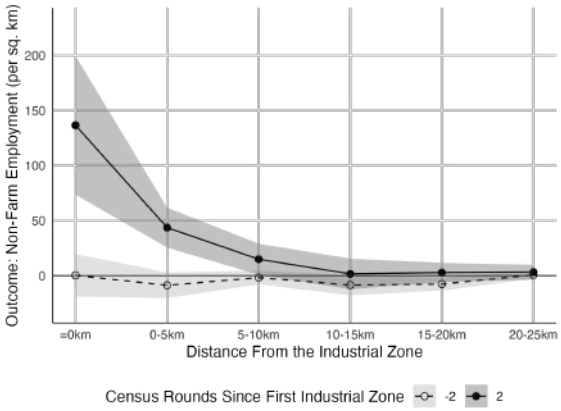
where i represents the municipality, b is the 5 km distance bin, and m (for market) is the catchment area. α_i is the municipality fixed effect, and $\gamma_{g(m)b(i)t}$ captures common shocks within each distance bin $b(i)$ in the P-Score group $g(m)$. X_{it} controls for the proximity to roads which changes over time as new roads are built. Details on the digitization of the historical roads network are provided in Appendix B.6. Identification of the spillover effects, $\beta_{DiD,bs}$, relies on the assumption that municipalities within the control catchment areas serve as valid counterfactuals for municipalities in the corresponding distance bin of the treated catchment area, after accounting for time-invariant differences and common shocks. As in the previous section, the absence of differential pre-trends between treated and control catchments is crucial for validating this approach.

Results. Figure 6 plots the DiD estimates $\beta_{DiD,b,s}$ from the model in equation 4 for $s = -2, 2$ with distance bins b depicted on the x axis. Reassuringly, I observe a lack of preexisting differential trends between treated and control municipalities two census rounds (15-16) years before the treatment. Two rounds after the treatment, the municipalities that do not get the zone but are close to the zone see a long-run increase in non-farm employment, suggesting spillovers of the zone beyond the targeted location. The construction of new roads cannot explain the spillovers, as proximity to roads is controlled for in the regression. Suppose the zone does not directly change the fundamental productivity of the nearby locations except through the construction of new roads. In that case, this result is suggestive of scale economies within the non-farm sector. The effects decrease in the distance from the zone, suggesting spatial decay of the scale economies.

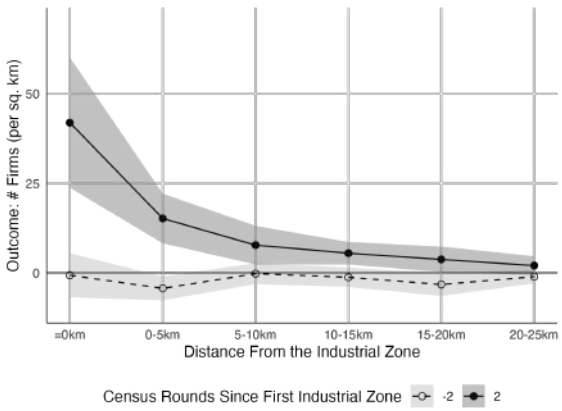
The results cannot be interpreted as indicative of scale economies if industrial zones provide direct support to the firms outside the zones. To allay that concern, I focus on firms in sectors that are not targeted by the zones. Even though all the zones in the sample target manufacturing, we see an increase in employment in the service sector in Figure 7, providing further evidence

Figure 6: Industrial Zones Increase Non-Farm Activity in Non-Targeted Municipalities

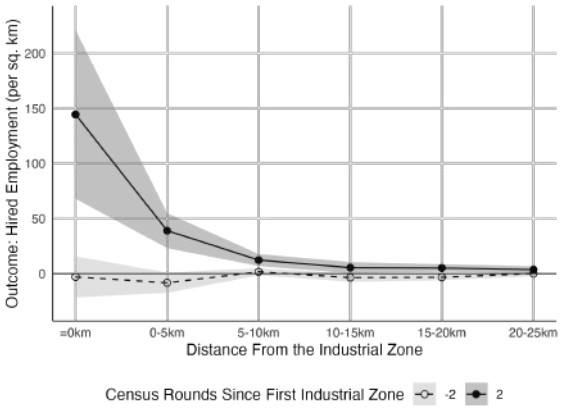
(a) Non-Farm Employment



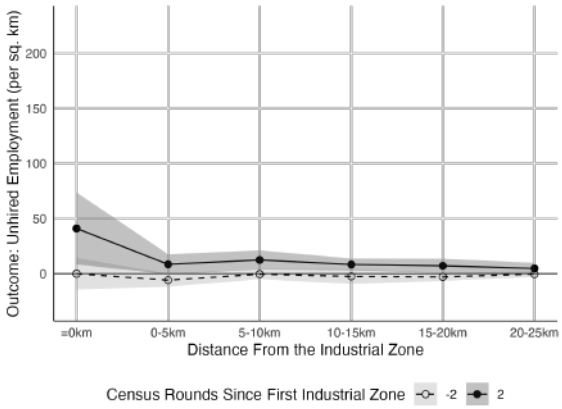
(b) Number of Non-Farm Establishments



(c) Non-Farm Employment (Hired)



(d) Non-Farm Employment (Unhired)



Notes: The figures show event study coefficients from a difference-in-difference regression for two periods before and after the treatment in untargeted municipalities, across 5 km bins around industrial zones. The comparison group is constructed using Propensity Score Matching. Each period spans 7 or 8 years. Outcomes are number of (a) workers, (b) establishments, (c) hired workers, and (d) non-hired workers (including family members, unpaid apprentices, and owners), restricted to the non-farm sector. Outcomes are normalized by the area of the municipality measured in square kilometers. The event studies include fixed effects for municipality, pscore-group $\times$ period, state $\times$ period, and highways within 25 km $\times$ period. Standard errors are clustered at the zone-catchment level (25 km buffer).

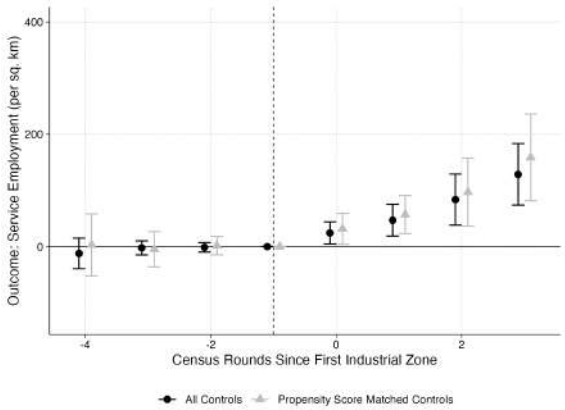
of spillovers within the overall non-farm sector.

Fact III: Distribution of Treatment Effects is Multi-modal

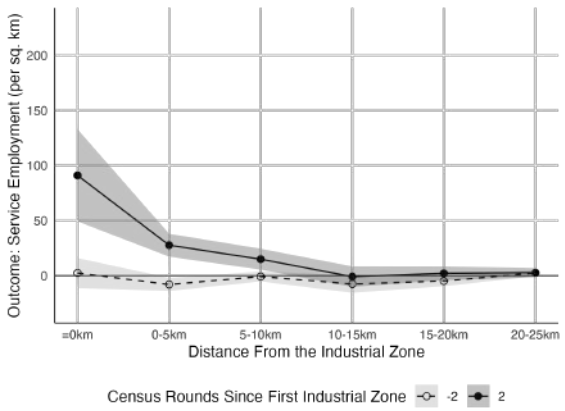
In this section, I provide additional suggestive evidence on the role of coordination in explaining the reduced-form impacts of industrial zones. I analyze the distribution of the treatment effect of industrial zones. The motivation behind this exercise is as follows: when zones affect the

Figure 7: Industrial Zones Increase Non-Farm Activity in Non-Targeted Sectors

(a) Service Sector Employment



(b) Service Sector Employment



Notes: The figures show event study coefficients from a difference-in-difference regression for two periods before and after the treatment in untargeted sectors, across 5 km bins around industrial zones. The comparison group is constructed using Propensity Score Matching. Each period spans 7 or 8 years. The outcome is the number of service sector workers as reported in the Economic Census. All outcomes are normalized by the area of the municipality. The event studies include fixed effects for municipality, pscore-group $\times$ period, and state $\times$ period. The right panel includes fixed effects for highways within 25 km $\times$ period. Standard errors are clustered at the municipality level in the left panel and zone-catchment level (25 km buffer) in the right panel.

probability of selecting a certain equilibrium in addition to influencing fundamentals, we expect the distribution of treatment effects to have multiple peaks.[40] One peak is associated with movement along fundamentals, and the other peaks are associated with markets where an equilibrium switch occurs.

Empirical Strategy. For each P-Score group g, I calculate the average outcomes in treated and control municipalities, both before and after the treatment. The treatment effect for P-Score group g is then defined as the difference in outcome changes between treated and control municipalities. Formally,

$$TE_g = \underbrace{(y_g^{treated,post} - y_g^{treated,pre})}_{\equiv \Delta_g^{treated}} - \underbrace{(y_g^{control,post} - y_g^{control,pre})}_{\equiv \Delta_g^{control}} \tag{5}$$

where $y_g^{treated,post}$ denotes the average non-farm employment or number of establishments in the treated municipality of group g *after* the treatment. $y_g^{control,post}$ is the average employment in the control municipalities of group g in the corresponding periods.[41] $y_g^{treated,pre}$ and $y_g^{control,pre}$ are

[40] This argument relies on a few assumptions. First, it assumes that both the distribution of underlying fundamentals and the distribution of the treatment effect of zones on fundamentals are single-peaked. Second, it assumes that the probability of switching equilibria upon receiving a zone is less than one.

[41] The periods after the treatment ($s > 0$) are pooled for power reasons, as not all villages have observations corresponding to every period s.

defined analogously. Note that $\mathbf{E}_g TE_g$ corresponds to the DiD estimate of the average treatment effect on the treated.

Results. Figures 8a and 8b display the distribution of changes in non-farm employment density and firm density in treated municipalities relative to the control municipalities.[42] The observed treatment effect distributions for both outcomes appear to be generated from a bimodal distribution or a mixture of multiple single-peaked distributions, suggesting equilibrium switching. However, since outcomes depend on fundamentals, the bimodality of outcomes could be explained by the bimodality of fundamentals. To test this, I plot the distribution of outcomes against the baseline fundamentals as summarized by the level of baseline outcomes. The results indicate that while industrial zones generally increase non-farm employment and firm entry, municipalities with similar baseline fundamentals may experience different outcomes upon receiving an industrial zone, warranting further investigation into the coordination channel.

4 Recovering Equilibrium Types

In Section 2, I outlined a three-step estimation procedure to recover equilibrium types in a general framework, illustrating each step with a simple example. Building on that foundation, this section analyzes the effects of industrial zones on employment outcomes within a specific framework where i) equilibrium employment outcomes across sectors and space emerge from the intersection of labor demand and supply curves, and ii) equilibrium multiplicity arises from scale economies in the non-farm sector. Consistent with previous notation, a catchment area, indexed by m, represents a single labor market comprising multiple municipalities, indexed by i, and multiple sectors, indexed by k. Census periods are indexed by t. The total number of municipalities in catchment area m at time t is denoted by I_{mt}, and the total number of sectors at any time is K.

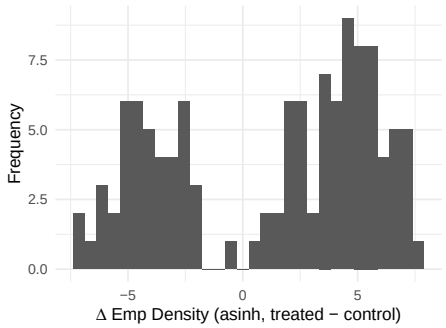
Labor Supply. Each market m is endowed with a fixed mass of labor L_{mt} in period t. Within a market, labor is supplied across locations and sectors with less than perfect elasticity. Workers in market m at time t choose a location $i \in I_{mt}$ and a sector $k \in K$ to work in. Let the pay-off received by worker $l \in L_{mt}$ in market m, location i, sector k, and time t be denoted by $\nu_{ikl,mt}$. This captures all observed and unobserved amenities associated with the location-sector pair. I assume:

$$\nu_{ikl,mt} = v_{ik,mt} + \epsilon_{ikl,mt}, \quad \nu_{i_0k_0l,mt} = 1 + \epsilon_{i_0k_0l,mt}$$

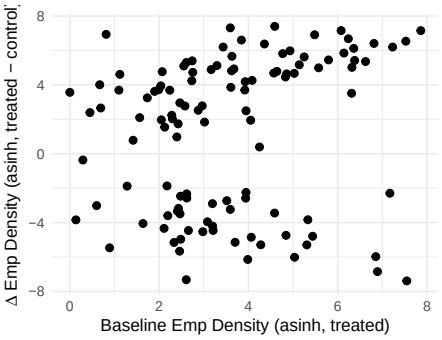
Here, $v_{ik,mt}$ represents the ex-ante average pay-off of working in location i and sector k, relative to an outside option indexed by i_0k_0. The term $\epsilon_{ikl,mt}$ captures an independent and identically

[42] The log/asinh scale helps visualize the distribution since the distributions in levels have thin tails. Note that the set of results in this section is for motivation only. These moments are not used in the structural estimation.

Figure 8: The Distribution of Treatment Effect of Industrial Estates

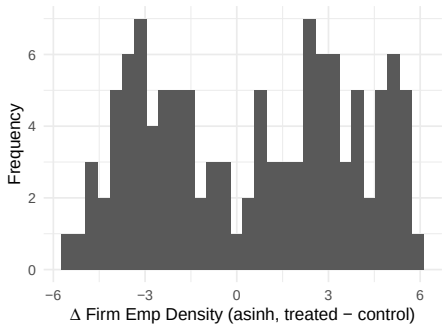


TE_g, Distribution

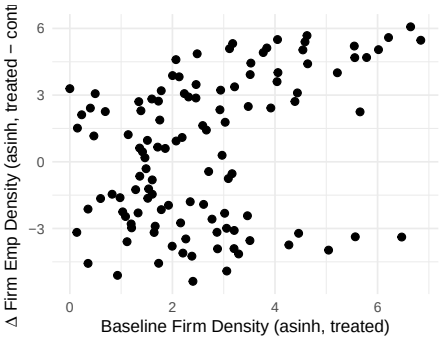


TE_g against $y_g^{treated,pre}$

(a) Non-Farm Employment



TE_g, Distribution



TE_g against $y_g^{treated,pre}$

(b) Number of Firms

Notes: The figures depict the unconditional distribution of TE_g as defined in the equation 5 on the left panel, and the scatter plot of TE_g against the baseline outcomes on the right panel. Each observation in the scatter plot corresponds to a propensity score group. The outcomes are number of non-farm workers and non-farm establishments, depicted in the top and the bottom panel respectively. The outcomes are normalized by the area of the municipality. The values for TE_g and $y_g^{treated,pre}$ are depicted on the Inverse Hyperbolic Sine Scale.

distributed (i.i.d.), ex-post, worker-specific Type-I Extreme Value (T1EV) preference shock. The distributional assumption on the ex-post heterogeneity allows us to derive an analytical expression for the ex-ante probability of a worker selecting a specific location-sector pair. Under symmetry in ex-ante pay-offs, this probability corresponds to the equilibrium share of workers

allocated to that location-sector pair:

$$\frac{L_{ik,mt}}{L_{mt}} = \frac{\nu_{ik,mt}^{\rho}}{1+\sum_{i'k'}\nu_{i'k',mt}^{\rho}} \tag{6}$$

In this expression, ρ governs both the variance of the T1EV shock and the elasticity of labor supply. Workers prefer a location-sector pair either because it offers, relative to the outside option, higher wages or better amenities. The relative pay-off of working in a non-farm sector can be decomposed into a wage component and other supply-side factors:

$$\ln\nu_{ik,mt} = \ln w_{ik,mt} + \ln\eta_{ik,mt} \tag{7}$$

Here, $w_{ik,mt}$ denotes the (relative) wage paid in sector k at location i, determined endogenously by labor supply and demand conditions. The term $\eta_{ik,mt}$ captures unobserved supply-side shifters, including the (relative) cost of living and other amenities.

Labor Demand. Each sector-location pair is populated by a representative firm that uses labor as its sole input, with internal returns to scale parametrized by α.[43] The representative firm takes wages as given, implying that the wage paid equals the marginal product of labor in that sector-location pair, scaled by a productivity term $A_{ik,mt}$:

$$w_{ik,mt} = A_{ik,mt}\left(\frac{L_{ik,mt}}{L_{i_0k_0,mt}}\right)^{\alpha} \tag{8}$$

Here, $A_{ik,mt}$ incorporates both exogenous and endogenous demand conditions, including nationally set prices and the labor share $1+\alpha$.[44] The outside sector—represented by the farm sector in the application—does not exhibit externalities, while non-farm sectors may experience aggregate increasing returns due to external economies of scale.[45]

The productivity term can be decomposed as follows:

$$\ln A_{ik,mt} = g_{ik}(\mathbf{L}_{mt}) + \ln a_{ik,mt} \tag{9}$$

In this expression, $\ln a_{ik,mt}$ represents an unobserved demand shifter that captures the exogenous productivity of a given location-sector pair. The function $g_{ik}(\mathbf{L}_{mt})$ captures externalities driven

[43] This simplification aligns with the classic models of Arthur (1986) and Krugman (1991). While not essential for identification, it simplifies the model and estimation for clarity. Extensions to multiple factors are unlikely to alter the results significantly. For instance, a model with multiple factors but proportional ownership across them is isomorphic to the current model.

[44] If production is given by $\tilde{A}_{ik,mt}L_{ik,mt}^{1+\alpha}$, and the sector faces a price p_k, the labor demand function can be derived by setting $A_{ik,mt} = \tilde{A}_{ik,mt}(1+\alpha)p_k$.

[45] If the farm sector (k_0) also exhibits scale economies, the estimated scale parameters for non-farm sectors (k) will be interpreted relative to the farm sector's scale parameter. Therefore, allowing scale economies in the farm sector does not affect the results, only the interpretation of the parameters.

by aggregate labor across the market. The first step in the estimation procedure focuses on identifying the spillover function g_{ik} along with other model parameters.

Equilibrium System. Equations (6)–(7) define labor supply as a function of relative wages $w_{ik,mt}$ and amenities $\nu_{ik,mt}$, where the latter subsumes unknown labor supply shifters. Equation (8) defines labor demand as a function of wages and underlying productivities. Labor demand $L_{ik,mt}$ decreases with wages $w_{ik,mt}$ if $\alpha < 0$. However, labor demand also responds endogenously to equilibrium labor shares through productivity spillovers.

Equations (8)–(9) together describe the aggregate labor demand function, which can exhibit an upward-sloping relationship due to these spillovers. The equilibrium vector of labor shares $\mathbf{L}_{mt} = \{L_{ik,mt}\}_{ik \in I_{mt} \times K}$ satisfies Equations (6)–(9) for a given vector of fundamentals $\{a_{ik,mt}, \eta_{ik,mt}\}$, labor supply and demand elasticities $\rho, \{\alpha_{ik}\}$, and the spillover functions $\{g_{ik}\}$.[46]

4.1 Step 1: Model Estimation and Inversion using Generalized Method of Moments

The equilibrium system of equations can be parsimoniously written as:

$$\ln L_{ik,mt} = \tilde{g}_{ik}(\mathbf{L}_{mt}) + \ln L_{i_0 k_0,mt} + s_{ik,mt} \tag{10}$$

where

$$\tilde{g}_{ik,mt}(\mathbf{L}) = \frac{\rho}{1-\rho\alpha} g_{ik,mt}(\mathbf{L}); \quad s_{ik,mt} = \frac{\rho}{1-\rho\alpha}(\ln a_{ik,mt} + \ln \eta_{ik,mt}).$$

The derivation of these equations is provided in Appendix C.1. Appendix C.2 shows that the system of equations satisfy Assumption 1.

The function $\tilde{g}_{ik,mt}$ captures the dependence of $L_{ik,mt}$ on the equilibrium vector of labor shares through supply, demand, and spillover channels. The term $s_{ik,mt}$ reflects the fundamental attractiveness of a location, arising from both productivity and amenity advantages. It is important to note that counterfactual predictions for equilibrium outcomes do not rely on disentangling the demand and supply channels.

