Identification and Estimation of Panel Regression Models with Unknown Aggregated Spatial Units*

Hideyuki Tomiyama
Department of Economics
The Pennsylvania State University
hkt5140@psu.edu

This version: October 25, 2026

Click here for the latest version

Abstract

This paper studies panel regression models in which each unit corresponds to an aggregated spatial unit. Each aggregated unit is a collection of multiple finer units in the raw dataset. The variables of each aggregated unit can be constructed as the sum of finer unit-level variables in the aggregated unit. However, the researcher does not know the correct delineation of aggregated units under which the structural equation holds. Examples include regression models of agglomeration economies and game theoretic models of spatial entry competitions. We show that the derivative of the structural function can be nonparametrically identified from observational panel data on finer units under certain conditions. The key identification condition is mean independence between the structural error terms and the vector of finer unit-level exogenous variables. The correct delineation can be also identified if the structural function is nonaffine with respect to the regressors. We propose several computationally feasible inference procedures for parametric models via GMM-type objective functions. As an empirical illustration, we apply the methodology to the economics of agglomeration using U.S. data.

*First version: October 25, 2026. I am grateful to Marc Henry, Keisuke Hirano, and Andres Aradillas-Lopez for their helpful discussions.

1 Introduction

Statistical inference based on many aggregated spatial units prevails in a wide range of fields in economics. For instance, in empirical industrial organization, researchers often construct many spatial markets based on political jurisdictions to estimate firms' payoff functions and conduct some welfare analysis (e.g., Bresnahan and Reiss, 1991; Ryan, 2012; Collard-Wexler, 2013). However, economic markets are usually not observed in data and such political jurisdictions rarely reflect them (see Berry et al., 2019, for instance). Also, in urban economics, spatial economics, and economics of geography, researchers need to define spatial units in order to estimate the impact of the economics of density, agglomeration effects, and knowledge spillover. However, in many cases, there is no consensus on how to delineate spatial units, which is known as modifiable areal units problem (MAUP) in the literature (e.g., Briant et al., 2010; Duranton, 2021). It is important to identify the correct delineation since the estimates of the structural function are not reliable with an incorrect delineation.

It is also common to have raw data that is finer than the aggregate data used in statistical models. For instance, when researchers study policies or issues in labor markets, aggregate data is often constructed based on local labor markets from finer municipality-level raw data. In the context of the United States, commuting zones (CZ) proposed by Tolbert and Killian, 1987 and Tolbert and Sizer, 1996 are often used as the units of local labor markets.[1] Thus, the CZ-level dataset is constructed from the county-level dataset. Also, in empirical industrial organization, market-level dataset is often constructed from detailed information on each firm's exact location of entry (i.e., longitude and latitude) when researchers study geographical competitions among plants, retailers, and among others.

This paper studies identification and inference of a panel regression model where each unit corresponds to an aggregated spatial unit. We consider a scenario in which a researcher observes data for each of finer spatial units (e.g., counties or boroughs) and constructs a sample by partitioning the set of finer units into many aggregated units (e.g., states, MSAs, or CZs). The variables of each aggregated unit can be constructed as the sum of finer unit-level variables in the aggregated unit. However, the researcher does not know the correct delineation of aggregated units under which the structural equation holds.

We show that the derivative of the structural function (i.e., the partial effect of regressors) can

[1] Examples include but not limited to Autor et al., 2013 and Acemoglu and Restrepo, 2020.

be nonparametrically identified from observational panel data on finer spatial units under certain conditions. In particular, we consider a setup in which the variables of each aggregated unit can be constructed as the sum of finer unit-level variables in the aggregated unit. This aggregation formula is satisfied in many occasions such as when population and the number of manufactures are considered as aggregated variables. Another key condition is the mean independence between the structural error terms and the vector of finer unit-level exogenous variables. The mean independence condition is stronger than the conventional mean independence condition between the structural error terms and aggregated exogenous variables, which is often assumed when the correct delineation is known. Thus, the stronger exogeneity condition provides additional conditional moment restrictions, which identifies the derivative of the structural function even though the true delineation is unknown.

The correct delineation may not be identified in general. However, we show that it can also be identified with an additional assumption that the structural function is nonaffine with respect to the regressors. In other words, the conditional moment restrictions based on the stronger exogeneity condition cannot be satisfied with any incorrect delineation under the nonaffinity assumption, and hence the correct delineation can be identified.

We propose several inference procedures for parametric models based on the identification result. As stated above, the mean independence condition between the structural error terms and the vector of finer unit-level exogenous variables gives conditional moment restrictions. Thus, the vector of finer unit-level exogenous variables serves as instrumental variables. Then we can construct an estimator by transforming conditional moment restrictions to unconditional ones without losing information for the identification. There are many papers on these methodologies in both parametric models (e.g., Bierens, 1982; Bierens, 1990; Hong and White, 1995; Bierens and Ploberger, 1997; Domínguez and Lobato, 2004) and nonparametric models (e.g., Newey and Powell, 2003; Horowitz, 2011; Chen and Christensen, 2018; Chen et al., 2025). Under some additional regularity conditions, we show the consistency and asymptotic normality of our proposed estimator as the number of time periods increases.

This paper relates to the literature on ecological fallacy (e.g., Robinson, 1950; Duncan and Davis, 1953; King, 1997). The literature considers a similar problem that aggregate variables can exhibit different relationships from the same individual-level variables, depending on the level of aggregation. Econometricians have studied the problem through the lens of data combination. Examples include but are not limited to Cross and Manski, 2002, Molinari and Peski, 2006, Ridder and Moffitt, 2007,

Fan et al., 2014, Fan et al., 2016, Buchinsky et al., 2022, and Moon, 2024. The prime interest of the literature is parameters which determine the individual-level response. Our model is different in that we are interested in structural functions of which arguments are aggregated variables and the structural equation holds with some unknown level of aggregation. This is an empirically relevant scenario in many fields of economics, especially when aggregated outcome or equilibrium outcomes in local markets are of interest (e.g., spatial competition among firms in industrial organization, agglomeration effects in urban economics, and so on).

This paper also contributes to the vast literature on MAUP in statistics, geography, and economics. For example, Gehlke and Biehl, 1934 is a seminal work that shows that statistical results can vary greatly depending on the delimitation chosen by the researcher. Briant et al., 2010 investigates the magnitude of the distortions resulting from the choice of aggregated spatial units in the context of spatial concentration, economics of agglomeration, and gravity equations. For a more detailed review on MAUP, see Briant et al., 2010 and the reference therein. To the best of my knowledge, this is the first paper to consider the identification and inference of both the correct delineation and structural function from observational panel data on finer spatial units. In particular, we show that the correct delineation can be identified from the nonaffinity of the structural function.

The role of the nonaffinity assumption is related to the classical literature on aggregation (e.g., Gorman, 1953; Theil, 1954; see Blundell and Stoker, 2005 for a review). This literature considers when the structural equation continues to hold after aggregation, and shows that it does only when the structural function is affine. We instead show that the nonaffinity of the structural function identifies how the finer units are aggregated, which is unknown to the researcher.

Some recent papers consider the identification of network structures from observational panel data in various setups. For instance, Manresa, 2013, Lam and Souza, 2016, 2020, and de Paula et al., 2024 study the identification problem in spatial regression and social interaction models. Our model has some similarities with them because we aim to detect the structure of connections between finer spatial units from observational panel data. However, this paper is different in that we are interested in the relationship between endogenous and exogenous variables within each aggregated unit and spillover effects across aggregated units are ruled out. Such specifications are common in many applied fields due to the interpretability and tractability of the models.