The functions $\{\tilde{g}_{ik}(.)\}$ (and consequently, the fundamentals $\{s_{ik,mt}\}$) are non-parametrically identified using instruments under appropriate completeness and exclusion restrictions (Proposition 1). Intuitively, shifters for $s_{i'k',mt}$ can serve as valid instruments for estimating g_{ik} if these shifters: i) induce sufficient variation in $L_{i'k',mt}$ (completeness), and ii) are orthogonal to $s_{ik,mt}$ (exclusion). However, I proceed by imposing parametric assumptions on $\tilde{g}_{ik}$ to enable

[46] I do not need to close the goods side of the model since prices in both farm and non-farm sectors are assumed to be set nationally. For the same reason, I do not need to specify how firm profits, if any, are distributed among agents in the economy. While this assumption ensures exogenous good prices, endogenous prices are straightforward to incorporate.

estimation with finite data. Specifically, I assume:

$$\tilde{g}_{ik}(\mathbf{L}_{mt}) = \delta \sum_{k \neq k_0, i' \in I_m} \exp(-\eta d_{ii'}) \frac{L_{ik,mt}}{L_{mt}} \qquad \forall k \in K/k_0 \tag{11}$$

Here, $d_{ii'}$ represents the distance between locations i and i', while η captures the spatial decay of spillovers. When $d_{ii'} > 0$, $\exp(-\eta d_{ii'}) < 1$, implying that an increase in non-farm employment in a nearby municipality has a smaller impact on productivity compared to a local increase. The term $\delta \cdot 100 \cdot L_{mt}^{-1}$ represents the percentage increase in productivity resulting from the addition of a non-farm worker locally. Normalizing the scale elasticity by the total size of the labor market L_{mt} ensures that the productivity impact of an additional worker is greater in smaller labor markets.[47]

Empirical Strategy. To estimate Equations (10)-(11), we require shifters for the allocation of labor $\mathbf{L}_{mt}$ that are orthogonal to $s_{ik,mt}$, conditional on the size of the farm sector in the reference location $L_{i_0 k_0, mt}$. I use industrial zones that target sectors $k' \neq k$ and locations $i' \neq i$ as instruments for this estimation. These industrial zones must induce independent variation in the vector $\{L_{ik,mt}\}_{ik}$, a first stage suggested by Fact I (for $i = i'$ and $k = k'$) and Fact II (for $i \neq i'$ and $k \neq k'$), while being orthogonal to the unobserved fundamentals of non-targeted sectors and locations—a condition validated by the absence of pre-trends in Fact II. In other words, conditional on controls, we require that (a) the sectoral and locational targeting of industrial zones is orthogonal to trends in non-targeted sectors and locations, and (b) industrial zones do not directly affect the fundamentals of those locations.

Motivated by the event studies presented in Section 3, I saturate Equation 10 with a rich set of fixed effects and focus on the matched sample. Formally, I estimate the following equation:

$$\ln L_{ik,mt} = \delta \underbrace{\sum_{\tilde{k} \neq k_0, i' \in I_m} \exp\{(-\eta d_{ii'})\} \frac{L_{i\tilde{k},mt}}{L_{i,mt}}}_{\equiv y_{ik,mt}(\eta)} + x_{ik,m} + x_{b(i)k,g(m)t} + x_{k,mt} + X_{ik,mt} + \xi_{ik,mt}$$

where k represents the untargeted service sector. As before, $g(m)$ refers to the P-Score group corresponding to market m, and $b(i)$ represents the 5km distance bin where municipality i is located. The term $x_{ik,m}$ captures municipality-specific components, $x_{k,mt}$ accounts for market-specific common shocks, and $x_{b(i)k,g(m)t}$ represents common shocks affecting municipalities

[47] I make two simplifying assumptions regarding the spillover structure. First, I assume that the endogenous component of sectoral productivity depends solely on total non-farm employment in an area. Although productivity in any non-farm subsector k could theoretically depend on the size and spatial distribution of the non-farm sector, I abstract from dependencies on sectoral composition within the non-farm sector. Second, I assume that, within a location, endogenous spillovers are uniform across non-farm subsectors, that is, $g_{ik} = g_{ik'}$ $\forall k \neq k' \neq k_0$. For example, an increase in the labor share of the non-farm sector benefits industries like steel and oil equally. These assumptions streamline notation, simplify estimation, and make equilibrium computations feasible. Importantly, these assumptions are not critical for the core identification arguments presented in the paper.

within a distance bin and a P-Score group. Additionally, $X_{ik,mt}$ controls for infrastructure improvements through which industrial zones might directly impact non-targeted sectors in non-targeted locations. The endogenous variable $\ln L_{i_0 k_0,mt}$ is absorbed into the market-time fixed effect $x_{k,mt}$. Intuitively, by controlling for market-specific shocks—effectively adding a third difference that compares locations within a market—I isolate the agglomeration forces captured by the spillover function $\tilde{g}_{ik}$ from the dispersion forces driven by the fixed labor market size L_{mt}.

Let the total number of fixed effects to be estimated be F, and let the dummy variable corresponding to each fixed effect be denoted by D^f.[48] In addition to the F moments, $\{E(D^f_{ik,mt}\xi_{ik,mt})=0\}_{f\in F}$, we require at least two additional moments to estimate δ and η. Intuitively, the local effects of industrial zones on the non-targeted sector identify δ, while spatial spillovers on nearby non-targeted locations discipline the spatial decay parameter η.Define the instruments as:

$$z^{bs}_{ik,mt} = \mathbf{1}[b(i)=b]\mathbf{1}[t-TreatRound_{g(m)}=s]Zone_m$$

Here, $z^{bs}_{ik,mt}$ is an indicator variable that equals one if municipality i received an industrial zone, located b distance away, and s census periods ago. These instruments correspond directly to the key regressors used in the event study regressions presented in Section 3, Fact II. The parameters δ and η are identified using the following moment conditions:

$$E(z^{bs}_{ik,mt}\xi_{ik,mt})=0, \qquad b\in\{\text{0–5km, 5–10km, 10–15km}\}, \quad s=2$$

Using the 15–20km bin as the reference bin and $s=-1$ as the reference period, I estimate the spillover function based on the effect of industrial zones two census rounds (or 15 years) after treatment. This ensures consistency with the static model, where the spillover function captures the long-run effects of labor allocation on productivity. This setup provides three instruments—each corresponding to one of the three distance bins—for estimating two parameters (δ, η), making the model over-identified. To account for the relative precision of the moment conditions, I employ the standard two-step GMM procedure. In the first step, all moments are weighted equally. In the second step, moments are weighted by the inverse of the variance-covariance matrix estimated from the first step. The reduced-form version of the equation—where the endogenous term in Equation 10 is replaced by the instruments—aligns directly with the spillover specification in Equation 4 from the empirical facts discussed in Section 3.[49] Consequently, the validity of the exclusion restrictions is supported by the absence of preexisting differential trends, as illustrated in Figure 7, right panel.

[48] There are $F=1+(\sum_m I_m-1)+(M\times T-1)+(B\times G\times T-1)$ fixed effects to be estimated, corresponding to the intercept, location fixed effects, market-specific time fixed effects, and bin-group-specific time fixed effects.

[49] The only difference is the inclusion of three-way fixed effects instead of two-way fixed effects. The third comparison contrasts distance bins close to the industrial zone with those further away. In practice, adding this third comparison does not materially affect the results, as shown in Fact II, where spillovers to peripheral bins are statistically indistinguishable from zero.

Since there are a large number of coefficients to be estimated—F fixed effects, along with (δ, η)—I follow a sequential approach. First, I guess a value of η so that the structural equation becomes linear in the regressor $y_{ik,mt}(\eta)$ and the fixed effects. Second, I residualize the fixed effects from the outcome variable $\ln L_{ik,mt}$, the regressor $y_{ik,mt}(\eta)$, and the instruments $\{z^{bs}_{ik,mt}\}$. Third, given a guess of η, the parameter $\hat{\delta}(\eta)$ is estimated from a linear IV regression of the residualized outcome $\ln L^{RE}_{ik,mt}$ on the residualized regressor $y^{RE}_{ik,mt}(\eta)$ using the residualized instruments $z^{bs,RE}_{ik,mt}$. This approach allows me to construct the moments

$$g^{bs}_{ik,mt}(\eta) = \left(\ln L^{RE}_{ik,mt} - \hat{\delta}(\eta) y^{RE}_{ik,mt}(\eta)\right) z^{bs,RE}_{ik,mt}$$

and the corresponding objective function

$$g(\eta)' W g(\eta),$$

where $g(\eta)$ is a three-dimensional vector of moments defined as

$$g(\eta) = [g^{b,s=2}_{ik,mt}(\eta)]_{b \in \{0-5km, 5-10km, 10-15km\}},$$

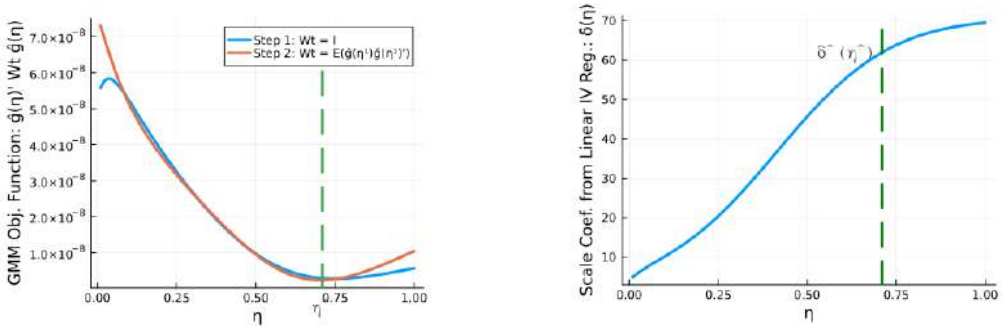
and W is a matrix that weights the moments. Minimizing this objective function provides an estimate of η. Figure 9a presents the empirical relationship between the objective function and η. Once η is estimated, I estimate δ using a linear IV regression. A higher estimate of η requires a correspondingly higher estimate of δ (Figure 9b) to rationalize the observed effects. Intuitively, the scale economies must be sufficiently strong to explain the observed effects in the catchment area of the zone despite sharp spatial decay.

Results. I estimate $\hat{\delta}, \hat{\eta} = 61.94, 0.71$, significant at the 95% confidence interval. The estimates are not sensitive to the choice of the weight matrix, as shown in Figures 9a and 9b. Given that the average market size is 150,512 workers, an addition of 100 non-farm workers in a municipality increases non-farm productivity by 3.14 percent ($61.94 \times 100 \times 100/150{,}512$) locally, on average. For the median market, the same increase in non-farm employment raises productivity by 4.12 percent. I also find sharp spatial decay in spillovers: an additional worker in a municipality one kilometer away contributes only half as much to productivity as an additional worker locally.[50]

Once δ, η and hence $\tilde{g}_{ik}(\mathbf{L}_{mt})$ is identified, I recover $s_{ik,mt}$ as the residuals of the model:

$$s_{ik,mt} = \ln L_{ik,mt} - \ln L_{i_0 k_0, mt} - \tilde{g}_{ik}(\mathbf{L}_{mt}) \tag{12}$$

[50] It would be incorrect to conclude that a reduction in employment effects by a factor of half in villages 5 km away implies η is approximately equal to $-\frac{\ln 0.5}{5} = 0.13$. This reasoning overlooks the higher-order effects of increased employment. Productivity spillovers are amplified by higher-order linkages among firms across space. To rationalize the same observed spatial decay in employment increase, a much sharper degree of spatial decay in productivity is required.



(a) $\hat{\eta}=0.71(0.0004)$ (b) $\hat{\delta}=61.94(31.29)$

Figure 9: $(\hat{\delta},\hat{\eta})$ computed with two-step GMM

4.2 Step 2 and 3: Equilibrium Enumeration and Type Assignment

We can write the equilibrium system for market m in time period t succinctly as a system of $I_m \times K$ equations $\mathbf{L}_{mt}=F(\mathbf{L}_{mt},\mathbf{s}_{mt})$ where[51]

$$L_{ik,mt}=F_{ik}(\mathbf{L}_{mt},\mathbf{s}_{mt})=\begin{cases}\exp(\tilde{g}_{ik}(\mathbf{L}_{mt})s_{ik,mt})L_{i_0k_0,mt} & \text{for } ik\neq i_0k_0\\ L_{mt}-\sum_{ik\in I_m\times K}L_{ik,mt} & \text{for } ik=i_0k_0.\end{cases}$$

Even though I assumed $\tilde{g}_{ik}(\mathbf{L})$ is linear in $\mathbf{L}$, the equilibrium system remains non-linear due to the presence of the exponential term. Solving for all equilibria in a non-linear system is challenging, but the same problem for polynomial systems is well-studied in the field of numerical Algebraic Geometry. In Proposition 2, I showed that the roots of the original system F can be brought arbitrarily close to those of its polynomial approximation. Therefore, I approximate each structural equation in the equilibrium system F with a 15th-degree polynomial denoted by $\tilde{F}$.[52] Despite these simplifications, the total degree of $\tilde{F}$ remains greater than $16^{I_m\times(K-1)}$. Given that computational time and memory requirements scale linearly with the total degree of the system—which itself grows exponentially with the number of locations and sectors—I group municipalities into broader locations within each catchment.[53] Motivated by the empirical findings from Section 3, which suggest that spillovers do not extend beyond 10km, I partition

[51] Appendix C.3 shows that the system of equations satisfies the hypothesis of Proposition 3 generically, for two location case.

[52] I replace the exponential term with its 15th-degree Taylor expansion around 0. Define the n-th degree approximation of $\exp(x)$ by $\tilde{\exp}(x,n)$. Formally, I define

$$\tilde{F}_{ik}(\mathbf{L},\mathbf{s})=\tilde{\exp}(g_{ik}(\mathbf{L}_{mt}),15)\exp\{s_{ik,mt}\}L_{i_0k_0,mt}$$

for $ik\neq i_0,k_0$ and $\tilde{F}_{i_0,k_0}(\mathbf{y},\mathbf{s})=F_{i_0,k_0}(\mathbf{L},\mathbf{s})$.

[53] Formally, we have $\ln\frac{\sum_{k\neq k_0}L_{ik,mt}}{L_{i_0k_0,mt}}=\delta\sum_{i'}\exp(-\eta d_{ii'})\frac{\sum_{k\neq k_0}L_{i'k,mt}}{L_{mt}}+\tilde{s}_{i,mt}$, where $\tilde{s}_{i,mt}=\ln\sum_{k\neq k_0}\exp\{s_{ik,mt}\}$.

each catchment area into five equal-area inner locations and one large peripheral location, as depicted in Figure 10. The non-farm sector in the five inner locations exhibits scale economies, while the inclusion of the hinterland ensures that labor supply remains elastic to wages in the inner locations. I restrict the sample to catchments where there is at least one municipality in one of the six sub-regions. This leaves a total of 627 regions, of which 230 contain an industrial zone and are classified as treated. Among these, 174 treated regions and 335 control regions are observed across all Census rounds.

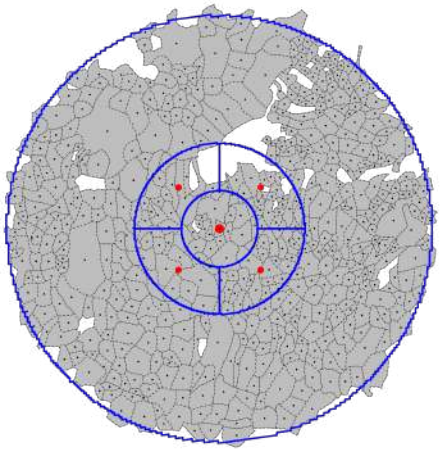


Figure 10: Set of Location-Groups for Equilibrium Enumeration

Notes: The figure shows the definition of the action space and the labor market for an example zone. At the center is the set of villages less than 4.472 km from the park. The 4.472km to 10km concentric ring is split into four parts of equal sizes. These five equally sized regions form the part of the catchment area within which the spillovers operate. 10-25km ring denotes the hinterland.

For the remainder of this section, I set $I_m = 6$ and $K = 2$. Using the algorithm outlined in Section 2, I solve for all the roots of the equilibrium system across $T \times M$ markets. The algorithm proceeds in two steps. First, I use total degree homotopy continuation (Algorithm 1) to identify the equilibria of a generic market. Second, I apply parametric continuation (Algorithm 2) to efficiently compute the equilibria for all $T \times M$ markets and assign equilibrium types accordingly.

Enumerating Equilibria of an Arbitrary Market

I apply Algorithm 1 to identify all equilibria of a generic market, where the market fundamentals $\mathbf{s}_{ik}$ are randomly selected from the $I \times K$ dimensional complex space. In this step, I solve the polynomial system $\tilde{F}(\mathbf{L}, \tilde{\mathbf{s}}) = 0$, where $\mathbf{L}$ and $\tilde{\mathbf{s}}$ are $I \times K$ vectors representing labor shares and randomly chosen fundamentals, respectively. The choice of the start system is critical for successful continuation. A start system with a lower total degree may lead to path bifurcations before reach-

ing the target system, resulting in failure. Following standard practice, I select the start system as:

$$S(\mathbf{L}) = \left[L_{ik}^{\text{Total Degree of } \tilde{F}_{ik}} - 1 \right].$$

Out of the 16 million paths tracked, approximately 15 million were extraneous, and a small fraction (around 0.02%) encountered points where H_y was ill-conditioned.

Enumerating Equilibria of all Markets and Assigning Equilibrium Types

We could use Total Degree Homotopy Continuation to solve for the equilibria of all markets $mt \in M \times T$ with fundamentals $\mathbf{s}_{mt}$, not just a generic market with random fundamentals $\tilde{\mathbf{s}}$. With approximately 627 markets across 4 time periods, this computation would take close to 50 days. However, tracking all paths for each market is unnecessary. As demonstrated earlier, around 15 million out of 16 million paths were extraneous, diverging to infinity. This inefficiency is a well-known limitation of total degree homotopy continuation. While the method ensures that all solutions are identified, it typically requires tracking far more paths than necessary, many of which do not contribute to the final set of real solutions.

To improve the efficiency of the algorithm and reduce the number of unnecessary paths tracked, I employ coefficient-parameter continuation, following the approach developed in Morgan and Sommese (1989). This method leverages information from the solution of a generic system $\tilde{F}(\mathbf{L}, \tilde{\mathbf{s}})$ to more effectively solve the target systems $\tilde{F}(\mathbf{L}, \mathbf{s}_{mt})$. Rather than treating each market as an entirely separate problem, the coefficient-parameter continuation algorithm, outlined in Algorithm 2, enables a smooth transition from the generic system to the target system by gradually varying the parameters. Specifically, after solving the generic system $\tilde{F}(\mathbf{L}, \tilde{\mathbf{s}})$, I use it as the start system to solve the target system $\tilde{F}(\mathbf{L}, \mathbf{s}_{mt})$. This involves tracking solution paths defined by the homotopy: $F(\mathbf{L}, h\tilde{\mathbf{s}} + (1-h)\mathbf{s}_{mt})$, as h transitions from 0 to 1. The choice of a random $\tilde{\mathbf{s}}$ from the space of complex vectors in the earlier step ensures that the homotopy is well-behaved, with a non-singular Jacobian throughout the deformation. This method is substantially more efficient because it exploits the continuity between the generic and target systems, reducing the number of paths that need to be tracked.

A total of 813,609 paths are tracked, of which only $2I+1$ correspond to real solutions.[54] Denote the set of these paths by $\mathcal{T}^{real}$, and let τ index each path. From Proposition 3, we know that the solution on each path corresponds to a type, allowing me to use the terms *paths* and *types* interchangeably.

Results. There are a total of $2I+1$ equilibrium types, with each market exhibiting an odd number of possible equilibria, ranging from 1 to $I-1$ types, depending on its underlying

[54] One might be tempted to track only the paths corresponding to the real solutions of the start system. This would be a mistake. It is essential to also track paths corresponding to complex solutions; otherwise, we risk missing real solutions of the target system.

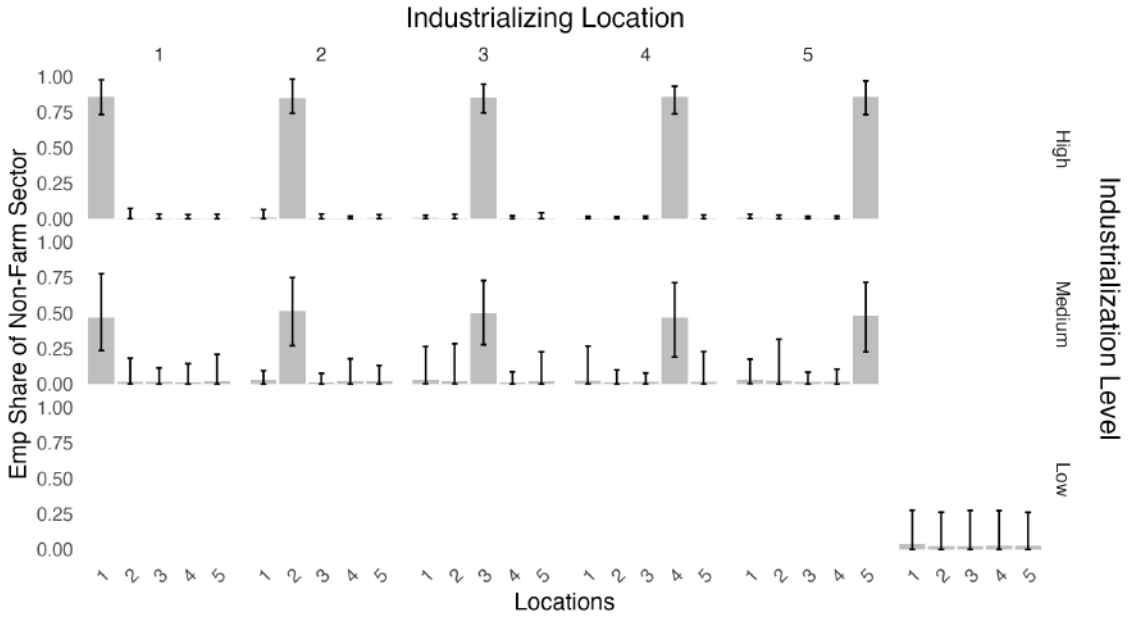


Figure 11: Distribution of Economic Activity Across Different Equilibrium Types

Notes: The figure displays the average employment share of the non-farm sector in the five central locations, with the first location being the centre that receives the industrial zone. Averages are calculated across all instances of each equilibrium type, whether factual or counterfactual. For instance, if the equilibrium labeled 'Type = 1, High' occurs in 200 out of 3000 markets, the average is taken across these 200 cases, regardless of whether the equilibrium is ultimately selected. Error bars indicate the minimum and maximum employment shares observed across these instances. Each panel corresponds to a distinct equilibrium type, labeled as 'Type = i, Level', where i represents the industrializing region (shown as columns), and 'Level' denotes the intensity of industrialization (shown as rows). The label 'Type = Low' represents no industrialization in the central locations, which accounts for 78.9% of the sample.

fundamentals. The $2I$ equilibrium types correspond to scenarios where one of the five central locations industrializes at varying levels (high or medium), while the final type represents low industrialization across all regions (Figure 11). Among markets that select a low industrialization equilibrium, 17.26 percent have the potential to achieve a high industrialization equilibrium. Conversely, among markets that select a high industrialization equilibrium, 13.63 percent have the possibility of selecting a low industrialization equilibrium.

The proportion of regions selecting the low industrialization equilibrium in each census year—1991, 2001, and 2011—is shown in Figure 12a. Over time, regions generally transition out of low-industrialization equilibrium traps. This trend is observed in both treated and control regions. However, the escape from the low-industrialization equilibrium trap is notably stronger in regions that receive the zones. The impact of these zones on equilibrium selection is systematically analyzed within an event study framework in the next subsection.

4.3 Impact of Industrial Policy on Equilibrium Selection

Once we recover the types of the factual and counterfactual equilibrium, the next step is to i) regress selected types on zones to study how zones affect equilibrium selection, and ii) decompose the observed equilibrium changes into due to fundamentals- and coordination driven. I first report results from the first set of analyses.

Empirical Strategy. Since we are interested in the effect of industrial zones on the ability of regions to escape low-industrialization equilibrium traps, I focus on regions that were selecting the low-industrialization equilibrium type in the baseline period $s=-1$. I estimate the effects of the introduction of an industrial zone on equilibrium selection using a difference-in-difference (DiD) design, similar to the approach in Section 3.

As before, the first difference compares regional outcomes in a given period with outcomes in the census period immediately preceding treatment. The second difference compares contemporaneous outcomes between treated regions and comparison regions belonging to the *same* P-Score Group as the treated region. Formally, the specification is given by:

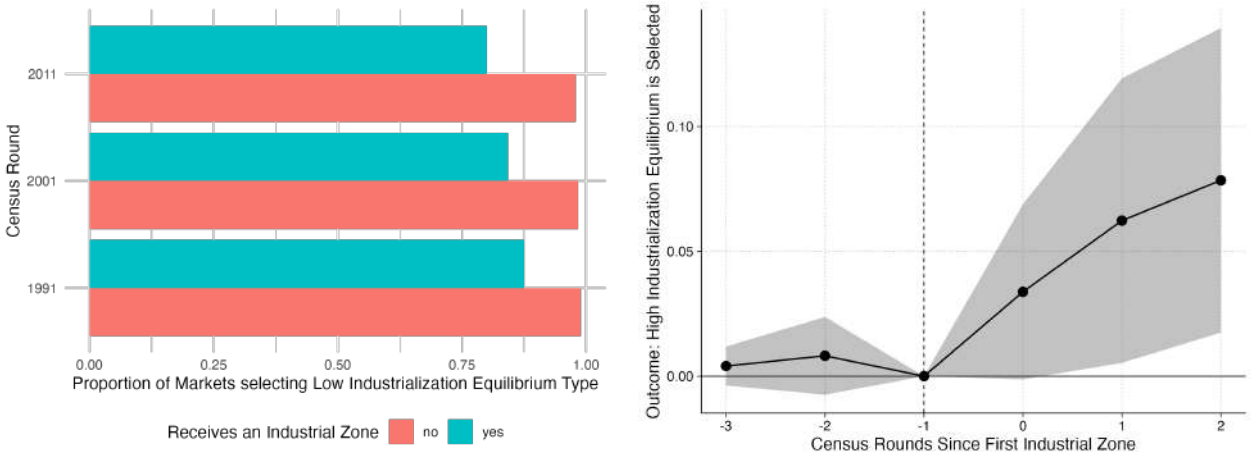
$$y_{mt} = \alpha_m + \beta_{g(m)t} + \sum_{s \neq -1} \beta_{DiD,s} \mathbf{1}[t - TreatRound_{g(m)} = s] Zone_m + \epsilon_{mt} \tag{13}$$

where m denotes the region, market, or catchment area, $g(m)$ indicates the P-Score Group of region m, and t represents the census round or time period. α_m captures market fixed effects, while $\beta_{g(m)t}$ represents P-Score Group$\times$period fixed effects that control for common shocks across all regions in a given P-Score Group. $TreatRound_{g(m)}$ denotes the census round in which the treated municipality in group $g(m)$ receives its first zone. The outcome variable y_{mt} is an indicator for whether a high-industrialization equilibrium is selected in market m at time t. Identification of the treatment effect relies on the assumption that comparison regions provide a suitable counterfactual for treated regions after accounting for time-invariant regional differences and common time-specific shocks. An important validation test for this assumption is the absence of differential pre-treatment trends between treated and comparison regions.

Results. Figure 12b presents the difference-in-differences (DiD) estimates from the model specified in Equation 13. These estimates correspond to the baseline comparison group—regions matched on a range of natural and economic attributes, including the equilibrium type observed in the data—and the baseline specification, which accounts for common shocks at the P-Score group level. Before the treatment, both treated and control groups exhibit a very low probability of selecting a high-industrialization equilibrium. However, ten years after treatment, regions receiving an industrial zone experience a substantial increase in the probability of selecting the high-industrialization equilibrium. This increase represents a 38% rise relative to the baseline probability of selecting a high-industrialization equilibrium among treated regions in the

unrestricted sample.

Figure 12: Determinants of Equilibrium Selection



(a) Equilibrium Selection over Time

(b) Impact of Zones on Equilibrium Selection

Notes: The figure depicts the event study coefficients from a difference-in-difference regression specification for different periods. Each period is equivalent to 10 years. The outcome is an indicator variable for whether a high-industrialization equilibrium even is realised in the market. The proportion of treated (control) markets selecting a high industrialization equilibrium in the baseline is 16.16% (2.3%) for the unrestricted sample. The sample is restricted to markets that are on low-industrialisation equilibrium in the baseline period -1. All the event studies include fixed effects for market, and Pscore-group$\times$year. Standard errors are clustered at the market level.

4.4 Decomposition of Effects of Industrial Policy

Given that we have identified the equilibrium types for both the observed (factual) and counterfactual equilibria, we can now decompose the observed effects of policy into two distinct components: those attributable to economic fundamentals and those arising from changes in equilibrium selection, referred to as the equilibrium switching or coordination effects.

Empirical Strategy. Denote the allocation of labor for the low-industrialization equilibrium type in market m at time t as $L_{ik,mt}^{\text{Type = Low}}$. This allocation is represented as an $L \times 2$ dimensional vector, capturing the distribution of labor across sectors and space, under the low-industrialization equilibrium. If the low equilibrium type is the equilibrium realized in market m at time t, then $L_{ik,mt}^{\text{Type = Low}} = L_{ik,mt}$, where $L_{ik,mt}$ is the labor allocation actually observed. However, if the observed equilibrium does not correspond to the low equilibrium type, the two allocations—$L_{ik,mt}^{\text{Type = Low}}$ and $L_{ik,mt}$—will differ, indicating the selection of a different equilibrium type.

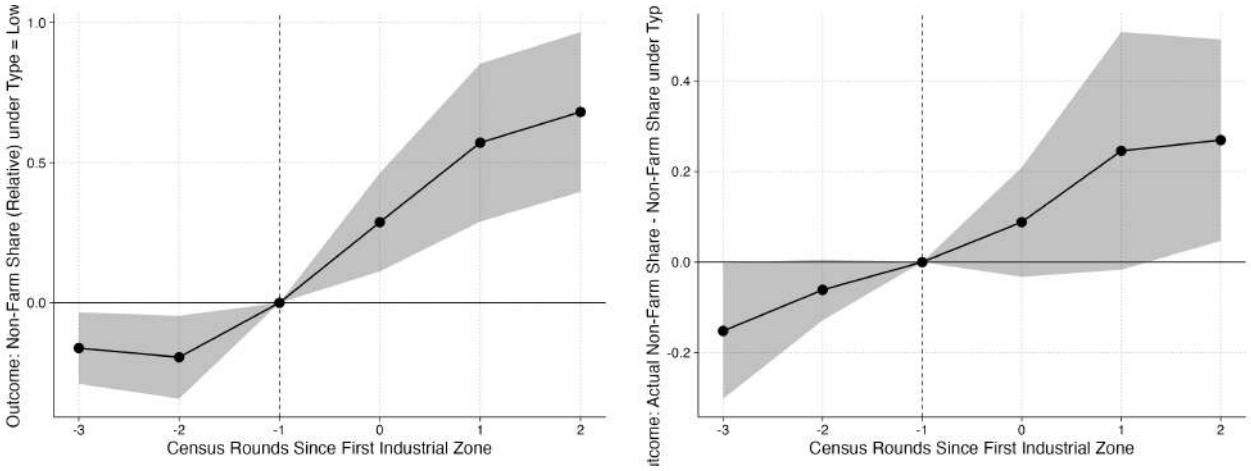
For each market, I compute the following measures

$$y_{mt} = \frac{L_{non-farm,mt}}{L_{farm,mt}}; \quad y_{mt}^{\text{Type = Low}} = \frac{L_{non-farm,mt}^{\text{Type = Low}}}{L_{farm,mt}^{\text{Type = Low}}}$$

where y_{mt} represents the relative labor share of the non-farm sector in the observed equilibrium, and $y_{mt}^{\text{Type = Low}}$ denotes the counterfactual equilibrium share, assuming that the equilibrium type remained consistent with the low-industrialization baseline. Here, $L_{non-farm,mt}$ and $L_{farm,mt}$ refer to the labor allocations in the observed equilibrium, while $L_{non-farm,mt}^{\text{Type = Low}}$ and $L_{farm,mt}^{\text{Type = Low}}$ reflect these allocations under the low equilibrium type. I use $y_{mt}^{\text{Type = Low}}$ and the difference $y_{mt} - y_{mt}^{\text{Type = Low}}$ as outcome variables in the regression specified by Equation 13, where the former isolates the effect due to fundamentals, while the latter captures the effect due to coordination, or equilibrium switching.

Results. The estimated effect of industrial zones on $y_{mt}^{\text{Type = Low}}$, interpreted as the effect attributable to fundamentals, is illustrated in Figure 13a. In contrast, Figure 13b shows the effect of zones on $y_{mt} - y_{mt}^{\text{Type = Low}}$, interpreted as the coordination effect. These results indicate that industrial zones influence equilibrium outcomes through both shifts in fundamentals and changes in equilibrium selection, with the coordination effect accounting for approximately 32.47% of the observed total effect. This decomposition underscores the dual role of industrial zones in both directly altering economic fundamentals and facilitating coordination that drives equilibrium switching.

Figure 13: Decomposition of Reduced Form Effects



(a) Impact on Non-Farm Share via Fundamentals (b) Impact on Non-Farm Share via Coordination

Notes: The figure depicts the event study coefficients from a difference-in-difference regression specification for different periods. Each period is equivalent to 10 years. The outcome on the left hand side is $y_{mt}^{Type=Low}$, and the outcome on the right hand side is $y_{mt} - y_{mt}^{Type=Low}$. All the event studies include fixed effects for market, and Pscore-group$\times$year. Standard errors are clustered at the market level.

5 Conclusion

In this paper, I introduce a new econometric framework for studying real-world settings with multiple equilibria, focusing on the role of industrial policy in a guiding an economy toward a more favorable equilibrium. The approach developed in the paper allows for the estimation of the equilibrium selection function imposing minimal restrictions on unobserved heterogeneity, making it possible to isolate the impact of policy shocks on equilibrium switching from their direct effects on underlying fundamentals. In a novel application to industrial zones in India, I find that the establishment of these zones increases the likelihood of switching from low- to high-industrialization equilibria by around 38% over 10 years. This quantification of a critical mechanism highlights the effectiveness of industrial zones as a policy tool.

An important contribution of this work is the novel data collection effort, which provides a foundation for further research on industrial zones. These zones are widespread across the developing world, including China, India, Southeast Asia, and Africa, where they have played and continue to play an important role in industrialization efforts. Industrial zones facilitate investment, promote agglomeration economies, and help mitigate market inefficiencies such as coordination failures. The dataset collected here enables a more rigorous investigation of these mechanisms, such as the source of agglomeration economies, offering a valuable resource for future research. Another avenue for future research is how parks facilitate coordination - is it

explicitly by bringing firms in a room together such as the Mesas Ejecutivas in Peru, are they providing implicit off-equilibrium guarantees of continued investment in public infrastructure, or are they benefitting from the past success of highly interventionist policies of the 1960s and 1970s? Understanding these mechanisms can further inform the optimal placement and design of these policies, which are critical to industrial development in these regions.

The primary contribution of this work, however, lies in the methods developed in Section 2, which are adaptable to a range of other quasi-experimental settings beyond industrial policy. Do large highway construction projects create self-fulfilling prophecies about the characteristics of a neighborhood? Do temporary supply chain disruptions permanently alter who firms trade with? Do large trade liberalization events coordinate labor markets on a different equilibrium?[55] This paper provides a parsimonious starting point towards understanding complementarities, policy multipliers, multiplicity and equilibrium selection in a rich set of real-world scenarios.

[55]Reduced form evidence consistent with the coordination aspect of highways in the US has been presented recently by Jefferey Lin and Kyle Mangum. Long-run persistence of the labor market effects has been noted in Dix-Carneiro and Kovak (2017) in the context of Brazilian trade liberalization

References

Adsera, Alicia and Debraj Ray, "History and coordination failure," *Journal of Economic Growth*, 1998, *3* (3), 267–276.

Aggarwal, Aradhna, "Promoting agglomeration economies and industrial clustering through SEZs: Evidence from India," *Journal of International Commerce, Economics and Policy*, 2011, *2* (02), 201–227.

Aguirregabiria, Victor, "A method for implementing counterfactual experiments in models with multiple equilibria," *Economics Letters*, 2012, *114* (2), 190–194.

_ , Allan Collard-Wexler, and Stephen P. Ryan, "Dynamic games in empirical industrial organization," in Kate Ho, Ali Hortaçsu, and Alessandro Lizzeri, eds., *Handbook of Industrial Organization*, Vol. 4, Elsevier, 2021, pp. 225–343.

_ and Pedro Mira, "Identification of games of incomplete information with multiple equilibria and unobserved heterogeneity," *Quantitative Economics*, 2019, *10* (4), 1659–1701.

Ahlfeldt, Gabriel M, Stephen J Redding, Daniel M Sturm, and Nikolaus Wolf, "The economics of density: Evidence from the Berlin Wall," *Econometrica*, 2015, *83* (6), 2127–2189.

Alexander, P. C., *Industrial Estates in India*, Asia Publishing House, 1963.

Allen, Treb, Costas Arkolakis, and Xiangliang Li, "On the equilibrium properties of network models with heterogeneous agents," Technical Report, National Bureau of Economic Research 2020.

Arthur, W Brian, *Industry location patterns and the importance of history*, Center for Economic Policy Research, Stanford University, 1986.

Asher, Sam, Tobias Lunt, Ryu Matsuura, and Paul Novosad, "Development research at high geographic resolution: an analysis of night-lights, firms, and poverty in India using the shrug open data platform," *The World Bank Economic Review*, 2021, *35* (4), 845–871.

Bajari, Patrick, Han Hong, and Denis Nekipelov, "The Econometrics of Static and Dynamic Models of Strategic Interaction," 2024. Working paper, under review. Solicited by the Handbook of Econometrics.

_ , _ , and Stephen P. Ryan, "Identification and estimation of a discrete game of complete information," *Econometrica*, 2010, *78* (5), 1529–1568.

_ , _ , **John Krainer, and Denis Nekipelov**, "Computing equilibria in static games of incomplete information using the all-solution homotopy," *Operations Research*, 2010, *58*, 237–245.

Barbanchon, Thomas Le, Roland Rathelot, and Alexandra Roulet, "Gender differences in job search: Trading off commute against wage," *The Quarterly Journal of Economics*, 2021, *136* (1), 381–426.

Barwick, Panle Jia, Myrto Kalouptsidi, and Nahim Zahur, "Industrial Policy Implementation: Empirical Evidence from China's Shipbuilding Industry," *Review of Economic Studies, Forthcoming*, December 2023.

Becko, John Sturm, "How to Fix a Coordination Failure: A Super-PigouvianÄpproach," Technical Report, Technical Report 2023.

Blakeslee, David, Ritam Chaurey, Ram Fishman, and Samreen Malik, "Land Rezoning and Structural Transformation in Rural India: Evidence from the Industrial Areas Program," *The World Bank Economic Review*, 02 2021, *36* (2), 488–513.

Brock, William A and Steven N Durlauf, "Discrete choice with social interactions," *The Review of Economic Studies*, 2001, *68* (2), 235–260.

Buera, Francisco J, Hugo Hopenhayn, Yongseok Shin, and Nicholas Trachter, "Big push in distorted economies," Technical Report, National Bureau of Economic Research 2021.

Chaisemartin, Clément De and Xavier d'Haultfoeuille, "Two-way fixed effects and differences-in-differences with heterogeneous treatment effects: A survey," *The Econometrics Journal*, 2023, *26* (3), C1–C30.

Choi, Jaedo and Andrei A Levchenko, "The long-term effects of industrial policy," Technical Report, National Bureau of Economic Research 2021.

_ **and Younghun Shim**, "Industrialization and the Big Push: Theory and Evidence from South Korea," 2024.

Ciccone, Antonio, "Input chains and industrialization," *The Review of Economic Studies*, 2002, *69* (3), 565–587.

Ciliberto, Federico and Elie Tamer, "Market structure and multiple equilibria in airline markets," *Econometrica*, 2009, *77* (6), 1791–1828.

Criscuolo, Chiara, Ralf Martin, Henry G. Overman, and John Van Reenen, "Some Causal Effects of an Industrial Policy," *American Economic Review*, 2019, *109* (1), 48–85.

Crouzet, Nicolas, Apoorv Gupta, and Filippo Mezzanotti, "Shocks and technology adoption: Evidence from electronic payment systems," *Journal of Political Economy*, 2023, *131* (11), 3003–3065.

Dix-Carneiro, Rafael and Brian K Kovak, "Trade liberalization and regional dynamics," *American Economic Review*, 2017, *107* (10), 2908–2946.

Eaves, B Curtis and Karl Schmedders, "General equilibrium models and homotopy methods," *Journal of Economic Dynamics and Control*, 1999, *23* (9-10), 1249–1279.

Frankel, David M. and Ady Pauzner, "Resolving Indeterminacy in Dynamic Settings: The Role of Shocks," *Quarterly Journal of Economics*, 2000, *115* (1), 285–304.

Gallé, Johannes, Daniel Overbeck, Nadine Riedel, and Tobias Seidel, "Place-based policies, structural change and female labor: Evidence from India's Special Economic Zones," *Available at SSRN 4210059*, 2022.

Ghezzi, Piero, "Mesas ejecutivas in Peru: lessons for productive development policies," *Global Policy*, 2017, *8* (3), 369–380.

Gindling, Thomas H and David Newhouse, "Self-employment in the developing world," *World development*, 2014, *56*, 313–331.

Grant, Matthew, "Why special economic zones? Using trade policy to discriminate across importers," *American Economic Review*, 2020, *110* (5), 1540–71.

Hanlon, W Walker, "The persistent effect of temporary input cost advantages in shipbuilding, 1850 to 1911," *Journal of the European Economic Association*, 2020, *18* (6), 3173–3209.

Higgins, Sean, "Financial Technology Adoption: Network Externalities of Cashless Payments in Mexico," *American Economic Review*, November 2024, *114* (11), 3469–3512.

Hirschman, Albert O., *The Strategy of Economic Development*, New Haven: Yale University Press, 1958.

Hoff, Karla and Joseph E Stiglitz, *Modern economic theory and development*, Citeseer, 2000.

Hornbeck, Richard, Shanon Hsuan-Ming Hsu, Anders Humlum, and Martin Rotemberg, "Gaining Steam: Incumbent Lock-in and Entrant Leapfrogging," Technical Report, National Bureau of Economic Research 2024.

III, Raymond Owens, Esteban Rossi-Hansberg, and Pierre-Daniel Sarte, "Rethinking detroit," *American Economic Journal: Economic Policy*, 2020, *12* (2), 258–305.

Jordan, Luke Simon and Yannick Saleman, "The Implementation of Industrial Parks: Some Lessons Learned in India," 2014. Disclosed, Country: India, Region: South Asia.