As an empirical illustration, we apply the methodology to the economics of agglomeration using U.S. data. We estimate the agglomeration parameter under three delineations: the finest (treating

each CZ as a separate unit), the coarsest (aggregating all CZs within each state), and our data-driven estimated delineation. The results show that the estimated delineation aggregates adjacent CZs in most states, although it is still much finer than the state-level delineation (i.e., each state is an aggregated unit). Compared with the estimate under the estimated delineation, the estimates under both the CZ-level and the state-level delineations underestimate the agglomeration effect.

The rest of the paper is organized as follows. Section 2 provides the setup of the model and nonparametric identification results. In Section 3, we consider inference of parametric regression models. Section 4 presents simulation results. In Section 5, we apply the methodology to the economics of agglomeration with US data. Section 6 concludes.

2 Nonparametric identification

2.1 Setup

We consider a pooled panel nonparametric regression model

$$Y_{it} = h(X_{it}) + \varepsilon_{it} \quad \text{and} \quad \mathbb{E}\left[\varepsilon_{it} \mid X_{it}\right] = 0,$$

where $h : \mathcal{X} \to \mathbb{R}$ is a unknown structural function, $Y_{it} \in \mathcal{Y} \subseteq \mathbb{R}$ is a dependent variable, $X_{it} \in \mathcal{X} \subseteq \mathbb{R}^{d_x}$ is a vector of exogenous regressors, and i and t correspond to an aggregated spatial unit and an instance (e.g., time periods) respectively. The model can be extended to a fixed effect model of the form; $Y_{it} = \alpha_i + h\left(X_{it}\right) + \varepsilon_{it}$. We discuss further details below.

We cannot directly observe (Y_{it}, X_{it}). Instead, many finer spatial units are observed. Henceforth, we call these finer units as blocks. Let $\mathcal{J} := \{1, ..., J\}$ be the set of blocks. For each block j, we observe a block-level dependent variable W_{jt} and a vector of block-level exogenous variables Z_{jt}. Let $\mathcal{W}_j \subseteq \mathbb{R}$ and $\mathcal{Z}_j \subseteq \mathbb{R}^{d_x}$ be the support of W_{jt} and Z_{jt} respectively. Let $W_t := \{W_{jt}\}_{j\in\mathcal{J}}$ and $Z_t := \{Z_{jt}\}_{j\in\mathcal{J}}$. We use γ_j to denote the parameter indicating the aggregated spatial unit to which block j belongs. Notice that the labels of each aggregated unit cannot be identified. Thus, $\gamma = \tilde{\gamma}$ means that γ and $\tilde{\gamma}$ coincide given a suitable choice of the labels.[2] Let $\mathcal{N}_\gamma$ be the set of aggregated units under γ. It is worth to mention that γ implicitly determines the number of aggregated units. Throughout this paper, the number of blocks J is fixed. We consider the asymptotics in which the number of periods T increases.

[2]Put differently, we consider that $\gamma = \tilde{\gamma}$ if these two parameters induce the same partition of $\mathcal{J}$.

Let X_{it}^{γ} and Y_{it}^{γ} be the dependent variable and the exogenous regressors of aggregated unit $i \in \mathcal{N}_{\gamma}$ under γ, respectively. Once γ is fixed, X_{it}^{γ} and Y_{it}^{γ} can be constructed from block-level data Z_t and W_t respectively for each aggregated unit $i \in \mathcal{N}_{\gamma}$. In particular, we make the following assumption on how aggregated variables are constructed from block-level variables.

Assumption 2.1 (Aggregation)**.** *For each $\gamma \in \Gamma$ and $i \in \mathcal{N}_{\gamma}$, Y_{it}^{γ} and X_{it}^{γ} can be obtained by the following aggregation formulae;*

$$\begin{aligned} Y_{it}^{\gamma} &= \sum_{j \in \mathcal{J}} \mathbb{1}(\gamma_j = i) W_{jt} \\ X_{it}^{\gamma} &= \sum_{j \in \mathcal{J}} \mathbb{1}(\gamma_j = i) Z_{jt}. \end{aligned} \tag{2.1}$$

Assumption 2.1 states that we can construct X_{it}^{γ} as the sum of Z_{jt} over every block j in aggregated unit i under the delineation γ. Similarly, Y_{it}^{γ} can be constructed as the sum of W_{jt} over every block j in aggregated unit i under γ. It is worth stressing that the aggregation formulae (2.1) hold for every generic γ which is not necessarily the true delineation. This assumption essentially requires the outcome to be measured in levels rather than the mean, the median, the logarithmic transformation, or the first difference of some variables. We will discuss the extensions for these cases in Section 2.4.

Next, we make the assumption on the data generating process. We use $h^0 \in \mathcal{H}$ and $\gamma^0 \in \Gamma$ to denote the true structural function and the true delineation parameter respectively. Let $\mathcal{N}^0 := \mathcal{N}_{\gamma^0}$ be the set of true aggregated units. We also use $Y_{it}^0 = Y_{it}^{\gamma^0}$ and $X_{it}^0 = X_{it}^{\gamma^0}$ to denote the dependent variable and the regressors of a true aggregated unit $i \in \mathcal{N}^0$.

Assumption 2.2 (Data generating process)**.** *The observed data (Z_t, W_t) and the structural error term ε_{it} satisfy the following conditions,*

$$Y_{it}^0 = h^0(X_{it}^0) + \varepsilon_{it} \quad \textit{and} \quad \mathbb{E}\left[\varepsilon_{it} \mid Z_t\right] = 0 \quad \textit{for every } i \in \mathcal{N}^0. \tag{2.2}$$

Assumption 2.2 states that the dependent variable Y_{it}^0 and regressors X_{it}^0 satisfy the structural equation $Y_{it}^0 = h^0(X_{it}) + \varepsilon_{it}$ for each aggregated unit $i \in \mathcal{N}^0$ under the true delineation γ^0. We also assume the mean independence between the structural error term ε_{it} and block-level exogenous variables Z_t. The mean independence condition implies the conventional mean independence condition $\mathbb{E}[\varepsilon_{it} \mid X_{it}^0] = 0$, which is often assumed when the true delineation γ^0 is known. In that sense, Assumption 2.2 requires stronger exogeneity conditions than conventional models with known aggre-

gated spatial units. Notice that Assumption 2.2 does not place any restrictions on the distribution of W_t except for the structural equation (2.2) at the true aggregation level. Next, we restrict the space of the structural functions in the following manner.

Assumption 2.3 (The space of h)**.** *We assume that*

$$h^0 \in \mathcal{H} \subseteq \{h : h \text{ is continuously differentiable and } \mathbb{E}\left[|h(X_{it}^{\gamma})|\right] < \infty \text{ for every } i \in \mathcal{N}_\gamma \text{ and } \gamma \in \Gamma\}. \tag{2.3}$$

In Assumption 2.3, we focus on the class of continuously differentiable functions because we are interested in the derivative of h^0 (i.e., partial effect of the regressors). Assumption 2.3 also requires that $h(X_{it}^{\gamma})$ has a finite first moment for each $h \in \mathcal{H}$, $i \in \mathcal{N}_\gamma$, and $\gamma \in \Gamma$. Next, we make the following support conditions for block-level exogenous regressors Z_t.