Jovanovic, Boyan, "Observable implications of models with multiple equilibria," *Econometrica: Journal of the Econometric Society*, 1989, pp. 1431–1437.

Juhász, Réka, "Temporary Protection and Technology Adoption: Evidence from the Napoleonic Blockade," *American Economic Review*, 2018, *108* (11), 3339–3376.

_ , Nathan Lane, and Dani Rodrik, "The New Economics of Industrial Policy," Technical Report, National Bureau of Economic Research 2023.

Kalouptsidi, Myrto, "Detection and Impact of Industrial Subsidies: The Case of Chinese Shipbuilding," *Review of Economic Studies*, 2018, *85* (2), 1111–1158.

Kim, Minho, Munseob Lee, and Yongseok Shin, "The Plant-Level View of an Industrial Policy: The Korean Heavy Industry Drive of 1973," NBER Working Paper 29252, National Bureau of Economic Research September 2021.

Kline, Patrick and Enrico Moretti, "Local economic development, agglomeration economies, and the big push: 100 years of evidence from the Tennessee Valley Authority," *The Quarterly journal of economics*, 2014, *129* (1), 275–331.

Krugman, Paul, "History Versus Expectations," *Quarterly Journal of Economics*, 1991, *106*, 651–667.

Lane, Nathan, "Manufacturing revolutions: Industrial policy and industrialization in South Korea," *Available at SSRN 3890311*, 2022.

Lu, Yi, Jin Wang, and Lianming Zhu, "Place-Based Policies, Creation, and Agglomeration Economies: Evidence from China's Economic Zone Program," *American Economic Journal: Economic Policy*, August 2019, *11* (3), 325–360.

Manelici, Isabela and Smaranda Pantea, "Industrial Policy at Work: Evidence from Romania's Income Tax Break for Workers in IT," *European Economic Review*, 2021, *133*, 103674.

Manski, Charles F, "Identification of endogenous social effects: The reflection problem," *The review of economic studies*, 1993, *60* (3), 531–542.

Mas-Colell, Andreu, Michael D. Whinston, and Jerry R. Green, *Microeconomic Theory*, 1st ed., New York, NY: Oxford University Press, 1995.

Matsuyama, Kiminori, "Increasing Returns, Industrialization, and Indeterminacy of Equilibrium," *Quarterly Journal of Economics*, 1991, *106*, 616–650.

Monte, Ferdinando, Charly Porcher, and Esteban Rossi-Hansberg, "Remote work and city structure," Technical Report, National Bureau of Economic Research 2023.

Morgan, Alexander P and Andrew J Sommese, "Coefficient-parameter polynomial continuation," *Applied Mathematics and Computation*, 1989, *29* (2), 123–160.

Murphy, Kevin M., Andrei Shleifer, and Robert W. Vishny, "Industrialization and the Big Push," *Journal of Political Economy*, 1989, *97*, 1003–1026.

Myrdal, Gunnar, *Economic Theory and Underdeveloped Regions*, London: Duckworth, 1957.

Neumark, David and Helen Simpson, "Place-based policies," in "Handbook of regional and urban economics," Vol. 5, Elsevier, 2015, pp. 1197–1287.

Newey, Whitney K and James L Powell, "Instrumental variable estimation of nonparametric models," *Econometrica*, 2003, *71* (5), 1565–1578.

Ouazad, Amine CL, "Equilibrium Multiplicity: A Systematic Approach using Homotopies, with an Application to Chicago," *arXiv preprint arXiv:2401.10181*, 2024.

Paula, Aureo De, "Econometric analysis of games with multiple equilibria," *Annu. Rev. Econ.*, 2013, *5* (1), 107–131.

_ **and Xiaoxia Tang**, "Inference of signs of interaction effects in simultaneous games with incomplete information," *Econometrica*, 2012, *80* (1), 143–172.

Petrongolo, Barbara and Maddalena Ronchi, "Gender gaps and the structure of local labor markets," *Labour Economics*, 2020, *64*, 101819.

Redding, Stephen J, Daniel M Sturm, and Nikolaus Wolf, "History and industry location: evidence from German airports," *Review of Economics and Statistics*, 2011, *93* (3), 814–831.

Rodriguez-Clare, Andres, "Coordination failures, clusters, and microeconomic interventions," *Economía*, 2005, *6* (1), 1–42.

Rodrik, Dani, "Getting Interventions Right: How South Korea and Taiwan Grew Rich," *Economic Policy*, 1995, *10* (20), 53–107.

_ , "Coordination Failures and Government Policy: A Model with Applications to East Asia and Eastern Europe," *Journal of International Economics*, 1996, *40* (1-2), 1–22.

Rosenstein-Rodan, Paul N, "Problems of industrialisation of eastern and south-eastern Europe," *The economic journal*, 1943, *53* (210-211), 202–211.

Rotemberg, Martin, "Equilibrium Effects of Firm Subsidies," *American Economic Review*, 2019, *109* (10), 3475–3513.

Roth, I., "Industrial Location and Indian Government Policy," *Asian Survey*, 1970, *10* (5), 383–396.

Roth, Jonathan, Pedro HC Sant'Anna, Alyssa Bilinski, and John Poe, "What's trending in difference-in-differences? A synthesis of the recent econometrics literature," *Journal of Econometrics*, 2023, *235* (2), 2218–2244.

Sanghvi, Rajnikant Laxmichand, "Role of industrial estates in a developing economy," *(No Title)*, 1979.

Sarma, NA, "Economic development in India: The first and second five year plans," *Staff Papers-International Monetary Fund*, 1958, *6* (2), 180–238.

Seim, Katja, "An empirical model of firm entry with endogenous product-type choices," *The RAND Journal of Economics*, 2006, *37* (3), 619–640.

Sommese, Andrew J. and Charles W. Wampler, *The Numerical Solution of Systems of Polynomials Arising in Engineering and Science*, 1st ed., Singapore: World Scientific, 2005.

Sommese, Andrew J, Jan Verschelde, and Charles W Wampler, "Introduction to numerical algebraic geometry," *Solving Polynomial Equations: Foundations, Algorithms, and Applications*, 2005, pp. 301–337.

Sun, Liyang and Sarah Abraham, "Estimating dynamic treatment effects in event studies with heterogeneous treatment effects," *Journal of econometrics*, 2021, *225* (2), 175–199.

Sweeting, Andrew, "The strategic timing incentives of commercial radio stations: An empirical analysis using multiple equilibria," *The RAND Journal of Economics*, 2009, *40* (4), 710–742.

Tamer, Elie, "Incomplete simultaneous discrete response model with multiple equilibria," *The Review of Economic Studies*, 2003, *70* (1), 147–165.

UNIDO, *Industrial Estates: Principles and Practice*, UNIDO, 1997. Cover title.

_ , "International Guidelines for Industrial Parks," Technical Report, United Nations Industrial Development Organization (UNIDO) 2020. Accessed: 18-October-2024.

Zheng, Siqi, Weizeng Sun, Jianfeng Wu, and Matthew E Kahn, "The birth of edge cities in China: Measuring the effects of industrial parks policy," *Journal of Urban Economics*, 2017, *100*, 80–103.

Zhou, Jie, "Firewall for Innovation," November 2024. Working paper.

A Appendix to Section 2

Example: Entry Game between Firms

The class of models considered in this section is general. Consider a model of firm entry into a market, where each firm, indexed by f, belongs to a specific sector, indexed by i. Firms simultaneously decide whether to enter the market. The payoff from entering depends on sector-specific, common-knowledge primitives, such as entry costs, denoted by $\mathbf{s}_i$, and the proportion of firms from other sectors entering the market, denoted by $\mathbf{p}=\{p_i\}$. Additionally, the payoff includes a private action-specific additive shock, represented by ϵ_{fi1} for entering and ϵ_{fi0} for not entering. The private shocks are assumed to be independent and identically distributed.

The firm's payoff from entering the market is given by $\pi_i(\mathbf{p},\mathbf{s}_i)+\epsilon_{fi1}$, while the payoff from not entering is ϵ_{fi0}. Importantly, the payoff structure is symmetric within each sector, aside from the private shock. This symmetry, along with the i.i.d. assumption on private shocks, allows us to focus on the average entry in each sector as the equilibrium outcome.

A firm f enters the market only if $\pi_i(\boldsymbol{p},s_i)+\epsilon_{fi1}\geq\epsilon_{fi0}$. Let's assume that the private shocks follow a Type 2 Extreme Value distribution. Since ϵ_{fi1} and ϵ_{fi0} are private information, the proportion of entrants is unknown. However, under consistent beliefs, the Bayes-Nash Equilibrium can be described by a vector of ex-ante choice probabilities $\boldsymbol{p}=\{p_i\}$ which solve the following system of equations:[56]

$$p_i=\frac{\exp(\pi_i(\boldsymbol{p},s_i))}{1+\exp(\pi_i(\boldsymbol{p},s_i))}\equiv F_i(\boldsymbol{p},s_i)\qquad\forall i\in I$$

As with Example 1, we can verify that this system satisfies Assumption 1 under mild assumptions on π.

A.1 Formal Statement of the Total-Degree Homotopy

Theorem 1 (Total-Degree Homotopy (Sommese and Wampler, 2005, Thm. 8.4.1))**.** *Given a system of polynomials*

$$f(z)=\{f_1(z),...,f_n(z)\}:\mathbb{C}^n\longrightarrow\mathbb{C}^n$$

with the degree of f_i equal to d_i, let $g(z)$ be any system of polynomials of matching degrees such that $g(z)=0$ has $d=\prod_i^n d_i$ nonsingular solutions. Then the d solution paths of the homotopy

$$h(z,t)=\gamma t g(z)+(1-t)f(z)=0,\qquad t\in[0,1].$$

[56] The mathematical structure of the equilibrium system is similar to that of Example 2. The only difference is in the denominator—in the entry game, firms in I sectors choose between binary actions, while in the economic geography model, one kind of worker chooses between I locations.

starting at the solutions of $g(z)=0$ are nonsingular for $t\in(0,1]$ and their endpoints as $t\to 0$ include all of the nonsingular solutions of $f(z)=0$ for almost all $\gamma\in\mathbb{C}$, are nonsingular for Then for all but finitely many $\gamma\in\mathbb{C}$, excepting a finite number of real-one-dimensional rays through the origin. In particular, restricting γ to the unit circle $\{\gamma=e^{i\theta}\}$ the exceptions are a finite number of points $\theta\in[0,2\pi)$.

A.2 Proofs

A.2.1 Proof of proposition 2

Proposition (1.1a)**.** *Almost all equilibria are regular.*

Proof. The result follows from the direct application of the Transversality Theorem (MWG Proposition 17.D.3) which I state below for completeness.

The Transversality Theorem If the $M\times(N+S)$ matrix $Df(v;q)$ has rank M whenever $f(v;q)=0$, then for almost every q, the $M\times N$ matrix $D_v f(v;q)$ has rank M whenever $f(v;q)=0$.

We have as many equilibrium equations as we have unknowns, that is $M=N$. Hence, under the assumption that $F_{\mathbf{y},\mathbf{s}}$ is full rank for all $\mathbf{y}\in\phi(\mathbf{s})$ (Assumption 1), the result follows. □

Proposition (1.1b)**.** *All regular equilibria are isolated.*

Proof. This follows directly from the fact that we have as many equations as unknowns and that at the regular equilibria, the Jacobian of the system is non-singular. This means that locally, the set of equilibria, that is the preimage of $I-F$ at $\mathbf{0}$, is of dimension $n-n=0$, implying that the equilibrium is isolated. This statement can also be proven directly using the implicit function theorem.

□

Lemma. *The set of equilibria $\phi(\boldsymbol{s})$ for $\boldsymbol{s}\in\bar{S}$ is finite.*

The previous result establishes that all regular equilibria are isolated points. Since we focus on $\mathbf{s}\in\bar{S}$, all equilibria are isolated. This ensures that around each equilibrium, there exists a neighborhood where no other equilibria can be found. Given that the set of equilibria lies within a bounded space, as stated in Assumption 1, combined with continuity of G, it follows that the set of equilibria for $\mathbf{s}\in\bar{S}$ is both closed and bounded. A closed and bounded set in $\mathbb{R}^n$ is compact.

To proceed, we construct an open cover for the set of equilibria by taking open balls centered around each equilibrium point. Since each equilibrium is isolated, it is possible to choose the radius of these balls such that each ball contains exactly one equilibrium, i.e., the intersection of each open ball with the set of equilibria is a singleton. By the compactness of the set, there exists a finite subcover, which implies that the set of equilibria is finite.

Proposition (1.2). *All regular isolated equilibria can be approximated arbitrarily well by a polynomial system.*

Proof. Note that $\mathbf{y} - F(\mathbf{y},\mathbf{s}) = 0$ is a system of n equations, where each function $F_i(\mathbf{y},\mathbf{s})$ is a real-valued, continuous function defined on a compact set. Since the theorem holds for a fixed $\mathbf{s}$, I will suppress the $\mathbf{s}$ argument for simplicity.

By the Stone-Weierstrass Theorem, for each F_i, there exists a sequence of polynomials $\tilde{F}_i(.,d)$ (where d denotes the degree of the polynomial) that uniformly approximates F_i on the compact set. Specifically, for every $\epsilon > 0$, there exists a degree d_i such that for all $\mathbf{y}$ and $d \geq d_i$, the polynomial $\tilde{F}_i(.,d)$ satisfies:

$$|F_i(\mathbf{y}) - \tilde{F}_i(\mathbf{y},d)| < \epsilon.$$

Thus, there exists a polynomial *system* $\tilde{F}(.,d^*)$, where $d^* = \max_{i \in 1 \cdots n} d_i$, that uniformly approximates the system $F(\mathbf{y})$ in all components. Now, consider an arbitrary regular root $\mathbf{y}_0$ such that $\mathbf{y}_0 - F(\mathbf{y}_0) = 0$, and $I - F_{\mathbf{y}}(\mathbf{y}_0)$ is nonsingular. By the uniform convergence of $\tilde{F}(\mathbf{y},d)$ to $F(\mathbf{y})$, for any $\epsilon > 0$, there exists a sufficiently large d such that:

$$|\tilde{F}(\mathbf{y}_0,d) - F(\mathbf{y}_0)| < \epsilon.$$

The difference simplifies as follows:

$$|\tilde{F}(\mathbf{y}_0,d) - F(\mathbf{y}_0)| = |(\mathbf{y}_0 - \tilde{F}(\mathbf{y}_0,d)) - (\mathbf{y}_0 - F(\mathbf{y}_0))| < \epsilon.$$

The term $\mathbf{y}_0 - F(\mathbf{y}_0)$ is zero since $\mathbf{y}_0$ is a root of the system, leaving us with:

$$|\mathbf{y}_0 - \tilde{F}(\mathbf{y}_0,d)| < \epsilon.$$

If ϵ is smaller than the precision threshold of the numerical algorithm used to approximate the roots, the numerical root of $\tilde{F}(\mathbf{y},d)$ will be close to $\mathbf{y}_0$.

Next, we prove that a real root of $\tilde{F}(\mathbf{y},d)$ can be made to be arbitrarilty close to $\mathbf{y}_0$ by choosing large enough d. Consider the closed neighborhood $\bar{B}_\delta(\mathbf{y}_0)$, where $\mathbf{y} - F(\mathbf{y})$ is nonzero. Such a neighborhood exists because we showed earlier that regular roots are isolated. $\mathbf{y} - F(\mathbf{y})$ is bounded in this neighborhood by the Weierstrass Theorem (a continuous function on a compact set is bounded). Let G be an n-dimensional cube that inscribes this neighborhood. The function $\mathbf{y} - F(\mathbf{y})$ must be nonzero on the boundary of the cube G. Furthermore, since $\mathbf{y}_0$ is a root and $I - F_{\mathbf{y}}(\mathbf{y}_0)$ is nonsingular, $F(\mathbf{y})$ must change signs within the neighborhood.

By uniform convergence, the polynomial system $\tilde{F}(\mathbf{y},d)$ inherits this sign change from $F(\mathbf{y})$ for sufficiently large d. Therefore, applying the Poincaré–Miranda Theorem[57] (a generalization of the Intermediate Value Theorem to higher dimensions), we conclude that $\tilde{F}(\mathbf{y},d)$ must have

[57] See Miranda, 1940

at least one root within the neighborhood $B_\delta(\mathbf{y}_0)$.

Thus, we have proven that $\mathbf{y}_0$, a regular root of $\mathbf{y}-F(\mathbf{y})=0$, can be approximated arbitrarily well by a root of the polynomial system of sufficiently high degree.

What remains to show is that *all* regular roots can be approximated arbitrarily well by a polynomial system. Recall that the set of regular roots is finite, as shown in the first part of the proposition. Suppose there are R regular roots, indexed by $r=1,...,R$. For each root $\mathbf{y}_r$, we can apply the above argument to obtain a polynomial approximation $\tilde{F}(\mathbf{y},d(r,\delta))$, where $d(r,\delta)$ is the degree of the polynomial required to approximate the system within a precision δ. To ensure uniform approximation across all roots, we choose the maximum approximation degree $d^*(\delta)=\max_{r=1,...,R} d(r,\delta)$. This guarantees that the polynomial system $\tilde{F}(\mathbf{y},d^*(\delta))$ approximates the system $F(\mathbf{y})$ uniformly across all regular roots, completing the proof.

□

A.2.2 Proof of Propistion 3

Proposition (2.1). *The 'Equilibrium Type' relation is an equivalence relation.*

That the types relation is reflexive and symmetric is obvious from the definition of Equilibrium Types. Therefore, all that remains to show is that the relation is transitive. Let us pick three arbitrary equilibria $\mathbf{e}^1=(\mathbf{s}^1,\mathbf{y}^1\in\psi(\mathbf{s}^1))$, $\mathbf{e}^2=(\mathbf{s}^2,\mathbf{y}^2\in\psi(\mathbf{s}^2))$, and $\mathbf{e}^3=(\mathbf{s}^3,\mathbf{y}^3\in\psi(\mathbf{s}^3))$ such that $\mathbf{e}^1$ belongs to the same type as $\mathbf{e}^2$ and $\mathbf{e}^2$ belongs to the same type as $\mathbf{e}^3$. That is, there are paths $\lambda^1(t)$ connecting $\mathbf{e}^1$ and $\mathbf{e}^2$, and $\lambda^2(r)$ connecting $\mathbf{e}^2$ and $\mathbf{e}^3$, such that:

$$\lambda^1(t)\in\psi\big(\mathbf{s}^1t+\mathbf{s}^2(1-t)\big) \quad \text{and} \quad \lambda^2(t)\in\psi\big(\mathbf{s}^2(1-t)+\mathbf{s}^3t\big).$$

We want to show that $\mathbf{e}^1$ is the same equilibrium type as $\mathbf{e}^3$. In other words, we want to show that there exists a *continuous* path $\lambda^3(r)$ that satisfies

$$\lambda^3(r)\in\psi\big(\mathbf{s}^1(1-r)+\mathbf{s}^3r\big); \qquad \lambda^3(0)=\mathbf{y}^1, \quad \lambda^3(1)=\mathbf{y}^3. \tag{14}$$

I will prove this by construction. First, I will define a two-dimensional homotopy that continuously deforms $\lambda^1(t)$ to $\lambda^2(t)$ as a parameter r moves from 0 to 1. Formally, let's define

$$H(r,t,\mathbf{y})=\mathbf{y}-F(\mathbf{s}(r,t),\mathbf{y}); \qquad \mathbf{s}(r,t)=\mathbf{s}^2(1-t)+(\mathbf{s}^1(1-r)+\mathbf{s}^3r)t.$$

This homotopy is visualized in Figure 14. Note that $H(0,t,\mathbf{y})=0$ is uniquely solved by $\mathbf{y}=\lambda^1(t)$, and $H(1,t,\mathbf{y})=0$ is uniquely solved by $\mathbf{y}=\lambda^2(t)$. I claim that for every $r\in[0,1]$, there exists a unique and continuous curve $\mathbf{y}(r,t)$ that solves $H(r,t,\mathbf{y})=0$. To do this, I first show that $H_\mathbf{y}(r,t,\mathbf{y})$ is nonsingular not only at the edges (for $r\in\{0,1\}$), but everywhere.[58]

[58] Note that this is not obvious—in general, when continuing from $(\mathbf{s},\mathbf{y})$ to $(\mathbf{s}',\mathbf{y}'\neq\mathbf{s},\mathbf{y})$, we may encounter a singular point. The fact that $H_\mathbf{y}$ is nonsingular at the edges is crucial to proving that it is nonsingular everywhere.

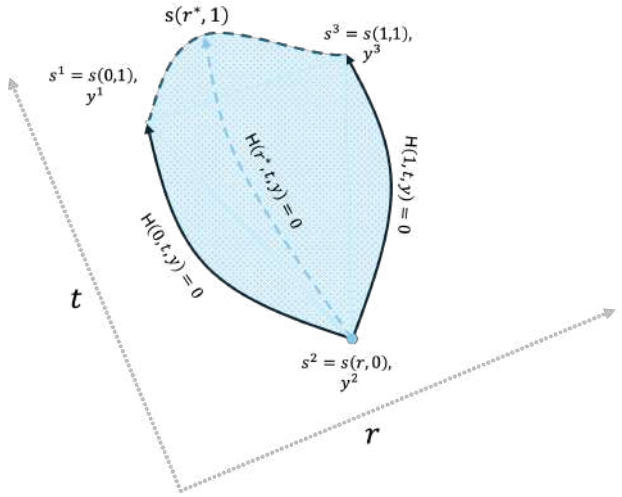


Figure 14: $H(r,t,y)$ continuously deforms $F(\mathbf{s}^1 t+\mathbf{s}^2(1-t),\mathbf{y})$ to $F(\mathbf{s}^3 t+\mathbf{s}^2(1-t),\mathbf{y})$ along r.

Lemma. *$H_{\mathbf{y}}(r,t,\mathbf{y})$ is nonsingular for all $r,t\in[0,1]^2$ and $\mathbf{y}\in Y$ that satisfy $H(r,t,\mathbf{y})=0$.*

Proof. I will prove this by contradiction. Consider an arbitrary point $r^*,t^*,\mathbf{y}^*$ that solve the following system of equations:

$$G(r,t,\mathbf{y})\equiv\begin{bmatrix} H(r,t,\mathbf{y}) \\ \det(H_{\mathbf{y}}(r,t,\mathbf{y})) \end{bmatrix}=0$$

By assumption, $H_{\mathbf{y}}(r,t,\mathbf{y}) = I - F_{\mathbf{y}}(\mathbf{s}(r,t),\mathbf{y})$ is non-singular along $\mathbf{y}(0,t) = \lambda^1(t)$ and $\mathbf{y}(1,t) = \lambda^2(t)$. Also, note that, by construction, $H(r,0,\mathbf{y}) = 0$ is satisfied at a single nonsingular point $\mathbf{y}=\mathbf{y}^2$ for all r. Therefore, it must be that $r^*\in(0,1)$ and $t^*\in(0,1]$.

Now let us ask, if we perturb t^* to $t^*-\delta$, does there exist a unique $r,\mathbf{y}$ near $r^*,\mathbf{y}^*$ that is a solution to the above system? The implicit function theorem says there does, if the Jacobian of $G(r^*,t^*,\mathbf{y}^*)$ with respect to $(r,\mathbf{y})$ is nonsingular. The Jacobian can be written as:

$$\frac{\partial G(r^*,t^*,\mathbf{y}^*)}{\partial(r,\mathbf{y})}=\begin{bmatrix} H_r(r^*,t^*,\mathbf{y}^*) & H_{\mathbf{y}}(r^*,t^*,\mathbf{y}^*) \\ \frac{\partial\det(H_{\mathbf{y}}(r^*,t^*,\mathbf{y}^*))}{\partial r} & \frac{\partial\det(H_{\mathbf{y}}(r^*,t^*,\mathbf{y}^*))}{\partial\mathbf{y}} \end{bmatrix}$$

Given $S\subset\mathbb{R}^m$ and $Y\subset\mathbb{R}^n$, it takes the form:

$$\frac{\partial G(r^*,t^*,\mathbf{y}^*)}{\partial(r,\mathbf{y})}=\begin{bmatrix} -F_{\mathbf{s}}(\mathbf{s}(r^*,t^*),\mathbf{y}^*)_{n\times m}t^*(\mathbf{s}_1-\mathbf{s}_3)_{m\times 1} & I-F_{\mathbf{y}}(\mathbf{s}(r^*,t^*),\mathbf{y}^*)_{n\times n} \\ \underbrace{\frac{\partial\det(I-F_{\mathbf{y}}(\mathbf{s}(r^*,t^*),\mathbf{y}^*))}{\partial r}}_{1\times 1} & \underbrace{\frac{\partial det(I-F_{\mathbf{y}}(\mathbf{s}(r^*,t^*),\mathbf{y}^*)}{\partial\mathbf{y}}}_{1\times n} \end{bmatrix}$$

This matrix is full rank by assumption, since $\mathbf{s}(r^*,t^*)\in S/\bar{S}$ and $\alpha\equiv\mathbf{s}_1-\mathbf{s}_3\in\mathbb{R}^m/\mathbf{0}$. This also implies that the matrix $[G_y\ G_r\ G_t]$ is full-rank. The latter, weaker statement, implies that $\mathbf{0}$ is

a regular value of the augmented system of equations G, and by the regular value theorem, the set of solutions of G is a manifold in $\mathbf{R}^{2+n}$ of co-dimension $n+1$, or a $1-$ dimensional manifold. The former, stronger statement, implies that this manifold does not "bend" as we decrease t^* towards zero, and, therefore, that the manifold intersects the edges as depicted in Figure 15 This contradicts the assumption that $H_{\mathbf{y}}$ is nonsingular at the edges.

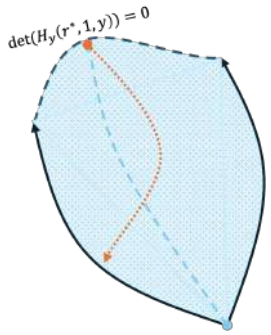


Figure 15: Path along which $H=0$ and $\det(H_{\mathbf{y}})=0$ can be continued until one of the "edges".

□

This allows us to define a family of systems of Ordinary Differential Equations (ODEs) along t (where r parametrizes the family):

$$\frac{\partial \mathbf{y}(r,t)}{\partial t} = -H_{\mathbf{y}}(r,t,\mathbf{y})H_t(r,t,\mathbf{y}) \equiv Q(r,t,\mathbf{y}); \qquad \mathbf{y}(r,0)=\mathbf{y}^2$$

Lemma. *Q is a) bounded, and b) Lipschitz continuous in all three arguments.*

Proof. H is a continuous function defined on a compact set: $(r,t,\mathbf{y}) \in [0,1]^2 \times Y$. Therefore, H must be bounded by the Weierstrass Theorem.

Note that F being C^1 in $S \times Y$ implies that H is C^1 in $[0,1]^2 \times Y$. Since ΔH is continuous and defined on a compact set, it is bounded. That is, there exists $K \in \mathbb{R}^{2+n}$ such that $\Delta H \leq K$. This implies that there exists a c such that $\|H(r,t,\mathbf{y}) - H(r',t',\mathbf{y}')\| \leq c\|(r,t,\mathbf{y}) - (r',t',\mathbf{y}')\|$.[59] □

Lipschitz continuity of $Q(r,t,\mathbf{y})$ in t and $\mathbf{y}$ implies that $\mathbf{y}(r,t)$ exists, is continuous in t, and uniquely solves $H(r,t,\mathbf{y})=0$. This defines a family of curves that takes us from $\mathbf{e}^2$, along t, to all convex combinations of $\mathbf{e}^1$ and $\mathbf{e}^3$. Moreover, because of Lipschitz continuity of Q in r, we have that $\mathbf{y}(r,t)$ is continuous in r (see the "continuous dependence of ODEs on parameters" result in Kelley and Peterson, 2010). Thus, the locus of the endpoints of these curves, $\mathbf{y}(r,1)$, forms a continuous path connecting $\mathbf{e}^1$ and $\mathbf{e}^3$. Defining $\lambda^3(r) = \mathbf{y}(r,1)$ completes the proof.

[59] See James R. Munkres, *Analysis on Manifolds (Addison-Wesley, 1991)*, page 127, for a proof.

Proposition (2.2). *$gr(\psi)$ can be partitioned into a countable number of equivalence classes where each class corresponds to an 'Equilibrium Type'.*

Proof. Since equilibrium type relation is an equivalence relation, it ensures that no equilibrium belongs to two different types. What remains to show is that the set of equilibrium types is countable. This follows from the fact that there are a finite number of regular solutions for every $\mathbf{s} \in S$, and that every regular solution in $\psi(\mathbf{s})$ has an equilibrium belong to the same type in a sufficiently small neighborhood of $\mathbf{s}$. The latter can be argued using the Implicit Function Theorem for $F(\mathbf{y},\mathbf{s})=0$ locally. □

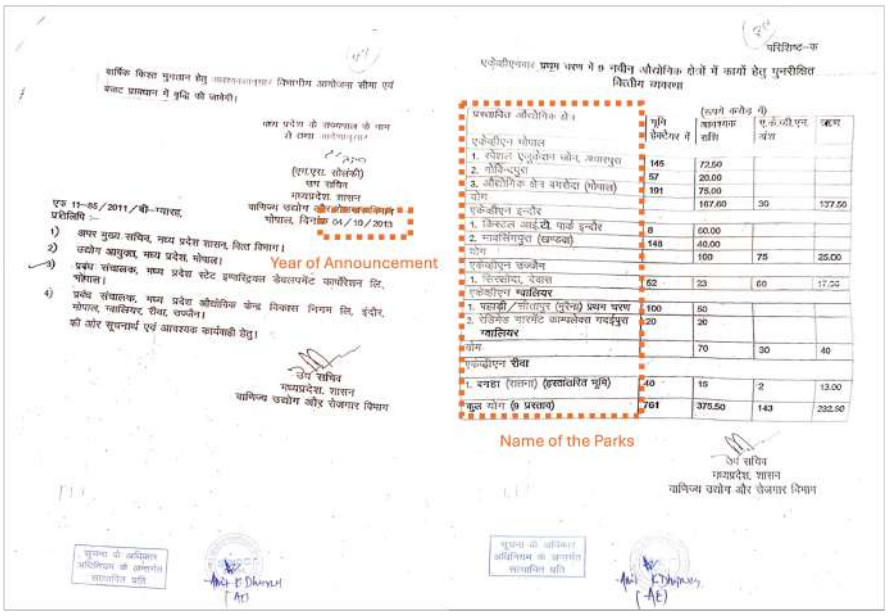
Proposition (2.3). *The type assignment rule $T(\boldsymbol{s},\boldsymbol{y})$ is invertible in $\boldsymbol{y}$. That is, $T(\boldsymbol{s},\boldsymbol{y})=T(\boldsymbol{s},\boldsymbol{y}')$ implies $\boldsymbol{y}=\boldsymbol{y}'$*

Proof. If $\mathbf{y}$ and $\mathbf{y}'$ belong to the same type, then by the definition of equilibrium types, every convex combination of $\mathbf{y}$ and $\mathbf{y}'$ must also belong to $\psi(\mathbf{s})$. This can be understood by observing that, by definition, there exists a continuous equilibrium path connecting $(\mathbf{y},\mathbf{s})$ and $(\mathbf{y}',\mathbf{s}')$, such that every point along the path is a regular solution. Since $\mathbf{s}=\mathbf{s}'$, any convex combination of $\mathbf{s}$ and $\mathbf{s}'$ simply equals $\mathbf{s}$, implying that the entire path lies within $\psi(\mathbf{s})$. This contradicts the fact that each regular equilibrium is isolated. □

B Appendix to Section 3

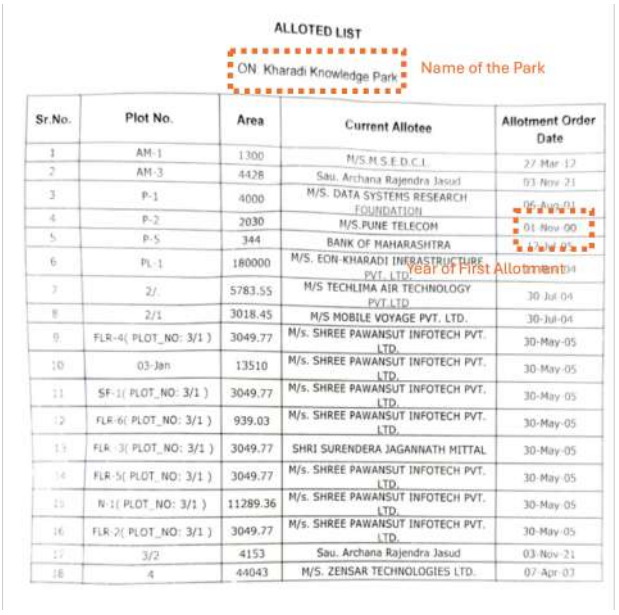
B.1 Example of Source Documents for Data Construction of Industrial Zone

Figure 16: Timing of Establishment of Industrial Zones



(a) Various, Madhya Pradesh

(b) Ahmedabad Apparel Park, Gujarat



Sr.No.	Plot No.	Area	Current Allotee	Allotment Order Date
1	AM-1	1300	M/S.M.S.E.D.C.L.	27-Mar-12
2	AM-3	4428	Sau. Archana Rajendra Jasud	03-Nov-21
3	P-1	4000	M/S. DATA SYSTEMS RESEARCH FOUNDATION	06-Aug-01
4	P-2	2030	M/S.PUNE TELECOM	01-Nov-00
5	P-5	344	BANK OF MAHARASHTRA	07-Jul-05
6	PL-1	180000	M/S. EON-KHARADI INFRASTRUCTURE PVT. LTD.	30-Jul-04
7	2/	5783.55	M/S TECHLIMA AIR TECHNOLOGY PVT.LTD	30-Jul-04
8	2/1	3018.45	M/S MOBILE VOYAGE PVT. LTD.	30-Jul-04
9	FLR-4(PLOT_NO: 3/1)	3049.77	M/s. SHREE PAWANSUT INFOTECH PVT. LTD.	30-May-05
10	03-Jan	13510	M/s. SHREE PAWANSUT INFOTECH PVT. LTD.	30-May-05
11	SF-1(PLOT_NO: 3/1)	3049.77	M/s. SHREE PAWANSUT INFOTECH PVT. LTD.	30-May-05
12	FLR-6(PLOT_NO: 3/1)	939.03	M/s. SHREE PAWANSUT INFOTECH PVT. LTD.	30-May-05
13	FLR -3(PLOT_NO: 3/1)	3049.77	SHRI SURENDERA JAGANNATH MITTAL	30-May-05
14	FLR-5(PLOT_NO: 3/1)	3049.77	M/s. SHREE PAWANSUT INFOTECH PVT. LTD.	30-May-05
15	N-1(PLOT_NO: 3/1)	11289.36	M/s. SHREE PAWANSUT INFOTECH PVT. LTD.	30-May-05
16	FLR-2(PLOT_NO: 3/1)	3049.77	M/s. SHREE PAWANSUT INFOTECH PVT. LTD.	30-May-05
17	3/2	4153	Sau. Archana Rajendra Jasud	03-Nov-21
18	4	44043	M/S. ZENSAR TECHNOLOGIES LTD.	07-Apr-03

(c) Kharadi Knowledge Park, Maharashtra

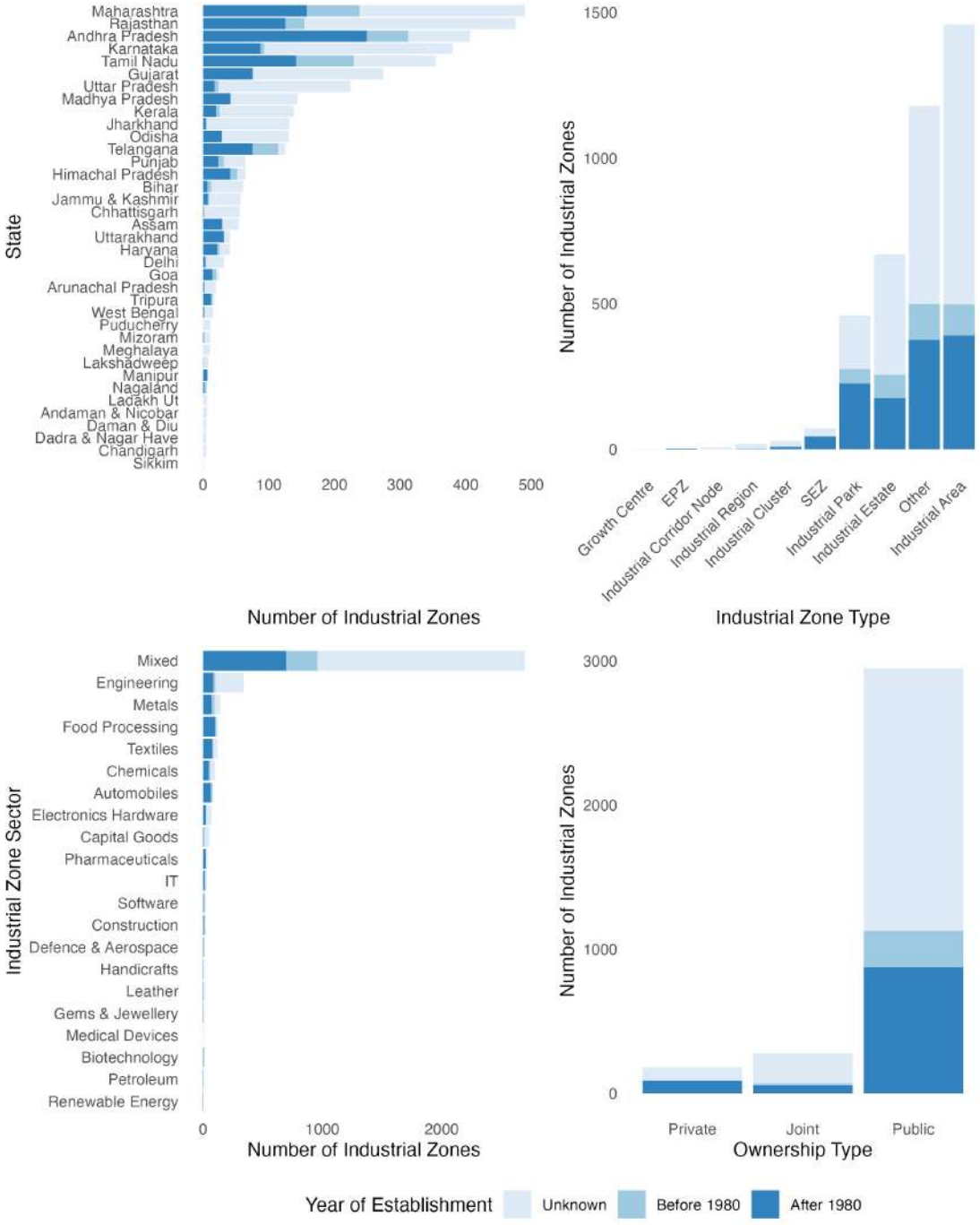
Notes: The figure shows example documents solicited under the Right To Information Act 2005 from various development corporations on the timing of establishment of industrial zones. For parks on the top panel, the dates of announcement were used as the time of establishment. For parks on the bottom panel, first dates of allotment were used as the time of establishment.

B.2 Industrial Zones Descriptive Facts

Figure 17 provides a summary of the key characteristics of industrial zones across India. These zones are predominantly located in large states, especially Maharashtra, Rajasthan, Karnataka, Tamil Nadu, and Gujarat. This distribution reflects both my sample and the national landscape of industrial zones. Most parks are mixed-use, catering to industries across a variety of sectors. However, a significant portion specializes in specific industries, particularly engineering, metals, food processing, textiles, chemicals, and automobiles. Results remain consistent even when sector-specific parks are excluded, with a slightly stronger treatment effect noted for mixed-use parks. The majority of parks are classified as Industrial Areas, Estates, or Parks, with a sizable portion labeled as "Others." Special Economic Zones (SEZs) and Export Promotion Zones (EPZs) are excluded from this sample due to their unique incentive structures. These zones constitute a small fraction of the overall distribution, and the results remain robust even when SEZs and EPZs are included. While most parks are publicly owned, recent years have seen an increase in private developments, which are included in this sample. Notably, the results are consistent when privately developed parks are excluded.

Table 1 presents detailed characteristics of the industrial zones. Some data is incomplete for specific zones, and potential issues of misreporting and measurement error could affect certain statistics. These characteristics were collected by the Department of Industrial Policy and Promotion from various development corporations, and there may be considerable variation in how the questions were interpreted across corporations. There is substantial variation in land sale and lease prices across industrial zones, with sale prices ranging from 100 rupees per acre to as high as 700,000 rupees per acre. The lower end of this range suggests that certain zones may offer land subsidies. Similar variability is observed in lease premiums. Industrial zones also vary widely in size, from as small as 1.7 hectares to as large as 500 hectares (approximately 5 square kilometers), with a median size around 30 hectares. Proximity to national highways is common among these zones, with a median distance of only 3 kilometers. This measure may reflect a survival bias, as proximity to infrastructure likely influences whether the proximity measure is reported. The number of vacant plots also varies greatly—while some parks are fully occupied, others have as many as 138 vacant plots.

Figure 17: Cross-sectional Distribution of Industrial Zones (June 2023)



Notes: This figure illustrates the distribution of 3,895 industrial zones across Indian States and Union Territories (top left), by sector (bottom left), by zone type (top right), and by ownership (bottom right). Of these, 1,584 parks have known establishment years, with 1,221 established post-1980, constituting the treatment group. IT zones focus on Information Technology and related services, while SEZ and EPZ indicate Special Economic Zones and Export Promotion Zones, respectively. Publicly owned zones include those managed by central public sector enterprises and public sector undertakings. "Joint" ownership refers to zones owned collaboratively by public and private sectors. Zones with missing information are excluded. Source: Department of Industrial Policy and Promotion, Government of India.