Assumption 2.4 (Support condition)**.** *supp*$(Z_t) = \prod_{j \in \mathcal{J}} \mathcal{Z}_j$*, where* $\mathcal{Z}_j := \text{supp}(Z_{jt})$ *is a closed and nondegenerate* d_x*-dimensional interval for each* $j \in \mathcal{J}$*.*

Assumption 2.4 states that Z_{jt} has a rectangular support for each block j, and the support of Z_t is the product of the supports of block-level exogenous variables. Note that the support of Z_j induces the support of regressors for each aggregated unit. Let $\mathcal{X}_i^{\gamma} := \text{supp}(X_{it}^{\gamma})$ be the support of the regressors for the unit $i \in \mathcal{N}_\gamma$ under γ. We define $\mathcal{X}^{\gamma} := \bigcup_{i \in \mathcal{N}_\gamma} \mathcal{X}_i^{\gamma}$ as the union of the support of the regressors across all aggregated units under γ. We assume that $\mathcal{X}^{\gamma}$ is a subset of the domain of the structural function for coherency, that is, $\mathcal{X}^{\gamma} \subseteq \mathcal{X}$ for every $\gamma \in \Gamma$. We also use $\mathcal{X}_i^0$ and $\mathcal{X}^0$ to denote $\mathcal{X}_i^{\gamma^0}$ and $\mathcal{X}^{\gamma^0}$ respectively.

Example 2.1 (Economics of agglomeration)**.** We are interested in a simple regression model $\tilde{Y}_{it} = \theta_1 + \theta_2 \tilde{X}_{it} + \tilde{\varepsilon}_{it}$, where $\tilde{Y}_{it}$ and $\tilde{X}_{it}$ denote the average wage and the density of the population in city i and period t respectively. That is, $\tilde{Y}_{it} = S_{it}/R_{it}$ and $\tilde{X}_{it} = R_{it}/A_i$, where S_{it} is the sum of wages over all workers, R_{it} is the number of workers, and A_i is the area, respectively, in city i. If $\theta_2 > 0$, there is an agglomeration effect; the higher the density of workers, the higher the productivity of the area. A typical problem is that it is unclear how to delineate spatial units (e.g., Briant et al., 2010; Duranton, 2021).

We observe block-level data W_{jt} and $Z_{jt} = (Z_{1jt}, Z_{2j})$, where W_{jt} is the sum of wages over all workers, Z_{1jt} is the number of workers, and Z_{2j} is the area, respectively, in block j. Notice that $\tilde{Y}_{it}$

and $\tilde{X}_{it}$ do not satisfy the aggregation formulae in Assumption 2.1. However, we can rewrite the structural equation as

$$S_{it} = \theta_1 R_{it} + \theta_2 \frac{R_{it}^2}{A_i} + \varepsilon_{it}, \tag{2.4}$$

where $\varepsilon_{it} = R_{it}\tilde{\varepsilon}_{it}$. Then, by setting $Y_{it} = S_{it}$, $X_{it} = (R_{it}, A_i)$, the aggregation formulae in Assumption 2.1 are satisfied. Since the area Z_{2j} is time-invariant, Assumption 2.4 is imposed only on the number of workers, and the area can be treated as a part of the structural function of each aggregated unit as discussed in Section 2.4. The exogeneity condition in Assumption 2.2 (i.e., $\mathbb{E}[\varepsilon_{it} \mid Z_t] = 0$) is also satisfied if we are willing to assume $\mathbb{E}[\tilde{\varepsilon}_{it} \mid Z_t] = 0$. Note that the structural function is nonaffine with respect to the regressors X_{it} unless $\theta_2 = 0$, which is important for the identification of the true delineation as detailed below. □

Example 2.2 (Spatial entry competition)**.** Another relevant example arises in the context of spatial competition among firms, as studied in the field of industrial organization.[3] For simplicity, we consider a static entry game between two chain stores (A and B). Each firm's payoff in spatial market i is given by

$$\begin{aligned}
\pi_{it}^A &= (\theta_1 + \theta_2 X_{it} + \theta_3 Y_{it}^B + u_{it}^A) Y_{it}^A \\
\pi_{it}^B &= (\theta_1 + \theta_2 X_{it} + \theta_3 Y_{it}^A + u_{it}^B) Y_{it}^B,
\end{aligned}$$

where Y_{it}^A (resp., Y_{it}^B) is an indicator function that takes 1 if firm A (resp., B) enters market i, X_{it} is the population in market i, and u_{it}^A (resp., u_{it}^B) is the unobserved payoff shock that follows some known distribution $F_{u|X_{it}}$. We assume that firm A (resp., B) does not observe u_{it}^B (resp., u_{it}^A), and payoff shocks are independent across firms, markets, and periods.

The two firms play a static incomplete information game in each market i and each period t repeatedly. We also assume that a single Bayesian Nash equilibrium is played across all markets and periods. Then the equilibrium can be characterized by the conditional choice probability of entry $\mathbb{P}(Y_{it}^A = 1 \mid X_{it})$ and $\mathbb{P}(Y_{it}^B = 1 \mid X_{it})$ such that

$$\begin{aligned}
\mathbb{P}\left(Y_{it}^A = 1 \mid X_{it}\right) &= \int_{-\infty}^{\infty} \mathbb{1}\left\{u^A \geq -(\theta_1 + \theta_2 X_{it} + \theta_3 \mathbb{P}(Y_{it}^B = 1 \mid X_{it}))\right\} dF_{u|X_{it}} \\
\mathbb{P}\left(Y_{it}^B = 1 \mid X_{it}\right) &= \int_{-\infty}^{\infty} \mathbb{1}\left\{u^B \geq -(\theta_1 + \theta_2 X_{it} + \theta_3 \mathbb{P}(Y_{it}^A = 1 \mid X_{it}))\right\} dF_{u|X_{it}}.
\end{aligned}$$

[3] Examples of relevant papers in the literature include, but are not limited to, Bresnahan and Reiss, 1991, Ryan, 2012, and Collard-Wexler, 2013).

For simplicity, we assume that the equilibrium is symmetric, that is, $\mathbb{P}(Y_{it}^A = 1 \mid X_{it}) = \mathbb{P}(Y_{it}^B = 1 \mid X_{it})$. Then, by a well known inversion result (e.g., Hotz and Miller, 1993; Aguirregabiria and Mira, 2002, 2007; Pesendorfer and Schmidt-Dengler, 2008), we can identify the structural parameters $(\theta_1, \theta_2, \theta_3)$ from the conditional choice probability of entry $\mathbb{P}(Y_{it}^A = 1 \mid X_{it})$. To this end, we consider a nonparametric regression model $Y_{it} = h(X_{it}) + \varepsilon_{it}$, where $Y_{it} := Y_{it}^A$ and $h(X_{it}) := \mathbb{E}[Y_{it} \mid X_{it}] = \mathbb{P}(Y_{it}^A = 1 \mid X_{it})$.[4]

A typical problem is that such economic markets are usually not observed in data (see Berry et al., 2019 for instance). Instead, we observe block-level data W_{jt} and Z_{jt}, where W_{jt} is the number of firm A's stores in block j, and Z_{jt} is population in block j. For simplicity, we assume that $Y_{it}^\gamma := \sum_{j \in \mathcal{J}} \mathbb{1}(\gamma_j = i) W_{jt} \in \{0, 1\}$ for any $i \in \mathcal{N}_\gamma$ and $\gamma \in \Gamma$. Furthermore, we assume $u_{it}^A \perp\!\!\!\perp Z_t \mid X_{it}$, which implies that the moment condition $\mathbb{E}[\varepsilon_{it} \mid Z_t] = 0$ is satisfied. Then Assumptions 2.1 and 2.2 are satisfied, and the setup falls within the scope of the framework. □

2.2 Identification results

In this subsection, we provide several identification results on h^0 and γ^0. The block-level exogeneity condition (2.2) induces the conditional moment restrictions;

$$\mathbb{E}\left[Y_{it}^\gamma - h(X_{it}^\gamma) \mid Z_t\right] = 0 \quad \text{(almost surely)} \quad \text{for every } i \in \mathcal{N}_\gamma. \tag{2.5}$$

The question is whether these conditional moment restrictions are satisfied only with the true structural function h^0 and delineation γ^0. First we make an assumption on the nonaffinity of h^0.