Table 1: Summary Statistics of Industrial Zone Characteristics

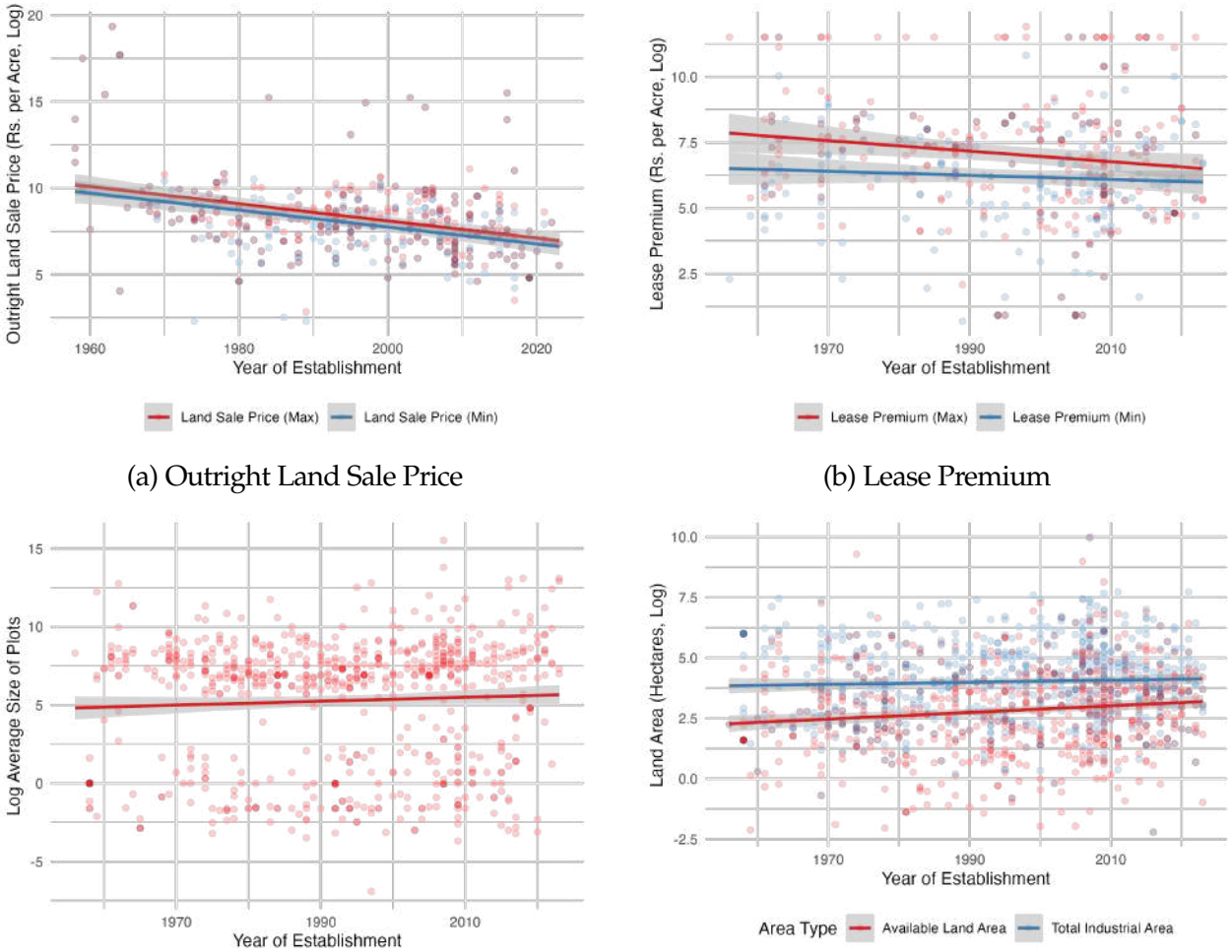
	Min	Median	Mean	Max	Observations
Outright Land Sale Price Range From	100.80	1995.00	5232.60	30610.00	1336
Outright Land Sale Price Range To	145.80	2600.00	10175.90	93189.90	975
Lease Premium From	1.00	504.10	46188.60	695658.20	810
Lease Premium To	2.00	968.00	359921.00	5427402.00	765
Average Size of Plots	0.086	727.00	2628.11	19806.66	1737
Total Industrial Area (ha.)	1.706	32.25	88.96	501.998	3113
Land Available (ha.)	0.32	12.30	38.99	237.18	1621
Distance from National Highway (km)	0.120	3.244	7.901	36.800	3455
Number of Plots	0.0	24.00	94.60	559.00	3879
Plots Allocated	0.0	2.00	61.67	399.25	3876
Vacant Plots	0.0	2.00	19.21	138.75	3876

Notes: All variables are winsorized at the 95 percent level. The zeros in Outright Land Sale Price Range From/To, Lease Premium From/To, Average Size of Plots, Total Industrial Area, and Distance from National Highways were recoded as N/A. All the prices are listed in Rupees per acre. All the variables are as of June 2023.

Figure 18 illustrates key characteristics of industrial zones as of June 2023, plotted against their years of establishment, focusing on zones with known establishment years. In panel (a), we observe that land sale prices in older, more established zones tend to be higher than those in newer zones, though this trend weakens when examining lease prices, as shown in panel (b). This difference may suggest that land ownership in established zones holds greater long-term value, while lease rates fluctuate more independently of the zone's age. Panel (c) displays the average plot size, revealing a bimodal distribution. This pattern suggests the existence of two categories of industrial zones: one with smaller plot sizes and another with larger plots. However, this may also be due to variations in reporting practices, where average plot sizes are either recorded in square kilometers or hectares. Additionally, there is a noticeable trend in more recent zones, where average plot size and total zone size tend to be larger, as shown in panel (d). This expansion likely reflects modern development trends favoring larger industrial areas to accommodate a broader range of industries. As expected, more recent zones also report a greater availability of land, reflecting the gradual uptake of land as zones become more established.

Although these descriptive statistics are not used directly in the main analysis due to potential reporting inconsistencies, they support the validity of the establishment year data. Despite some noise in other measures, the general relationship between zone characteristics and establishment year is logical and aligns with expected trends.

Figure 18: Industrial Zone Characteristics by Year of Establishment



(a) Outright Land Sale Price

(b) Lease Premium

(c) Average Plot Size

(d) Industrial Area

Notes: The figure plots the current characteristics of industrial zones against years of establishment for zones with known years of establishment. Figure (a) shows a scatter plot of the per acre minimum and maximum outright sale price (in rupees) of land in industrial zones, along with separate lines of best fit for the maximum and minimum prices. Figure (b) shows the lease price of land in industrial zones. Figure (c) shows the average size of plots (in hectares). Figure (d) shows the total land area and the land available to buy/lease separately.

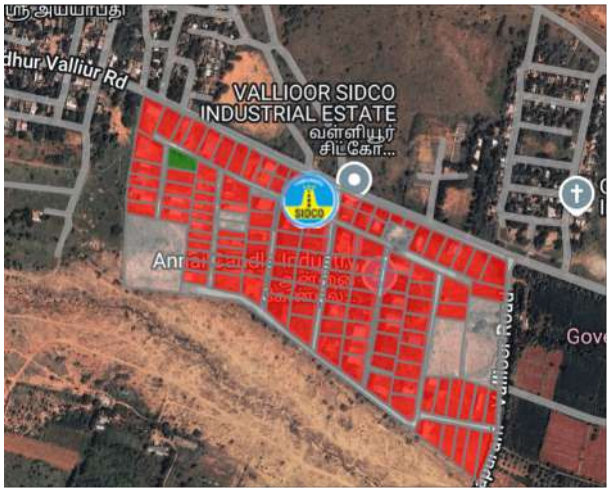
B.3 Industrial Zones Examples

Valliyoor Industrial Estate, Tamil Nadu: Located in the Tirunelveli district, Valliyoor Industrial Estate illustrates the inherent challenges faced by industrial zones in regions with low fundamentals, despite sustained government efforts. Established in 2005 by the Tamil Nadu Small Industries Development Corporation (SIDCO), the estate spans 39.91 acres but has struggled to attract substantial industrial activity. In fact, Valliyoor has the lowest treatment effect among the zones in my sample. A second phase was proposed in 2012, and in 2022, the government

announced further plans for a new 100-acre industrial estate nearby, targeted at small and medium enterprises.

Figure 19: Valliyur Industrial Estate (Announced in 2005)

(a) Position Relative to Tirunelveli



(b) Planned Layout

(c) May 2010 (Post Announcement)

(d) January 2024 (Post Announcement)

Geographically, Valliyoor is located around 50 kilometers from Tirunelveli, the nearest major town (see Figure 19a). The estate's distance from larger urban centers may contribute to its limited industrial uptake. Figure 19 illustrates the planned and actual development of the Estate. Contrasting the layout (Figure 19b) with the satellite images from 2010 (Figure 19c) and 2024 (Figure 19d) show limited infrastructure expansion and occupancy, indicating slow progress over nearly two decades.

The existing industrial landscape around Valliyoor includes establishments like Shalom Garments Pvt. Ltd., known for its Shelton brand of shirts and trousers. Established in 1999,

Shalom Garments has played a role in local employment and economic growth, contributing to the manufacturing sector's diversity in the region. Moreover, Valliyoor's proximity to Vadalivilai, the site of India's first 4.20-MW wind turbine generator installed in 2022, underscores its potential in renewable energy manufacturing.

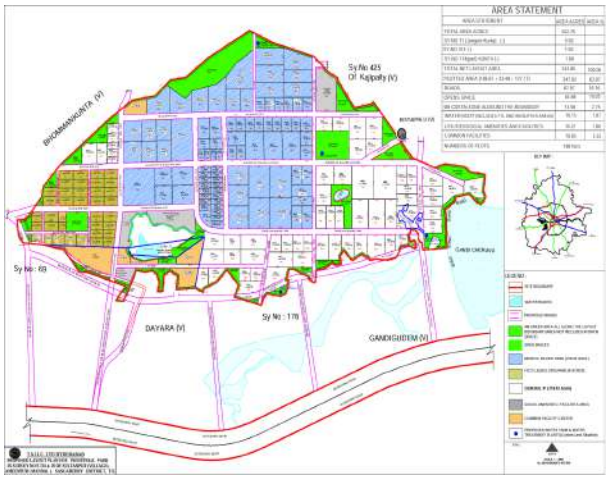
Sultanpur Medical Devices Park, Telangana: It is spread across 302 Acres or around 102 Hectares in the Rangareddy District of Telenaga, therefore it is on the larger side of the distribution. This park is part of Telanaga's Industrial Policy Initiatives. Telangana State identified medical devices and diagnostics as strategic growth sectors for Hyderabad. Since its launch, the park has received overwhelming response with more than 50 companies lining up to set up their manufacturing or research and development units here.

The location choice for the park is not random, but in part decided by the proximity to input markets and expertise from the large urban centre of Hyderabad. As shown in Figure 20a, the medical park is located on the eight-lane Hyderabad Outer Ring Road at Sultanpur, with a commute time of approximate 30 minutes (31km) to Hyderabad. Hyderabad has numerous small companies with expertise in plastics, precision engineering, electronics and others. All of these are vital suppliers for the medical device innovators and manufacturers. Hyderabad is also home to many precision engineering and plastic manufacturers capable of handling multiple types of materials. Also, availability of talent pool from academic and research and training centres also supports the medical devices segment. The park is around 40 minutes from the Indian Institute of Technology Hyderabad.

The proposed layout featured many basic amenities part of any typical urban planning: proximity to water source and water treatment plants, road connections to the highway network and common facility treatments.(see Figure 20b). Rest of the Figure 20 provides an overview of the park's development, starting from its pre-announcement phase in 2016 (Figure 20c) to its planned layout phase (Figures 20d, 20e) and the substantial infrastructure growth evident by October 2022 (Figure 20f).

Figure 20: Sultanpur Medical Devices Park (Announced in 2017)

(a) Position Relative to Hyderabad



(b) Planned Layout

(c) May 2016 (Pre Announcement)

(d) February 2020 (Post Announcement)

(e) May 2020 (Post Announcement)

(f) October 2022 (Post Announcement)

B.4 Treatment Definition and Propensity Score Matching

Treatment Definition. Municipalities are considered treated if (i) they receive an industrial zone, and (ii) any other industrial zone within the 25 km radius is established after that zone. Municipalities located within 25km of a zone with an unknown establishment year are excluded from the analysis. Control villages are defined as those situated outside a threshold distance (TT) from any treated village, with the baseline analysis using $TT = 25$ km. Robustness checks are conducted with various values of TT to confirm the robustness of this threshold.

Each treated village is linked to an Economic Census round (1982, 1990, 1998, 2005, 2013, 2021) and a Population Census round (1981, 1991, 2001, 2011, 2021) according to the year they receive their first industrial park. Note that we do not have outcomes corresponding to the first and rounds of the aforementioned censuses. When treatment occurs between two census rounds, the village is assigned to the latter census round. This implies that an increase of, say, 50% between two Economic Census rounds corresponds to an increase over about (7+8=) 14 to (7+8+8=) 24 years.

A key detail is that the first and last treated groups lack data on contemporaneous outcomes. However, these groups provide valuable insights: the last group serves as an additional check for pre-trends, and the first group offers insights into the long-term dynamics. These choices do not affect the structural analysis, as coefficients that use the first and last groups only are excluded from final estimations, and the other coefficients are not sensitive to the inclusion of these groups.

P-Score Matching. Treated units that received treatment between 1981 and 2021 are divided into four groups aligned with the population census rounds: 1991, 2001, 2011, and 2021. Matching is conducted separately for each group to account for shifts in policy regimes and to ensure that only baseline outcomes are used for matching. Since not all dynamic attributes are available for every treated unit, matching is separately performed for complete cases and non-complete cases. Non-complete cases are matched only on variables used to match units from 1981-1991.

Units treated between 1981 and 1991 are matched based solely on natural attributes (see Table 2 for details). For units treated between 1991 and 2001, economic attributes from the Economic Census 1990 are added to the matching criteria, specifically manufacturing employment (per sq km and log) and total non-farm Employment (per sq km and log). Units treated between 2001 and 2011 are further matched based on the 1998 economic census variables, distance to major population centers (populations above 10k, 50k, 100k, and 500k) in 1991, overall population (per sq km and log) in 1991, and proximity to highways in 1988, including whether a highway is within 25 km. For units treated between between 2011 and 2021, matching follows the criteria for 2001-2011 units, using corresponding baseline data from each period.

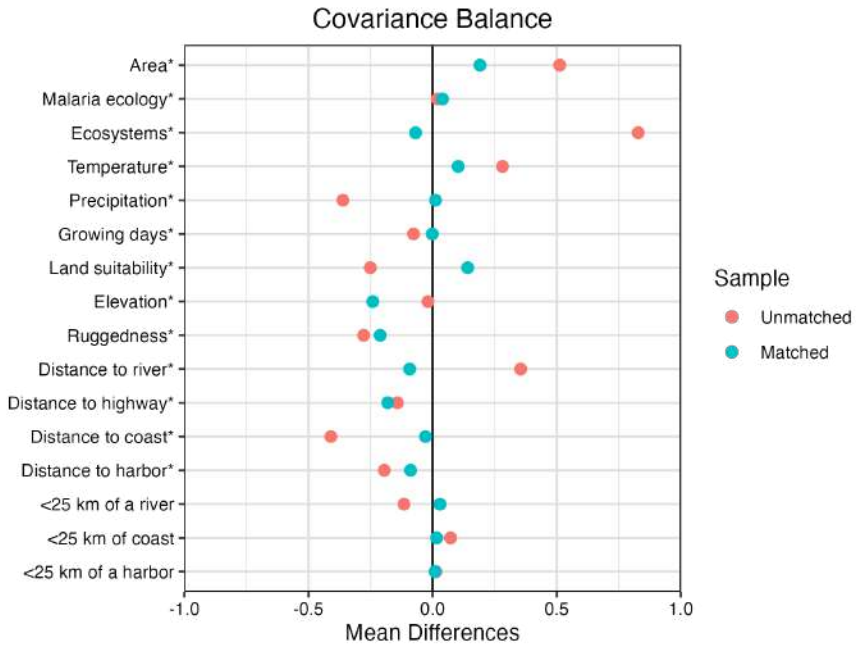
Variable Name	Description	Source
Malaria ecology	Index of the stability of malaria transmission, Kiszewski et al. (2004)	Henderson et al. (2018, QJE)
Ecosystems	Type of ecosystems	
Temperature	In Celsius (1960-1990 monthly average)	
Precipitation	mm/month (1960-1990 monthly average)	
Growing days	Length of growing period, days	
Land suitability	Probability that a cell is cultivated, Ramankutty (2002)	
Elevation	Mean elevation in location polygon (km)	SHRUG
Ruggedness	Mean TRI in polygon (000s of index)[60]	
Area	Units in sq. km, calculated with Lambert Azimuthal Equal Area Projection	
Distance to river	Units in km.	Author's calculations
Distance to coast	Units in km.	
<25 km of a river	1 (Distance to rivers $\leq$ 25 km)	
<25 km of a coast	1 (Distance to coastline $\leq$ 25 km)	

Table 2: Fixed Attributes

Variable Name	Description	Years; Source
Distance 10k	Distance (km) to nearest village with populatoon >10k	1991, 2001, 2011; Population Census
Distance 50k	Distance (km) to nearest village with population >50k	
Distance 100k	Distance (km) to nearest village with population >100k	
Distance 500k	Distance (km) to nearest village with population >500k	
Pop Density	Number of residents per sq km	
Pop Log	Number of residents, Log	
Non-Farm Emp Density	Number of non-farm workers per sq km	1990, 1998, 2001, 2013; Economic Census
Manuf Emp Density	Number of manufacturing workers per sq km	
Non-Farm Emp Log	Number of non-farm workers, Log	
Manuf Emp Log	Number of manufacturing workers, Log	
Distance to highway	Distance to nearest highway (km)	1988, 1996, 2004, 2011; Various Sources
<25 km of a highway	1 (Distance to highways $\leq$ 25 km)	

Table 3: Dynamic Attributes

Figure 21: Covariate Balance before and after P-Score Matching



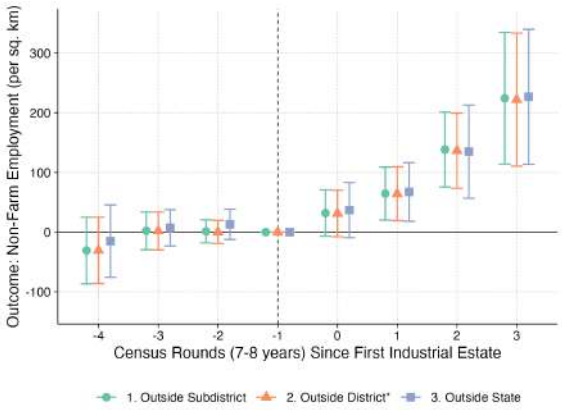
Notes: The figure depicts the difference in means for various attributes of villages/towns in India. The difference in means across the treated villages and all non-treated villages are depicted in red. Each treated village is matched with two control villages based on a series of village/town level characterisitcs. The difference in means across the treated villages and matched control villages is depicted in blue. Mean differences for continuous variables (marked with *) are normalized by the sample standard deviation of the treated group.

B.5 Robustness Checks

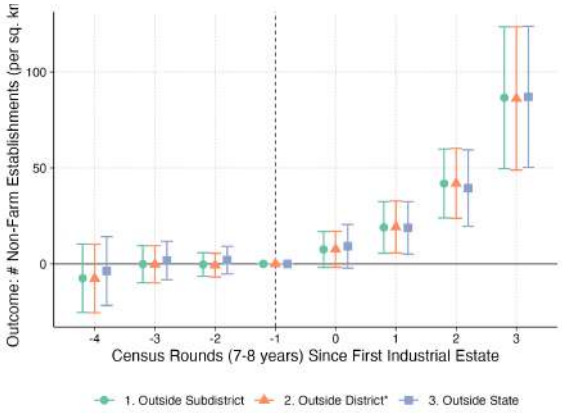
B.5.1 Alternative Propensity Score Matching Strategies

Selection of Control Villages Beyond "Subdistrict" or "State" Boundaries Instead of "District". I progressively restrict the control villages to be outside the boundaries of (i) the subdistrict, (ii) the district, and (iii) the state of the treated village. The results for the second restriction are presented as the baseline, as this boundary is often used to define labor markets in the Indian context (e.g., Rotemberg, AER). The results, shown in figure 22, demonstrate that the coefficients are robust to alternative geographic restrictions.