Assumption 2.5 (Nonaffinity)**.** *The structural function h^0 is nonaffine on $\mathcal{X}_i^0$ for every $i \in \mathcal{N}^0$.*

Assumption 2.5 states that h^0 is nonaffine on the support of the regressors for every *true* aggregated unit $i \in \mathcal{N}^0$. The first result claims that we can identify both the true delineation γ^0 and structural function h^0 from the joint distribution of block-level variables (Z_t, W_t) under the nonaffine assumption.

Theorem 2.1. *Suppose Assumption 2.1, 2.2, 2.3, 2.4, and 2.5 hold. If $(h, \gamma) \in \mathcal{H} \times \Gamma$ satisfies the conditional moment restrictions (2.5), then $\gamma^0 = \gamma$ and $h^0(x) = h(x)$ for every $x \in \mathcal{X}^0$.*

[4] The setup can be extended to a dynamic discrete game model as in Aguirregabiria and Mira, 2007 and Pesendorfer and Schmidt-Dengler, 2008.

Theorem 2.1 states that the conditional moment restrictions (2.5) derived from the block-level exogeneity condition (i.e., $\mathbb{E}[\varepsilon_i \mid Z_t] = 0$) provide enough information to identify both the delineation and the structural function under the nonaffinity assumption. It is worth mentioning that we consider the conditional moment for every $i \in \mathcal{N}_\gamma$ under a candidate γ, which is not necessarily the true γ^0.

It is clear that, once we identify the true delineation γ^0, the conditional moment restrictions (2.5) are satisfied only with the true structural function h^0 (more precisely, h and h^0 coincide on $\mathcal{X}^0$). Thus, the main claim of Theorem 2.1 is that the restrictions (2.5) cannot be satisfied with any $h \in \mathcal{H}$ if the delineation is incorrect, that is, $\gamma \neq \gamma^0$.

To see the intuition behind Theorem 2.1, let us consider a simple case in which there are only two blocks so that $\mathcal{J} = \{j_1, j_2\}$. We also assume $\mathcal{Z}_{j_1} = \mathcal{Z}_{j_2} = \mathbb{R}^{d_x}$. Suppose that j_1 and j_2 jointly constitute an aggregated unit i under γ^0. Then, Assumption 2.1 and Assumption 2.2 gives

$$\begin{aligned} \mathbb{E}\left[W_{j_1} + W_{j_2} \mid Z_{j_1}, Z_{j_2}\right] &= \mathbb{E}\left[Y_i^0 \mid Z_{j_1}, Z_{j_2}\right] \\ &= \mathbb{E}\left[h^0(X_i^0) \mid Z_{j_1}, Z_{j_2}\right] \\ &= h^0\left(Z_{j_1} + Z_{j_2}\right) \end{aligned} \tag{2.6}$$

Now suppose that, by contradiction, the conditional moment restrictions (2.5) are satisfied with some $h \in \mathcal{H}$ under the incorrect γ such that j_1 and j_2 constitute two aggregated units i_1 and i_2 separately. Then we have

$$\begin{aligned} \mathbb{E}\left[W_{j_1} + W_{j_2} \mid Z_{j_1}, Z_{j_2}\right] &= \mathbb{E}\left[Y_{i_1}^\gamma + Y_{i_2}^\gamma \mid Z_{j_1}, Z_{j_2}\right] \\ &= \mathbb{E}\left[h\left(X_{i_1}^\gamma\right) \mid Z_{j_1}, Z_{j_2}\right] + \mathbb{E}\left[h\left(X_{i_2}^\gamma\right) \mid Z_{j_1}, Z_{j_2}\right] \\ &= h\left(Z_{j_1}\right) + h\left(Z_{j_2}\right) \end{aligned} \tag{2.7}$$

Thus, by combining two equations above, we obtain $h^0\left(Z_{j_1} + Z_{j_2}\right) = h\left(Z_{j_1}\right) + h\left(Z_{j_2}\right)$ for every $Z_{j_1}, Z_{j_2} \in \mathbb{R}^{d_x}$, which is called the Pexider functional equation (a variant of the Cauchy functional equation). We can show that this functional equation can be satisfied only when h^0 is affine in the class of continuous functions. However, this is a contradiction to the assumption that h^0 is nonaffine, and hence γ^0 is identified. Notice that the linearity of the expectation operator and the aggregation formulae (2.1) are crucial to derive the Pexider functional equation.

We can analogously show that the conditional moment restrictions (2.5) can be satisfied only with γ^0 in the opposite case, that is, j_1 and j_2 constitute two aggregated units i_1 and i_2 separately

under γ^0. Theorem 2.1 states that a more general identification result holds even when there are more than two blocks with limitted supports. Appendix A.2 provides the formal proof.

The nonaffinity of the structural function may seem uncommon, especially when parametric models are considered. However, in many empirically relevant parametric models, the structural function is often nonaffine with respect to regressors after suitable rearrangement of the structural equation so that aggregation formulae in Assumption 2.1 hold. For instance, in Example 2.1, h^0 is nonaffine with respect to the regressors unless $\theta_2 = 0$.

The next result shows that the derivative of h^0 (i.e., partial effect of the regressors) can be still identified even when h^0 is an affine function.

Assumption 2.6 (Affinity)**.** *The structural function h^0 is affine on $\mathcal{X}^0$, that is, there exist $\alpha^0 \in \mathbb{R}$ and $\beta^0 \in \mathbb{R}^{d_x}$ such that $h^0(x) = \alpha^0 + \beta^0 x$ for every $x \in \mathcal{X}^0$.*

Theorem 2.2. *Suppose that Assumption 2.1, 2.2, 2.3, 2.4, and 2.6 hold. If $(h, \gamma) \in \mathcal{H} \times \Gamma$ satisfies the conditional moment restrictions (2.5), then $\partial h(x) = \beta^0$ for every $x \in \mathcal{X}^\gamma$.*

In particular, Theorem 2.2 claims that if $(\alpha, \beta) \in \mathbb{R}^{d_x+1}$ satisfies the conditional moment restrictions;

$$\mathbb{E}\left[Y_{it}^\gamma - \alpha - \beta X_{it}^\gamma \mid Z_t\right] = 0 \quad \text{(almost surely)} \quad \text{for every } i \in \mathcal{N}_\gamma, \tag{2.8}$$

then $\beta = \beta^0$ under the affinity assumption. To see the intuition, let us consider the simple example with two blocks again. Suppose that j_1 and j_2 jointly constitute an aggregated unit i under γ^0 and they constitute two aggregated units i_1 and i_2 separately under γ. If the conditional moment restrictions (2.8) are satisfied with the incorrect γ and some $(\alpha, \beta) \in \mathbb{R}^{d_x+1}$, then we can derive $\alpha^0 + \beta^0(Z_{j1} + Z_{j2}) = \alpha + \beta Z_{j1} + \alpha + \beta Z_{j2}$ by equation (2.6) and (2.7). It is easy to see that $\beta = \beta^0$ must hold, which means that the derivative of the structural function can be identified.