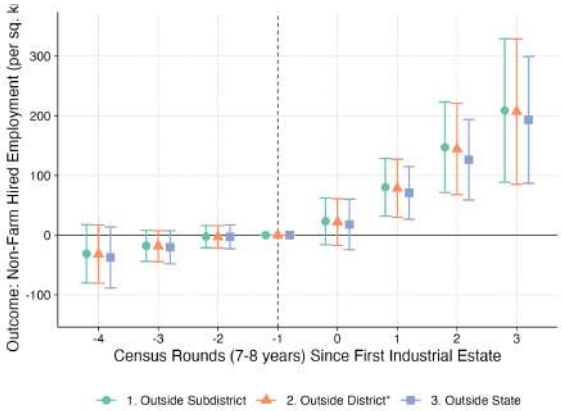
(a) Non-Farm Emp. Density (Total)



(b) Firm Density



(c) Non-Farm Emp. Density (Hired)



(d) Non-Farm Emp. Density (Unhired)

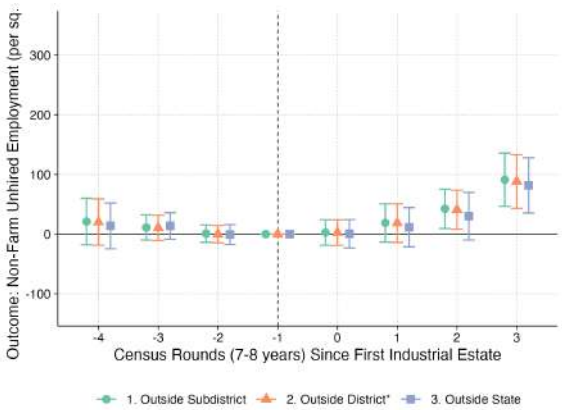
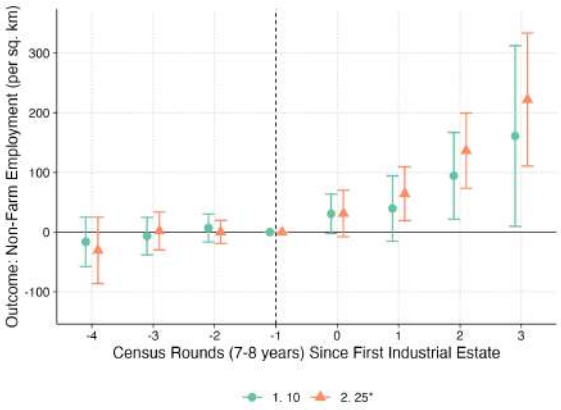


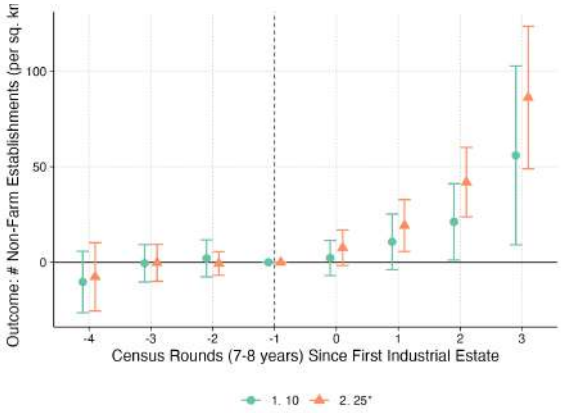
Figure 22: Different Geographical Restrictions on Controls: Administrative Boundaries

Control Villages Located Outside 10km Boundary Instead of 25 km. I restrict the control villages to be outside a (i) 10 km, and (ii) 25 km radius from the treated village. The results for the 25 km radius are presented as the baseline, as this boundary is also used to define the catchment area for Industrial Estates later in the spillover analyses. The results are reported in figure 23. Although not statistically distinguishable, the baseline coefficients tend to be slightly higher than the ones we get when we relax the boundary constraint. This is not surprising - as shown in Empirical Fact II, there are positive spillovers of the setting up of the Industrial Park up to 10-15 km.

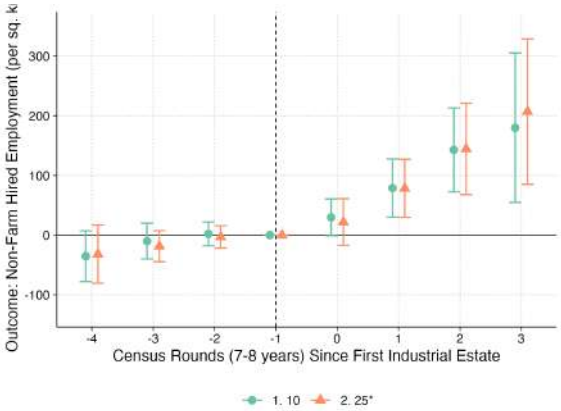
(a) Non-Farm Emp. Density (Total)



(b) Firm Density



(c) Non-Farm Emp. Density (Hired)



(d) Non-Farm Emp. Density (Unhired)

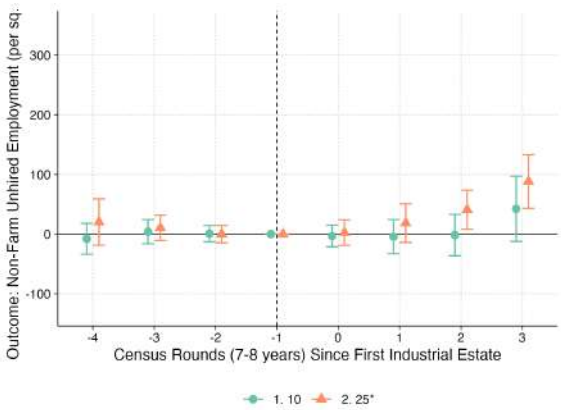
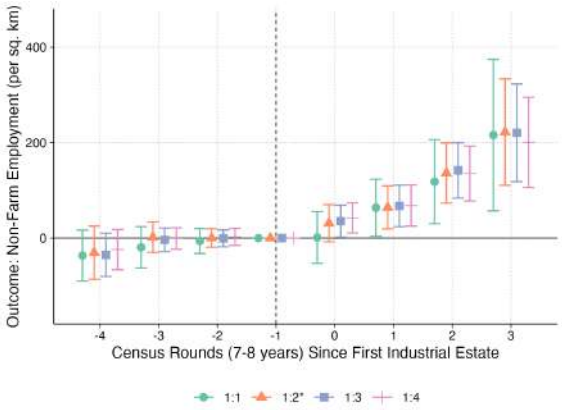


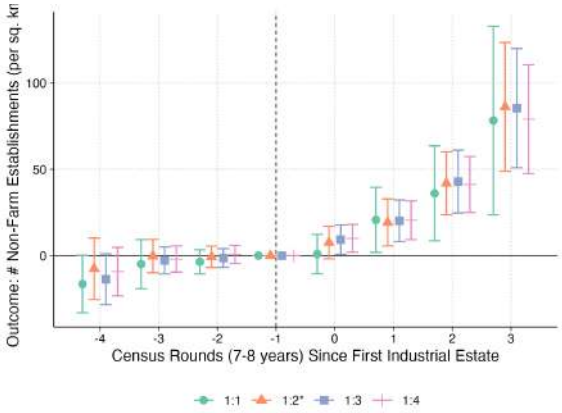
Figure 23: Different Geographical Restrictions on Controls: Distance from the Treated Unit

Selection of Three, or Four Control Villages per Treated Village Instead of Two. I progressively report results based on selecting 1, 2, 3, or 4 control villages for each treated village in figure 24. As expected, the precision of the estimates increases slightly in the number of control units, although this also increases computational demands in the structural estimation. Since the estimates do not change much, I selected two control units to contain the computational burden in the later steps.

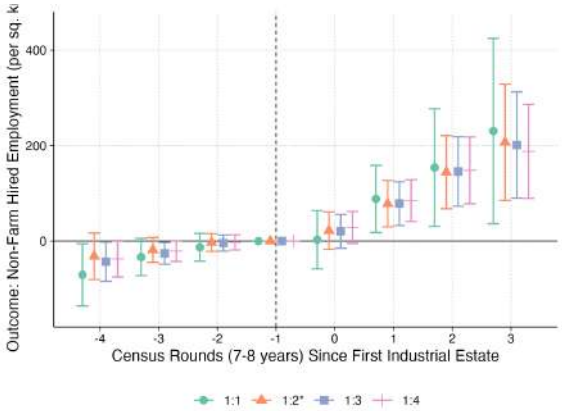
(a) Non-Farm Emp. Density (Total)



(b) Firm Density



(c) Non-Farm Emp. Density (Hired)



(d) Non-Farm Emp. Density (Unhired)

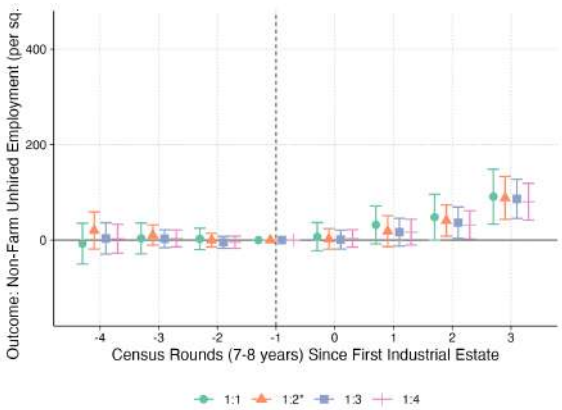


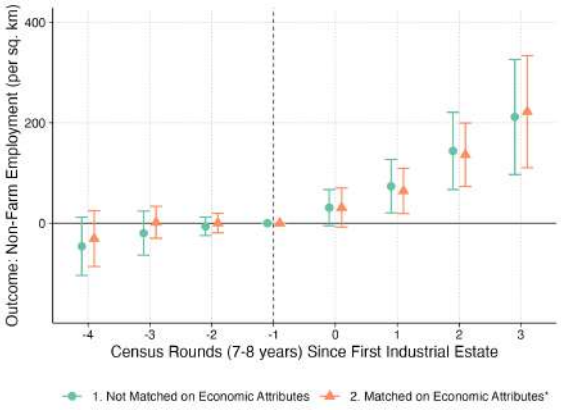
Figure 24: Different Number of Controls per Treated Unit

Propensity Score Construction Using Only Natural Attributes, Excluding Baseline Economic Attributes. In the baseline specification, treatment status is projected onto a set of natural and economic attributes (including some outcome variables), with different specifications for different treatment groups. The natural attributes are consistent across groups, while the economic attributes vary. For example, for the group treated between 2005 and 2013, I match treated units to control units based on economic attributes prior to 2005 while for the group treated between 1998 and 2005, I match units based on attributes prior to only 1998.

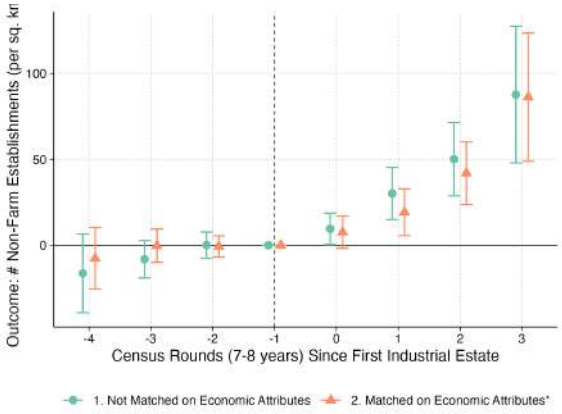
Although the inclusion of the economic attributes enhance comparability between treated and control units, I demonstrate that the lack of statistically significant pre-trends does not depend on their inclusion. I present results from an alternative matching strategy that relies solely on natural attributes in figure 25. The results demonstrate that any significant pre-trends are absent up to 3 rounds before the treatment, and while the coefficient for "round of treatment"

loses significance in some cases, the coefficient for "two census rounds after treatment"—the main long-run coefficient used in the structural estimation—remains robust.

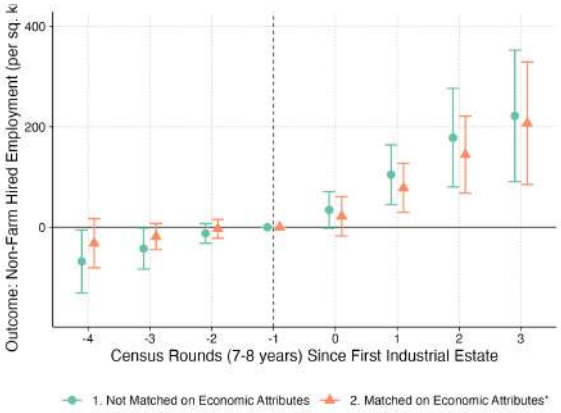
(a) Non-Farm Emp. Density (Total)



(b) Firm Density



(c) Non-Farm Emp. Density (Hired)



(d) Non-Farm Emp. Density (Unhired)

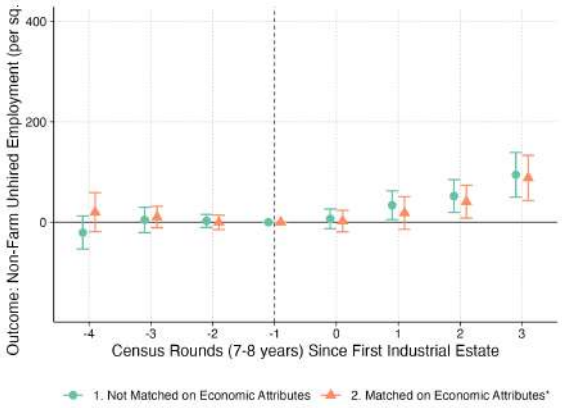
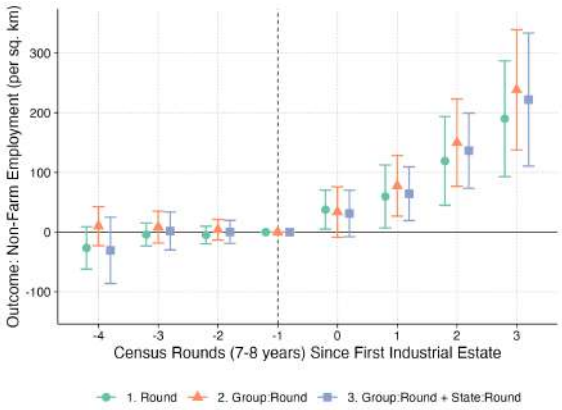


Figure 25: Different Matching Variables

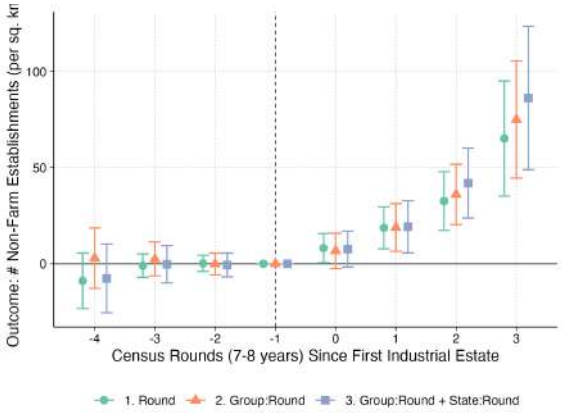
B.5.2 Alternative Model Specifications

Variation in Fixed Effects. In the baseline specification, I control for shocks that are common across all villages within a Propensity Score group and across all villages within a state. I demonstrate in figure 26 that the results are robust even when these fixed effects are excluded.

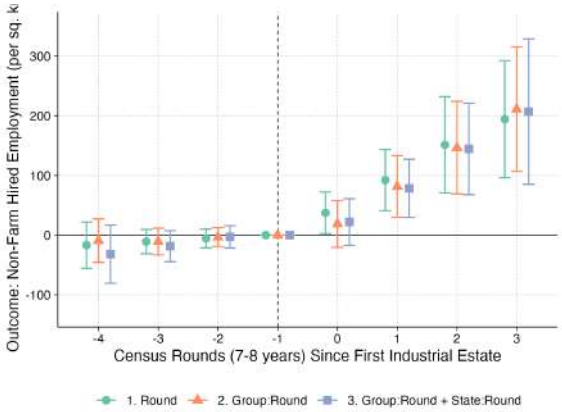
(a) Non-Farm Emp. Density (Total)



(b) Firm Density



(c) Non-Farm Emp. Density (Hired)



(d) Non-Farm Emp. Density (Unhired)

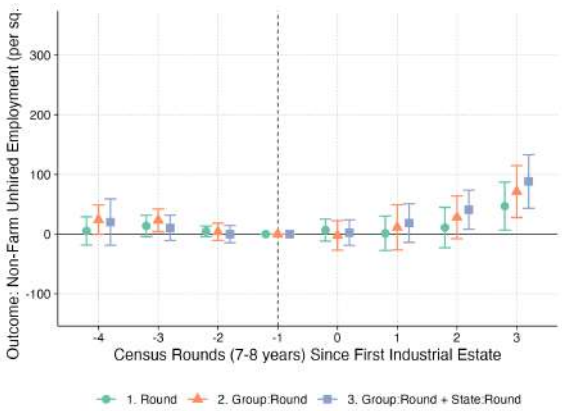
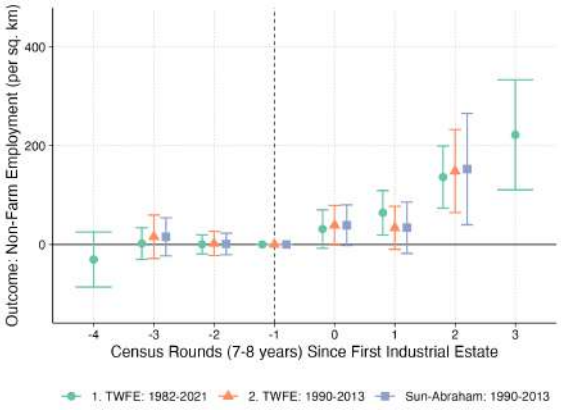


Figure 26: Different Fixed Effect Specifications

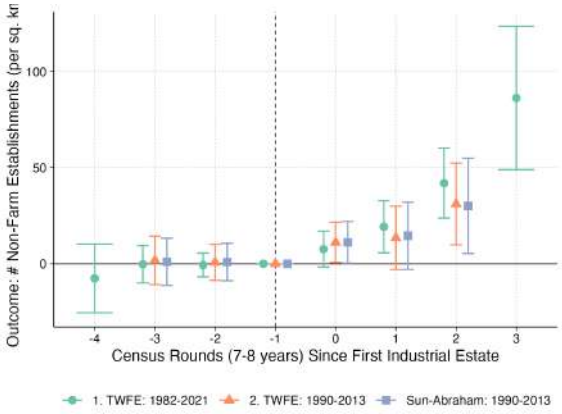
Addressing the 'Negative Weights' Problem. The baseline specification employs a Two-Way Fixed Effect (TWFE) strategy. Given the staggered nature of the treatment, this approach can be prone to the "negative weights" problem (Roth et al., 2023; De Chaisemartin and d'Haultfoeuille, 2023). However, this concern is mitigated in my analysis due to the presence of a large number of "pure" control units—those never treated. To illustrate, I use Sun and Abraham's Difference-in-Differences estimator (Sun and Abraham, 2021). The results are presented in Figure 27. Sun and Abraham's method only accommodates treatment groups for which contemporaneous outcomes are available. To facilitate a direct comparison with the TWFE results, I also present TWFE estimates that exclude treatment groups from pre-1990 and post-2013. The results demonstrate that although the standard errors are bit bigger, the point estimates are comparable to the TWFE estimates. Moreover, in most cases, the relevant long-run coefficient, the one corresponding to "two rounds after the treatment" remains statistically significant in most cases. I retain the

TWFE approach in the baseline because of its interpretability and due to its alignment with the structural framework equations.

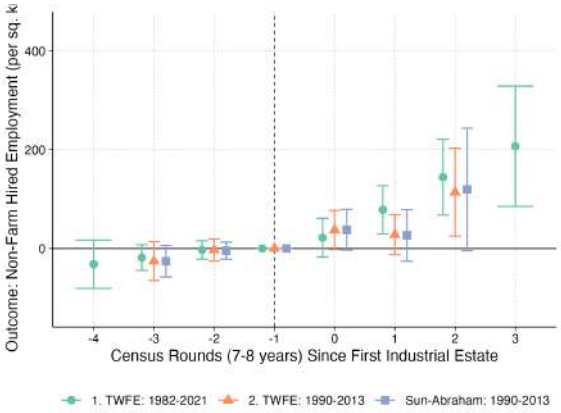
(a) Non-Farm Emp. Density (Total)



(b) Firm Density



(c) Non-Farm Emp. Density (Hired)



(d) Non-Farm Emp. Density (Unhired)

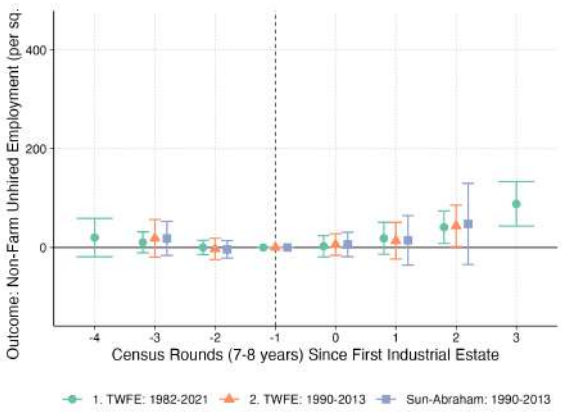


Figure 27: Different Empirical Specifications

B.5.3 Alternative Measures for Non-Traditional Activity

Figure illustrates the impact of industrial zones on various demographic and employment metrics. Panel (a) shows a notable increase in total population density, with approximately 500 additional people per square kilometer. This is accompanied by an increase in total employment, as depicted in panel (b), with around 200 additional workers per square kilometer. These patterns indicate that industrial parks are drawing in migrant workers, likely from nearby areas, to fill the job opportunities generated by the zones. In panel (c), there is a slight decrease in the share of farm sector employment within the municipalities hosting these zones, hinting at a mild shift from agricultural to modern sector employment. However, this does not

exclude the possibility of workers moving out of the agricultural sector in adjacent municipalities. Panel (d) shows a decline in the share of women in non-farm employment, a trend that aligns with evidence on shorter commuting preferences and migration frictions among women (Le Barbanchon et al., 2021; Petrongolo and Ronchi, 2020).