Both Theorem 2.1 and 2.2 rely on assumptions on the functional form of h^0 (i.e., Assumption 2.5 or 2.6) for the identification of h^0. However, we may not be willing to impose such a priori assumptions, especially when considering nonparametric regression models. Without either assumption, h^0 may be affine on the support of some true aggregated units and nonaffine on that of others. Moreover, h^0 may be affine with different slopes on the supports of two distinct true aggregated units when these supports do not overlap. The following assumption rules out the latter case.

Assumption 2.7 (Overlapping supports)**.** *For any $i, i' \in \mathcal{N}^0$,* $\operatorname{int}\left(\mathcal{X}_i^0 \cap \mathcal{X}_{i'}^0\right) \neq \emptyset$.

Assumption 2.7 requires that the supports of the regressors overlap across true aggregated units. It is weaker than the common support assumption (i.e., $\mathcal{X}_i^0 = \mathcal{X}^0$ for every $i \in \mathcal{N}^0$). A primitive sufficient condition is that $\underline{z}_j \geq 0$ for every $j \in \mathcal{J}$ and $\sum_{j\in\mathcal{J}} \underline{z}_j < \min_{j\in\mathcal{J}} \bar{z}_j$ (coordinate-wise), where $\underline{z}_j := \min \mathcal{Z}_j$ and $\bar{z}_j := \sup \mathcal{Z}_j$, with the convention that the inequality is satisfied in a coordinate in which $\bar{z}_j = +\infty$. In particular, the condition holds whenever $\underline{z}_j = 0$ for every $j \in \mathcal{J}$, which is natural when block-level variables are measured in levels such as population, employment, or the number of establishments. The next result is on the identification of the derivative of h^0 without requiring a priori imposition of the (non)affinity assumption on h^0.

Theorem 2.3. *Suppose that Assumption 2.1, 2.2, 2.3, 2.4, and 2.7 hold. If $(h, \gamma) \in \mathcal{H} \times \Gamma$ satisfies the conditional moment restrictions (2.5), then $\partial h^0(x) = \partial h(x)$ for every $x \in \mathcal{X}^0 \cap \mathcal{X}^\gamma$.*

Theorem 2.3 states that the derivative of the structural function can be identified from the conditional moment restrictions (2.5) without the (non)affinity assumption. A caveat of the result is that the satisfaction of (2.5) implies the identification of the derivative only at $x \in \mathcal{X}^0 \cap \mathcal{X}^\gamma$. Intuitively, it is impossible to know $h(x)$ for $x \notin \mathcal{X}^0$ since the observed data (W_t, Z_t) is generated according to the structural equation (2.2) under the correct delineation γ^0. Also, the conditional moment (2.5) can restrict $h(x)$ only for $x \in \mathcal{X}^\gamma$ under γ. Unfortunately it is impossible to know $\mathcal{X}^0$ in general if γ^0 cannot be identified. However, we can identify ∂h^0 on $\bigcap_{\tilde{\gamma}\in\Gamma} \mathcal{X}^{\tilde{\gamma}}$, which is a subset of $\mathcal{X}^0 \cap \mathcal{X}^\gamma$. For example, if $\mathcal{Z}_j$ has a full support on the entire d_x-dimensional real space (i.e., $\mathcal{Z}_j = \mathbb{R}^{d_x}$) for every $j \in \mathcal{J}$, ∂h^0 is identified on the entire $\mathbb{R}^{d_x}$ since $\mathcal{X}^0 \cap \mathcal{X}^\gamma = \mathbb{R}^{d_x}$.

2.3 Implications of the identification results

Now, let us consider implications of Theorem 2.1 and 2.2 when the support of regressors is common across all true aggregated units, that is, $\mathcal{X}^0 = \mathcal{X}_i^0$ for every $i \in \mathcal{N}^0$.[5] Notice that either Assumption 2.5 or 2.6 must hold under the common support assumption. Thus, by Theorem 2.1 and 2.2, if (h, γ) satisfies the conditional moment restrictions (2.5) and h is nonaffine on $\mathcal{X}_k^\gamma$ for some $k \in \mathcal{N}_\gamma$, then $\gamma = \gamma^0$ and $h(x) = h^0(x)$ for every $x \in \mathcal{X}^0$.[6] Conversely, if (h, γ) satisfies the restrictions (2.5) and h is affine on $\mathcal{X}_k^\gamma$ for every $k \in \mathcal{N}_\gamma$, then the derivative of h on $\mathcal{X}^\gamma$ and that of h^0 on $\mathcal{X}^0$ coincide. Therefore, by combining the two cases, we conclude that the partial effects can be

[5] For instance, the assumption is trivially satisfied if Z_j has a large support (i.e., $\mathcal{Z}_j = \mathbb{R}^{d_x}$ or $\mathcal{Z}_j = \mathbb{R}_+^{d_x}$) for every block $j \in \mathcal{J}$.

[6] Here we use the fact that $\mathcal{X}_k^\gamma$ is convex under Assumption 2.4. Specifically, if Assumption 2.6 held, then Theorem 2.2 would imply $\partial h(x) = \beta^0$ for every $x \in \mathcal{X}_k^\gamma$ and hence that h is affine on $\mathcal{X}_k^\gamma$.

identified no matter whether h^0 is affine or nonaffine if the common support assumption is satisfied. The result is complementary to Theorem 2.3 since both results do not require a priori imposition of the (non)affinity assumption. Although Theorem 2.3 provides an identification result without making the (non)affinity assumption, the derivative of the structural function can be identified only at $x \in \mathcal{X}^0 \cap \mathcal{X}^\gamma$, which is an unknown set. We discuss more implications without the common support assumption in Appendix A.5.

The identification results give straightforward implications in many parametric models without the common support assumption. Now, the space of the structural functions can be expressed as $\mathcal{H} = \{h(\cdot, \theta) : \theta \in \Theta\}$, where $\Theta \subseteq \mathbb{R}^{d_\theta}$ is the space of the structural parameters. Let $\theta^0 \in \Theta$ be the true structural parameter such that $h(\cdot, \theta^0) = h^0(\cdot)$. We assume that θ^0 is identified if the true delineation is known, that is, the conditional moment restrictions $\mathbb{E}\left[Y_{it}^0 - h(X_{it}^0, \theta) \mid X_{it}^0\right] = 0$ are satisfied if and only if $\theta = \theta^0$. Notice that the functional form of the structural function can tell us whether the (non)affinity assumption is satisfied. For instance, in Example 2.1, the structural function is nonaffine everywhere if $\theta_2^0 \neq 0$, and affine if $\theta_2^0 = 0$. Thus, Theorem 2.1 and Theorem 2.2 imply that if (θ, γ) satisfies the conditional moment restriction $\mathbb{E}\left[Y_{it}^\gamma - h(X_{it}^\gamma, \theta) \mid Z_t\right] = 0$ for every $i \in \mathcal{N}_\gamma$ and $\theta_2 \neq 0$, then $(\theta^0, \gamma^0) = (\theta, \gamma)$. Also, if (θ, γ) satisfies the conditional moment restriction and $\theta_2 = 0$, then $\theta_2^0 = 0$. Therefore, if (θ, γ) satisfies the conditional moment restriction, then we conclude $\theta_2^0 = \theta_2$. We can also see that $\theta_1^0 = \theta_1$ holds if (θ, γ) satisfies the conditional moment restriction.

2.4 Extensions

Recall that Assumption 2.1 requires the endogenous outcome to be measured in levels. Up to this point, we have concentrated on pooled panel regression models without incorporating fixed effects or the heterogeneity of the structural function. Note that the mean outcome can be allowed as in Example 2.1. Below, we will consider some other important extensions.