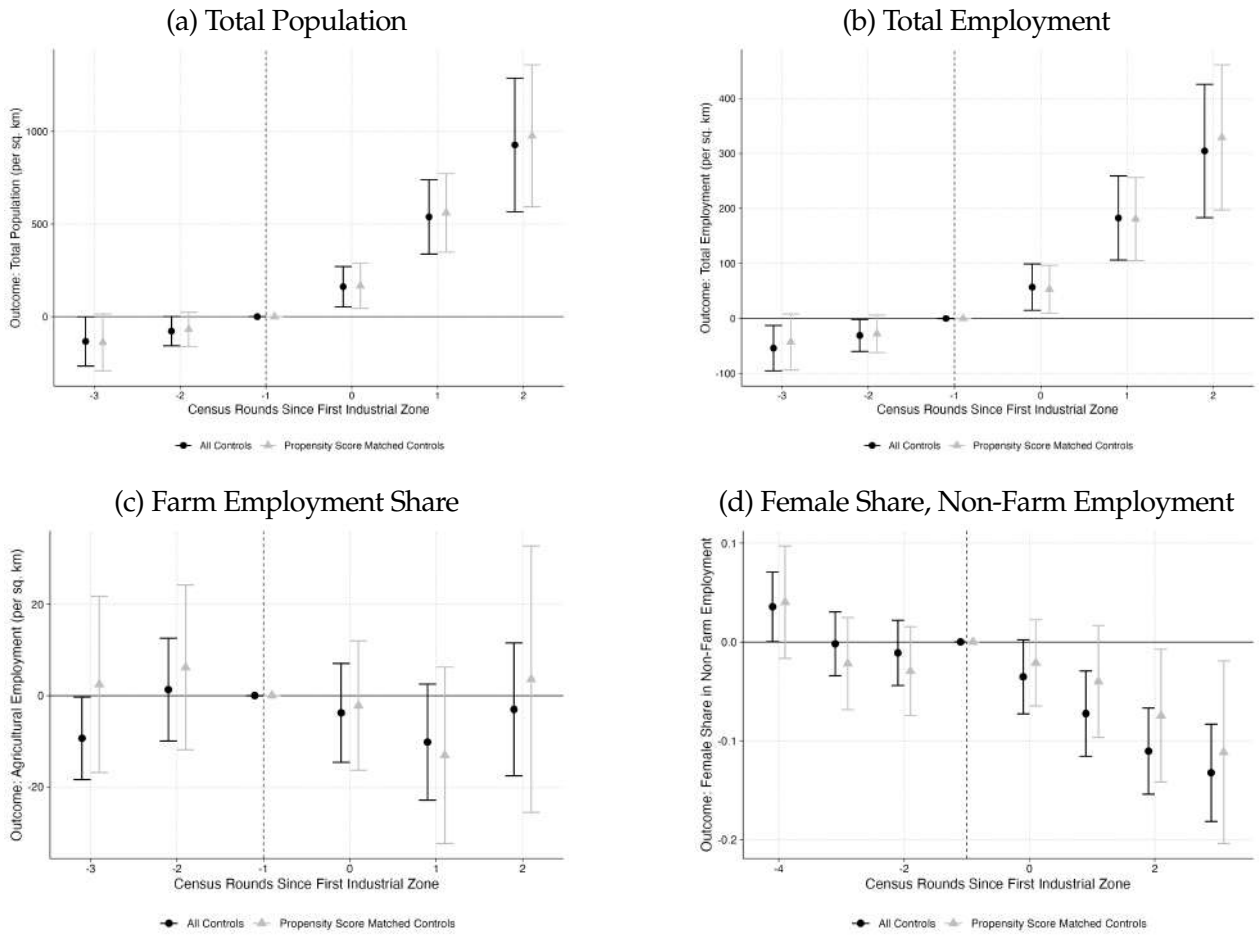


Figure 28: Additional Evidence of Industrialization

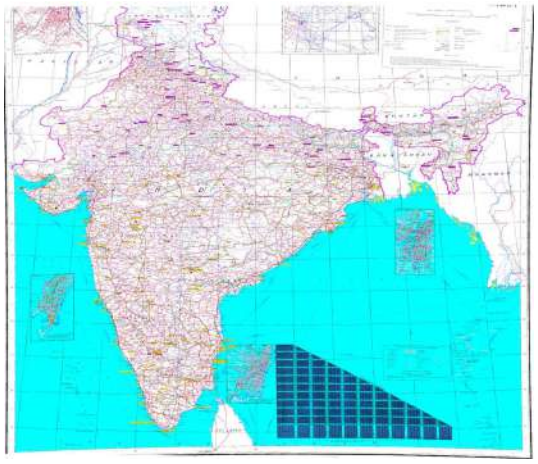
B.6 Construction of Historical Road Network Data

To accurately capture India's road network—including highways, medium roads, and minor roads—from map images (Figure 29), I led a research team to implement a comprehensive image processing workflow (described below) that integrates both MATLAB and Python tools for maximum precision. Th extracted road network is depicted in Figures 30 and 31.

(a) 1988

(b) 1996

(c) 2004



(d) 2011

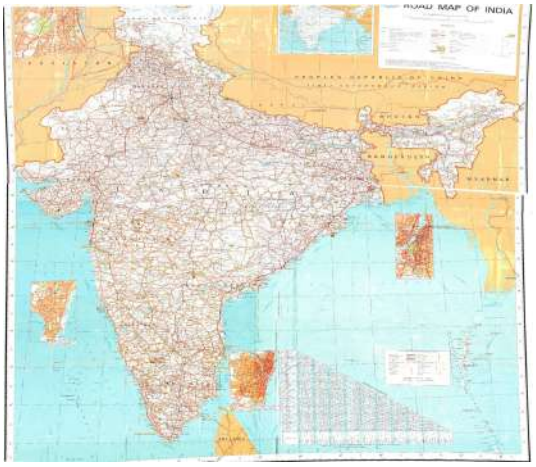


Figure 29: Historical Maps of India's Road Network.

Removal of Extraneous Features (Oceans) and Border Extraction

To ensure that only land areas were processed, we began by removing ocean regions. This step utilized color masking techniques specifically targeting cyan pixels to identify ocean areas, which were then removed from the image. After isolating the ocean regions, a morphological dilation was applied to the mask to ensure continuity along the coastlines. Next, we extracted India's national borders by creating a color mask for relevant RGB values indicative of the boundary areas. Morphological dilation was applied to further strengthen border continuity, followed by flood-filling operations to fill enclosed areas and finalize a precise land mask that accurately outlined India's territory.

Road Network Isolation and Masking

With the borders isolated, we turned to road extraction. Using the RGB values associated with each road type (highways, medium, and minor roads), we established defined tolerances to accommodate gradient variations in the maps. These tolerances were applied to create separate masks for each road category. The RGB values were refined by sampling approximately 25 points along each road type to set the maximum, minimum, and differential thresholds for each channel. Logical masks filtered the pixels matching these criteria, effectively isolating each road type in separate layers.

To refine the isolated pixels, we applied Gaussian and disk filters, which reduced noise. Morphological operations like dilation and erosion were crucial here: dilation connected any nearby road segments that were initially missed during extraction, while erosion restored the roads to their original widths. This ensured that road dimensions were preserved accurately and created a clean and continuous representation of the road network. Next, we used the skeletonization function from scikit-image, which reduced the road masks to their central lines, allowing for a more efficient contouring of the roads. Small, insignificant clusters outside India's borders were subsequently removed to ensure a cleaner road network restricted within the country's boundaries.

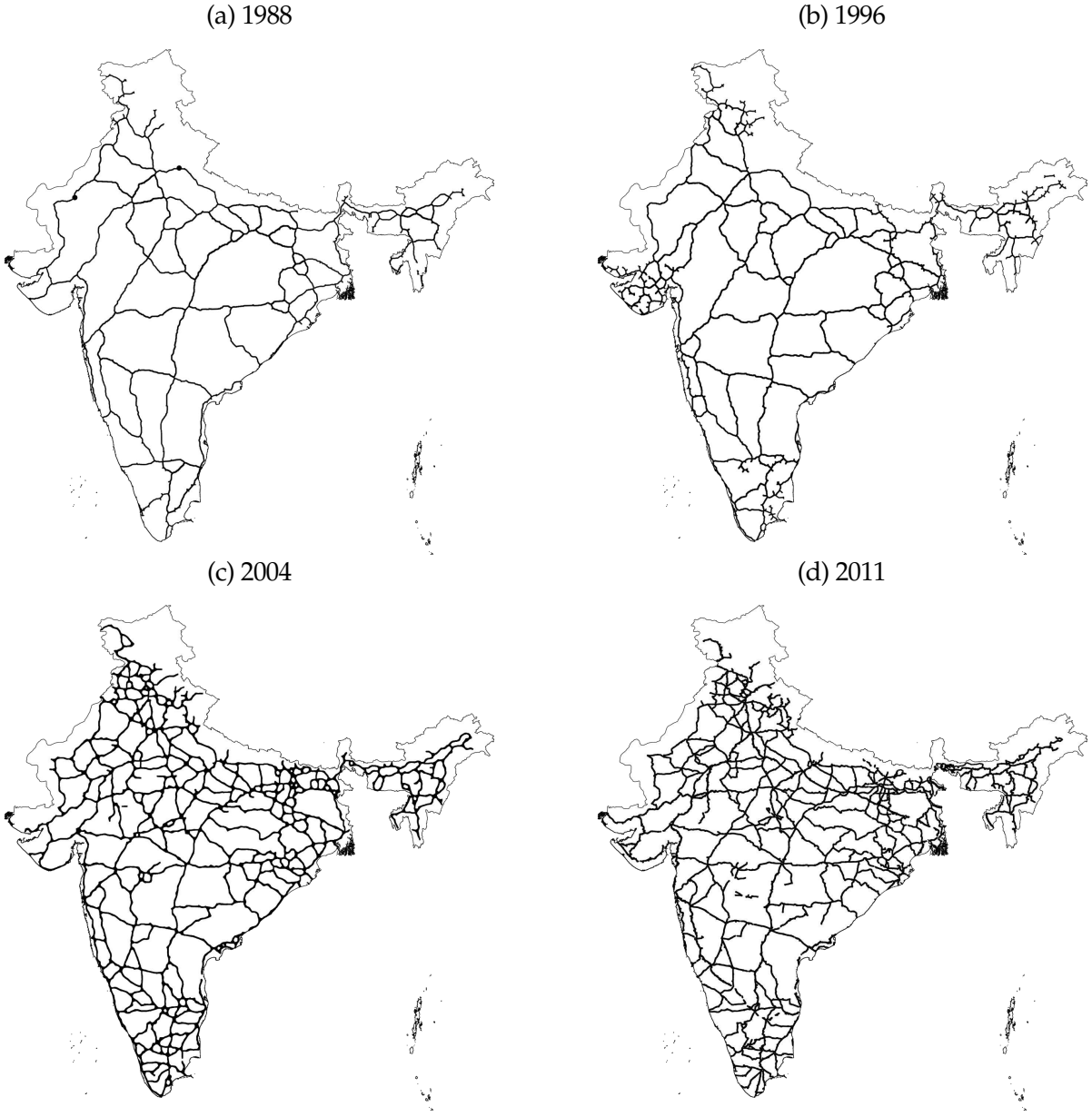
Shapefile Generation

In Python, contours within the processed road masks were converted into `LineString` objects using Shapely to represent road geometries accurately. These line geometries were compiled into a GeoDataFrame using GeoPandas, saved as shapefiles, and subjected to a final spatial check. Spatial indexing techniques were applied to filter out roads that extended beyond the defined borders. This step allowed us to combine the final road network and border layers into a single shapefile, preserving spatial alignment and ensuring the road network fit seamlessly within India's national boundaries.

Geocoding and Integration

Finally, the generated shapefiles were overlaid onto OpenStreetMap in QGIS, using India's borders to georeference the road network accurately. This alignment resulted in a geocoded shapefile suitable for spatial analyses, ensuring that each road segment was correctly mapped in its geographic location. As a last step, we designed the pipeline to be reusable for processing road map images across different years. Given that RGB values for roads vary across map editions, the workflow was adjusted by reducing or increasing color tolerances as needed, allowing for consistent extraction despite color variations.

Figure 30: Historical Network of Highways



Notes: The figure depicts the network of highways extracted from historical maps of the Indian road network for the years 1988, 1996, 2004, and 2011. These networks were digitized through an image-processing workflow that utilized color masking to identify roads based on their RGB values, followed by Gaussian filtering and morphological operations to enhance clarity. Each road category was isolated, skeletonized to reduce to central lines, and refined with spatial indexing to remove extraneous elements. The final shapefiles were georeferenced and layered onto India's national borders to ensure accurate spatial alignment across different years.

C Appendix: Estimation

Figure 31: Historical Network of Other Roads

Notes: The figure depicts the network of roads (all weather motorable and motorable under fair weather) extracted from historical maps of the Indian road network for the years 1988, 1996, 2004, and 2011. These networks were digitized through an image-processing workflow that utilized color masking to identify roads based on their RGB values, followed by Gaussian filtering and morphological operations to enhance clarity. Each road category was isolated, skeletonized to reduce to central lines, and refined with spatial indexing to remove extraneous elements. The final shapefiles were georeferenced and layered onto India's national borders to ensure accurate spatial alignment across different years.

C.1 Derivation of Equations 10 - 11

Denote $y_i = \frac{L_{i,mt}}{\sum_i L_{i,mt}}$. With some abuse of notation, lets use index i to index a location-sector pair. Lets denote with $n = IK - 1$ the dimension of the domian of $\mathbf{y}$. From Equation 6 and 2 , we have

$$\ln y_i - \ln y_0 = \rho \ln \nu_i = \rho(\ln w_i + \ln \eta_i)$$

From eq (8), we have

$$\ln w_i = \ln A_i - \alpha(\ln y_i - \ln y_0)$$

Substituting from Equation 8 we get

$$\ln y_i - \ln y_0 = \rho \ln \nu_i = \rho(\ln A_i - \alpha(\ln y_i - \ln y_0) + \ln \eta_i)$$

which can be further written as

$$\ln y_i - \ln y_0 = \frac{\rho}{1 - \rho\alpha}(\ln A_i + \ln \eta_i)$$

Substituting $\ln A_i$ from Equation 9 we have

$$\ln y_i - \ln y_0 = \frac{\rho}{1 - \rho\alpha}(g_i(\mathbf{y}) + \ln a_i + \ln \eta_i)$$

which leads to Equations 10 - 11 which is only in terms of δ, η, y_i, s_i:

$$\ln y_i - \ln\left(1 - \sum_i y_i\right) = \tilde{g}_i(\mathbf{y}; \delta, \eta) + s_i$$

C.2 Verifying assumption 1 for Equations 10 - 11

The system of Equations 10 - 11 can be described as

$$J_i \equiv \ln y_i - \sum_j M_{ij} y_j - \ln\left(1 - \sum_j y_j\right) - s_i = 0$$

Proposition. *System of equations 10 - 11 satisfies Assumption 1*

First, we know that $\mathbf{y}$ lie in a compact and convex space since it denotes employment shares. It is also easy to verify that any vector $\mathbf{y}$ that satisfies the above system of equations must lie strictly in the interior of the unit simplex. To verify the transversality condition, note that the Jacobian of the system with respect to $\mathbf{y}$ and $\mathbf{s}$ is given by the $n \times 2n$ dimensional matrix

$$J_{\mathbf{y},\mathbf{s}} = [J_{\mathbf{y}}\ I]$$

which is always full rank, irrespective of the values of $J_{\mathbf{y}}$.

C.3 Verifying the hypothesis of Proposition 3 for Equations 10 - 11

Lemma. $\frac{\partial det(J_{\boldsymbol{y}}(\boldsymbol{y},\boldsymbol{s}))}{\partial \boldsymbol{y}} \neq \boldsymbol{0}$ *for all irregular solutions* $(\boldsymbol{y},\boldsymbol{s})$ *implies that the system of equations 10 - 11 satisfies the assumption in Proposition 3*

Proof. To check this assumption, we need to examine the rank of the matrix

$$\begin{bmatrix} J_{\mathbf{y}}(\mathbf{y},\mathbf{s}) & I \\ \frac{\partial det(J_{\mathbf{y}}(\mathbf{y},\mathbf{s}))}{\partial \mathbf{y}} & \frac{\partial det(J_{\mathbf{y}}(\mathbf{y},\mathbf{s}))}{\partial \mathbf{s}} \end{bmatrix}$$

for all $\mathbf{y},\mathbf{s}$ for which $det(J_{\mathbf{y}}(\mathbf{y},\mathbf{s}))=0$. Note that $\frac{\partial det(J_{\mathbf{y}}(\mathbf{y},\mathbf{s}))}{\partial \mathbf{s}}=0$, and the $n\times 2n$ matrix $[J_{\mathbf{y}}\ I]$ is full rank. Hence, we need to examine whether the row $[\frac{\partial det(J_{\mathbf{y}})}{\partial \mathbf{y}} \frac{\partial det(J_{\mathbf{y}})}{\partial \mathbf{s}}]$ is in the span of $[J_{\mathbf{y}}\ I]$. That is, we ask, whether there exists an $\alpha \in R^n$ such that

$$\alpha * [J_{\mathbf{y}}\ I] = [\frac{\partial det(J_{\mathbf{y}})}{\partial \mathbf{y}}\ \mathbf{0}].$$

This equation implies that $\alpha=0$, which implies that $\frac{\partial det(J_{\mathbf{y}})}{\partial \mathbf{y}}=\mathbf{0}$, completing the proof. □

$\frac{\partial det(J_{\mathbf{y}})}{\partial \mathbf{y}} \neq \mathbf{0}$ is generically true can be easily verified in the case with two locations. With two locations, we have

$$J_{\mathbf{y}} = \begin{bmatrix} \frac{1}{y_1} + \frac{1}{1-\sum_j y_j} - \delta & \frac{1}{1-\sum_j y_j} - \delta' \\ \frac{1}{1-\sum_j y_j} - \delta' & \frac{1}{y_2} + \frac{1}{1-\sum_j y_j} - \delta \end{bmatrix}$$

where $\delta = M_{11} = M_{22}$ and $\delta' < \delta = M_{12} = M_{21}$. For a 2×2 matrix, it is easy to write the expression for the determinant:

$$det(J_{\mathbf{y}}) = (\frac{1}{y_1} + \frac{1}{1-\sum_j y_j} - \delta)(\frac{1}{y_2} + \frac{1}{1-\sum_j y_j} - \delta) - (\frac{1}{1-\sum_j y_j} - \delta')^2$$

Let's consider the cases under which this is zero.
Case 1. $\left(\frac{1}{y_i} + \frac{1}{1-\sum_j y_j} - \delta\right) = 0$ for some i. This implies $y_0 = \frac{1}{\delta'}$, $y_i = \delta - \delta'$ and $y_{-i} = 1 - \frac{1}{\delta'} - (\delta - \delta')$
Case 2. $\left(\frac{1}{y_i} + \frac{1}{1-\sum_j y_j} - \delta\right) \neq 0$ for all i, which implies that $\Pi_i \left(\frac{1}{y_i} + \frac{1}{1-\sum_j y_j} - \delta\right) > 0$
Differential further with respect to $\mathbf{y}$, we get

$$\frac{\partial det(J_{\mathbf{y}})}{\partial y_i} = \left(\frac{1}{y_{-i}} + \frac{1}{1-\sum_j y_j} - \delta\right)\left(\frac{-1}{y_i^2} + \frac{1}{(1-\sum_j y_j)^2}\right) \tag{15}$$

Let's consider the cases under which this is zero.

Case 1. $\left(\frac{1}{y_i} + \frac{1}{1-\sum_j y_j} - \delta\right) \neq 0$ for all i. Then only way for 15 to be zero for all i is for $y_i = \frac{1}{n+1}$, which is true (from the zero determinant condition) only if

$$\frac{2}{n} - \delta = \frac{1}{n} - \delta' \neq 0$$

Case 2. $\left(\frac{1}{y_i} + \frac{1}{1-\sum_j y_j} - \delta\right) = 0$ for some i which we saw implies $y_0 = \frac{1}{\delta'}$, $y_i = \delta - \delta'$ and $y_{-i} = 1 - \frac{1}{\delta'} - (\delta - \delta')$.

Subcase 1. $\left(\frac{1}{y_{-i}} + \frac{1}{1-\sum_j y_j} - \delta\right) \neq 0$. This implies $y_{-i} = y_0 \neq y_i$ which is true only if

$$1 - \frac{1}{\delta'} - (\delta - \delta') = \frac{1}{\delta'} \neq \delta - \delta'$$

Subcase 2. $\left(\frac{1}{y_{-i}} + \frac{1}{1-\sum_j y_j} - \delta\right) = 0$. This implies that $y_i = y_{-i}$ which is true only if

$$0.5 - \frac{0.5}{\delta'} = \delta - \delta'$$