2.4.1 Heterogeneous structural function

We can allow that the structural function depends on each aggregated unit, that is, $Y_{it}^0 = h_i^0(X_{it}^0) + \varepsilon_{it}$ and $\mathbb{E}\left[\varepsilon_{it} \mid Z_t\right] = 0$ for every $i \in \mathcal{N}^0$. More specifically, under the assumption that h_i^0 is nonaffine

on $\mathcal{X}_i^0$ for every $i \in \mathcal{N}^0$,

$$\mathbb{E}\left[Y_{it}^{\gamma} - h_i(X_{it}^{\gamma}) \mid Z_t\right] = 0 \quad \text{(almost surely)} \quad \text{for every } i \in \mathcal{N}_{\gamma}$$

is satisfied if and only if $\gamma = \gamma^0$ and $h_i = h_i^0$ for every $i \in \mathcal{N}^0$. The proofs in Appendix A.2 can be easily adjusted for this case.

2.4.2 Fixed effects

We can easily extend the model to a fixed effect model of the form

$$Y_{it} = \alpha_i + h(X_{it}) + \varepsilon_{it}.$$

Instead of the exogeneity condition in Assumption 2.2, we make the strict block-level exogeneity condition $\mathbb{E}[\varepsilon_{it} \mid \{Z_s\}_{s=1}^T] = 0$. Then, we can consider the first difference of the model

$$\tilde{Y}_{it}^{\gamma} = \tilde{h}(X_{it}^{\gamma}, X_{it-1}^{\gamma}) + \tilde{\varepsilon}_{it},$$

where $\tilde{Y}_{it}^{\gamma} := Y_{it}^{\gamma} - Y_{it-1}^{\gamma}$, $\tilde{h}(X_{it}^{\gamma}, X_{it-1}^{\gamma}) := h(X_{it}^{\gamma}) - h(X_{it-1}^{\gamma})$, and $\tilde{\varepsilon}_{it} := \varepsilon_{it} - \varepsilon_{it-1}$. Note that $\tilde{Y}_{it}^{\gamma}$ satisfies the aggregation formula in Assumption 2.1 with $\tilde{W}_{jt} := W_{jt} - W_{jt-1}$. We also have $\mathbb{E}[\tilde{\varepsilon}_{it} | Z_t, Z_{t-1}] = 0$ under the strict block-level exogeneity condition. Now we assume h is nonaffine. Since $\tilde{h}$ is nonaffine under the assumption, $\tilde{h}$ and γ can be identified by Theorem 2.1, which in turn implies that the derivative of h can be identified.

2.4.3 Logarithmic transformation

Economists are often interested in the model of the form

$$\log Y_{it} = h(X_{it}) + \varepsilon_{it} \quad \text{and} \quad \mathbb{E}[\varepsilon_{it} \mid X_{it}] = 0,$$

where X_{it} and Y_{it} are aggregated variables satisfying the aggregation formulae in Assumption 2.1. Note that we can rewrite the structural function as $Y_{it} = \exp(h(X_{it})) \cdot \exp(\varepsilon_{it})$. Then, under the additional assumption that $\varepsilon_{it} \perp\!\!\!\perp Z_t \mid X_{it}$, we have a new regression model:

$$Y_{it} = \tilde{h}(X_{it}) + \tilde{\varepsilon}_{it} \quad \text{and} \quad \mathbb{E}[\tilde{\varepsilon}_{it} \mid Z_t] = 0,$$

where $\tilde{h}(X_{it}) := \exp(h(X_{it})) \cdot \mathbb{E}[\exp(\varepsilon_{it}) \mid X_{it}]$. Since the endogenous variable on the left-hand side satisfies the aggregation formula (2.1), $\tilde{h}$ and γ can be identified as long as $\tilde{h}$ is nonaffine. Once γ is identified, we can identify the original structural function h obviously.

2.4.4 Log-transformed mean outcome with fixed effects

Finally let us consider the following regression:

$$\log\left(\frac{Y_{it}}{R_{it}}\right) = \alpha_i + h(X_{it}) + \varepsilon_{it},$$

where X_{it} represents the exogenous regressors including the population R_{it}. Y_{it}, X_{it}, and R_{it} satisfy the aggregation formulae in Assumption 2.1. In this specification, the outcome of interest is the log transformed mean outcome of Y_{it} and fixed effects are incorporated. The model can be rewritten as $Y_{it} = R_{it} \cdot \exp(\alpha_i + h(X_{it})) \cdot \exp(\varepsilon_{it})$. Then, under the additional assumption that $\varepsilon_{it} \perp\!\!\!\perp Z_t \mid X_{it}$, we have a new regression model:

$$Y_{it} = \tilde{h}_i(X_{it}) + \tilde{\varepsilon}_{it} \quad \text{and} \quad \mathbb{E}[\tilde{\varepsilon}_{it} \mid Z_t] = 0,$$

where $\tilde{h}_i(X_{it}) := R_{it} \cdot \exp(\alpha_i + h(X_{it})) \cdot \mathbb{E}[\exp(\varepsilon_{it}) \mid X_{it}]$. Although the structural function depends on each aggregated unit i, as explained above, $\tilde{h}_i$ and γ can be identified as long as $\tilde{h}_i$ is nonaffine. Again, once the correct delineation is identified, we can identify the original structural function h and the fixed effects α_i.

3 Estimation of parametric models

This section discusses estimation and inference procedures for parametric regression models. We observe the block-level data $(Z_t, W_t) = (\{Z_{jt}\}_{j\in\mathcal{J}}, \{W_{jt}\}_{j\in\mathcal{J}})$ for each period $t \in \{1, \dots, T\}$. Let $\mathcal{H} = \{h(\cdot, \theta) : \theta \in \Theta\}$ be the space of structural functions, where $\Theta \subseteq \mathbb{R}^{d_\theta}$ is the space of the structural parameters. Let $\theta^0 \in \Theta$ be the true structural parameter such that $h(\cdot, \theta^0) = h^0(\cdot)$. We focus on the nonaffinity case, that is, $h(\cdot, \theta^0)$ is nonaffine with respect to the regressors.

3.1 Estimator

Recall that Theorem 2.1 states $\mathbb{E}\left[Y_{it}^{\gamma} - h(X_{it}^{\gamma}, \theta) \mid Z_t\right] = 0$ for every $i \in \mathcal{N}_\gamma$ if and only if $\theta = \theta^0$ and $\gamma = \gamma^0$ under the nonaffinity assumption. If we assume that some instrumental function b can uniquely pin down (θ^0, γ^0), that is, $\mathbb{E}\left[b\left(Z_t\right)\left(Y_{it}^{\gamma} - h\left(X_{it}^{\gamma}, \theta\right)\right)\right] = 0$ is satisfied only with the true (θ^0, γ^0), then we can construct the generalized method of moments (GMM) estimator in a natural way. Intuitively, we can use the sample analogue of these moment restrictions to estimate θ and γ. However, an arbitrary choice of b may lead to the loss of enough restrictions for identification.

Below we consider transforming the conditional moments into unconditional ones without losing enough restrictions for identification. There are many papers on these methodologies in both parametric models (e.g., Bierens, 1982; Bierens, 1990; Bierens and Ploberger, 1997; Domínguez and Lobato, 2004) and nonparametric models (e.g., Newey and Powell, 2003; Horowitz, 2011; Chen and Christensen, 2018; Chen et al., 2025). In particular, we consider series approximation such as power series or splines following Newey and Powell, 2003.

Specifically, we propose the following estimator $(\hat{\theta}, \hat{\gamma})$ based on the identification result in Theorem 2.1.

$$(\hat{\theta}, \hat{\gamma}) = \underset{(\theta,\gamma)\in\Theta\times\Gamma}{\arg\min}\, \hat{Q}(\theta, \gamma), \tag{3.1}$$

where $\hat{Q}(\theta, \gamma)$ is the objective function

$$\hat{Q}(\theta, \gamma) = \frac{1}{|\mathcal{J}|} \sum_{j\in\mathcal{J}} \bar{m}_j(\theta, \gamma)' \, \hat{\Xi} \, \bar{m}_j(\theta, \gamma),$$

$\bar{m}_j(\theta, \gamma)$ is the sample moment for aggregated unit $i := \gamma_j \in \mathcal{N}_\gamma$

$$\bar{m}_j(\theta, \gamma) = \frac{1}{T} \sum_{t=1}^{T} b^K(Z_t) \left(Y_{it}^{\gamma} - h\left(X_{it}^{\gamma}, \theta\right)\right),$$

and $\hat{\Xi}$ is the weight matrix

$$\hat{\Xi} = \left(\frac{1}{T} \sum_{t=1}^{T} b^K(Z_t)\, b^K(Z_t)'\right)^{-}.$$

We use $b^K(Z_t) = (b_1(Z_t), ..., b_K(Z_t))'$ to denote an instrumental function which transforms the conditional moment restrictions into unconditional ones, where K is the dimension of the instrumental function growing as T increases. Next, we assume the following regularity conditions for the consistency of the proposed estimator $(\hat{\theta}, \hat{\gamma})$.

Assumption 3.1.

(i) (Z_t, W_t) *is strictly stationary and ergodic.*

(ii) Θ *is compact.*

(iii) For the true delineation γ^0*, the conditional moment restrictions*

$$\mathbb{E}\left[Y_{it}^0 - h(X_{it}^0, \theta) \mid X_{it}^0\right] = 0 \quad \textit{(almost surely)} \quad \textit{for every } i \in \mathcal{N}^0 \tag{3.2}$$

are satisfied if and only if $\theta = \theta^0$*.*

(iv) For any $b(Z_t)$ *with* $\mathbb{E}\left[b(Z_t)^2\right] < \infty$*, there exists a sequence of coefficients* π_K *with* $\mathbb{E}[\{b(Z_t) - b^K(Z_t)'\pi_K\}^2] \to 0, K \to \infty$*, and* $K/T \to 0$*.*

(v) For every $\gamma \in \Gamma$ *and* $i \in \mathcal{N}_\gamma$*, there exists a constant* $C < \infty$ *which does not depend on* T *such that*

$$\sup_{\theta \in \Theta} \max_{t \leq T} \sum_{s=1}^{T} \left|\mathbb{E}\left[\{\rho_{it}^\gamma(\theta) - \bar{\rho}_{it}^\gamma(\theta)\}\{\rho_{is}^\gamma(\theta) - \bar{\rho}_{is}^\gamma(\theta)\} \mid Z_1, \ldots, Z_T\right]\right| \leq C$$

almost surely, where $\rho_{it}^\gamma(\theta) := Y_{it}^\gamma - h(X_{it}^\gamma, \theta)$ *and* $\bar{\rho}_{it}^\gamma(\theta) := \mathbb{E}[\rho_{it}^\gamma(\theta) \mid Z_t]$

(vi) For every $\gamma \in \Gamma$ *and* $i \in \mathcal{N}_\gamma$*, there exists* $M : \mathcal{X} \to \mathbb{R}$ *and continuous* $g : \mathbb{R} \to \mathbb{R}$ *with* $g(0) = 0$ *such that* $|h(X_{it}^\gamma, \theta) - h(X_{it}^\gamma, \tilde{\theta})| \leq M(X_{it}^\gamma) g(\|\tilde{\theta} - \theta\|)$ *for all* $\tilde{\theta}, \theta \in \Theta$ *and* $\mathbb{E}\left[M(X_{it}^\gamma)^2\right]$ *is bounded.*

(vii) For every $\gamma \in \Gamma$ *and* $i \in \mathcal{N}_\gamma$*,* $\mathbb{E}\left[(Y_{it}^\gamma)^2\right] < \infty$ *and* $\mathbb{E}\left[h(X_{it}^\gamma, \theta^0)^2\right] < \infty$*.*

Assumption 3.1 (i) requires that the strict stationarity and ergodicity hold for block-level data, which is stronger than assuming them for aggregated variables. Assumption 3.1 (ii) is a standard assumption about the compactness of the space of θ. Note that Γ is also compact since J is fixed. Assumption 3.1 (iii) states that θ^0 can be identified from conventional conditional moment restrictions based on aggregated variables (Y_{it}^0, X_{it}^0) if the true delineation is known. Assumption 3.1 (iv) requires that the instrumental function $b^K(Z_t)$ becomes richer as the number of periods T increases and regulates the rate between K and T. The condition is satisfied with power or Fourier series, or B-splines. Assumption 3.1 (v) requires the conditional autocovariances of the residual to be absolutely summable uniformly in T, which bounds the largest eigenvalue of the conditional second moment matrix of the residual vector. Assumption 3.1 (vi) imposes a Hölder continuity property for

the structural function. Assumption 3.1 (vii) requires the aggregated outcome and the structural function to be square integrable, which guarantees that the population objective function $Q(\theta, \gamma)$ is finite for every (θ, γ). Then we have the following asymptotic result.

Proposition 3.1. *Suppose Assumption 2.1, 2.2, 2.3, 2.4, 2.5, and 3.1 hold. Then,* $(\hat{\theta}, \hat{\gamma}) \xrightarrow{p} (\theta^0, \gamma^0)$ *as* $T \to \infty$.

Proposition 3.1 shows the consistency of both structural parameters and delineations as the number of time periods T increases. Although we focus on the nonaffinity case in Proposition 3.1, we can similarly show the consistency of the slope parameter θ in the affinity case (i.e, $Y_{it} = \alpha + \beta X_{it} + \varepsilon_{it}$) under Assumption 2.6.

3.2 Asymptotic distribution

Next, we consider the asymptotic distribution of an estimator of θ. To this end, given the first stage estimator $\hat{\gamma}$, we construct the second stage estimator of θ. We use $\hat{Y}_{it} := Y_{it}^{\hat{\gamma}}$ and $\hat{X}_{it} := X_{it}^{\hat{\gamma}}$ to denote the dependent variable and regressors of aggregated unit $i \in \mathcal{N}_{\hat{\gamma}}$ respectively. Then we define the following least squares estimator.

$$\tilde{\theta} = \arg\min_{\theta \in \Theta} \tilde{Q}(\theta; \hat{\gamma}), \tag{3.3}$$

where $\tilde{Q}(\theta; \hat{\gamma})$ is the objective function

$$\tilde{Q}(\theta; \hat{\gamma}) = \frac{1}{|\mathcal{N}_{\hat{\gamma}}|} \sum_{i \in \mathcal{N}_{\hat{\gamma}}} \tilde{Q}_i(\theta; \hat{\gamma}),$$

and $\tilde{Q}_i(\theta; \hat{\gamma})$ is the sum of squared residuals over T periods for aggregated unit $i \in \mathcal{N}_{\hat{\gamma}}$

$$\tilde{Q}_i(\theta; \hat{\gamma}) := \frac{1}{T} \sum_{t=1}^{T} \left(\hat{Y}_{it} - h(\hat{X}_{it}, \theta) \right)^2.$$

In words, $\tilde{\theta}$ is the conventional least squares estimator assuming that data is generated according to the delineation $\hat{\gamma}$, which is obtained in the first stage. Next, we assume the following regularity conditions for establishing the asymptotic normality of $\tilde{\theta}$.

Assumption 3.2.

(i) $\theta^0 \in \operatorname{int}\Theta$.

(ii) $h(x,\theta)$ *is twice continuously differentiable with respect to* θ *for each* $x \in \mathcal{X}^0$.

(iii) For every $i \in \mathcal{N}^0$,

$$\mathbb{E}\left[\sup_{\theta\in\Theta(\theta^0)}\left[\left\|\partial_\theta h(X_{it}^0,\theta)\partial_\theta h(X_{it}^0,\theta)' - \left\{Y_{it}^0 - h(X_{it}^0,\theta)\right\}\partial_{\theta\theta} h(X_{it}^0,\theta)\right\|\right]\right] < \infty,$$

where $\Theta(\theta^0)$ *is an open neighborhood of* θ^0.

(iv) For a finite and positive definite V,

$$\frac{1}{\sqrt{|\mathcal{N}^0|T}}\sum_{i\in\mathcal{N}^0}\sum_{t=1}^{T}\partial_\theta h(X_{it}^0,\theta^0)\varepsilon_{it} \xrightarrow{d} N(0,V).$$

Assumption 3.2 imposes standard regularity conditions typically required to establish the asymptotic properties of the nonlinear least squares estimator. In Assumption 3.2 (i), we require that the true structural parameter is in the interior of the parameter space. Assumption 3.2 (ii) and (iii) are conventional differentiability and boundedness conditions. Assumption 3.2 (iv) is a high-level condition to ensure the asymptotic normality. We can begin with more primitive conditions such as mixing conditions and apply a suitable central limit theorem. Notice that we can consider the variance V under both temporal and spatial correlation (e.g., Conley, 1999; Sarafidis and Wansbeek, 2012). Then we have the following results.

Proposition 3.2. *Suppose Assumption 2.1, 2.2, 2.3, 2.4, 2.5, 3.1, and 3.2 hold. Then, the following (i) and (ii) hold.*

(i) $\tilde{\theta} \xrightarrow{p} \theta^0$ *as* $T \to \infty$.

(ii)

$$\sqrt{|\mathcal{N}_{\hat{\gamma}}|T}(\tilde{\theta} - \theta^0) \xrightarrow{d} N(0, \Lambda^{-1}V\Lambda^{-1})$$

as $T \to \infty$, *where*

$$\Lambda = \frac{1}{|\mathcal{N}^0|}\sum_{i\in\mathcal{N}^0}\mathbb{E}\left[\partial_\theta h(X_{it}^0,\theta^0)\partial_\theta h(X_{it}^0,\theta^0)'\right].$$

Proposition 3.2 (i) and (ii) show the consistency and asymptotic normality of $\tilde{\theta}$ respectively. Although Proposition 3.2 focuses on parametric models, we can consider nonparametric models

based on the nonparametric identification result on h^0 and γ^0 in Theorem 2.1. As noted above, we can rely on the literature on nonparametric instrumental variable estimation (e.g., Newey and Powell, 2003; Horowitz, 2011; Chen and Christensen, 2018; Chen et al., 2025).

3.3 Discussion

We have derived the consistency of $(\hat{\theta}, \hat{\gamma})$ and the asymptotic distribution of $\tilde{\theta}$ under the assumption that $h^0(\cdot) = h(\cdot, \theta^0)$ is nonaffine. In practice, even when the structural function is parametrized, the researcher does not know whether h^0 is nonaffine since θ^0 is unknown. However, as discussed in Section 2.3, the identification of θ^0 does not rely on the (non)affinity assumption in many parametric models. For instance, in Example 2.1, θ^0 is identified from the conditional moment restrictions whether or not $\theta_2^0 = 0$. In contrast, γ^0 is not identified when $\theta_2^0 = 0$ since the conditional moment restrictions are satisfied with θ^0 under any γ obtained by merging true aggregated units. Note that the asymptotic results above are not directly applicable to the case of $\theta_2^0 = 0$ since $\hat{\gamma}$ is not consistent in that case.

Although this paper does not pursue the statistical properties in detail, one would expect that the identification results enable us to conduct a specification test to assess whether the delineation imposed by the researcher is correct or not. We begin by assuming that the aggregation formulae and the block-level exogeneity condition are correctly specified (i.e., Assumptions 2.1 and 2.2). Under this premise, we consider the null hypothesis that the delineation γ specified by the researcher is correct (i.e., $H_0 : \gamma^0 = \gamma$). Under H_0, the conditional moment restrictions (2.5) are satisfied with $h = h^0$. Hence, the absence of any $h \in \mathcal{H}$ satisfying the restrictions (2.5) under γ leads to rejection of H_0. Conversely, if such an h exists, the interpretation depends on the functional form of the structural function. Under the nonaffinity assumption, Theorem 2.1 implies that $\gamma = \gamma^0$ and $h(x) = h^0(x)$ for every $x \in \mathcal{X}^0$. Without the (non)affinity assumption, Theorem 2.3 only implies that the derivative of h coincides with that of h^0 on $\mathcal{X}^0 \cap \mathcal{X}^\gamma$ under Assumption 2.7. Although the nonaffinity of h^0 cannot be verified a priori, the results in Section 2.3 allow us to rely on the nonaffinity of h instead of that of h^0. Under the common support assumption, if h is nonaffine on $\mathcal{X}_k^\gamma$ for some $k \in \mathcal{N}_\gamma$, then $\gamma = \gamma^0$. In Example 2.1, the existence of θ with $\theta_2 \neq 0$ satisfying the conditional moment restrictions implies $\gamma = \gamma^0$ even without the common support assumption. We expect that the methodologies in the literature of specification testing can be used in principle (e.g., Bierens, 1982; Bierens, 1990; Hong and White, 1995; Bierens and Ploberger, 1997).

3.4 Computational issues

In order to obtain $(\hat{\theta}, \hat{\gamma})$, we have to solve the minimization problem (3.1). Unfortunately, searching over the set of all possible partitions is often computationally infeasible. For example, if $|\mathcal{J}| = 100$, the number of possible partitions is over 2^{191} without any restrictions. This section discusses some plausible methods to mitigate the computational problem.

3.4.1 Multiple clusters

It is important to make reasonable restrictions on the space of possible delineations Γ. We consider the case in which the set of blocks $\mathcal{J}$ can be partitioned into multiple known small clusters beforehand. That is, the researcher knows that there is no aggregated unit across different clusters. We provide a computationally feasible algorithm to estimate $(\hat{\theta}, \hat{\gamma})$ for this scenario.

Let $\mathcal{C}$ be the set of independent and isolated clusters. Let $\mathcal{J}_c$ be the set of blocks in cluster $c \in \mathcal{C}$. For each $c \in \mathcal{C}$, we observe the block-level data $(Z_{ct}, W_{ct}) = (\{Z_{cjt}\}_{j \in \mathcal{J}_c}, \{W_{cjt}\}_{j \in \mathcal{J}_c})$. We assume that $\{(Z_{ct}, W_{ct})\}_{t=1}^{T}$ are independent across clusters. We follow the notations in Section 2 except for an additional subscript c for cluster c. For instance, $\gamma_c = \{\gamma_{cj}\}_{j \in \mathcal{J}_c} \in \Gamma_c$ is a delineation in cluster c, where Γ_c is the set of possible delineations in cluster c. Similarly, $\mathcal{N}_{\gamma_c}$ is the set of aggregated units in cluster c under γ_c.

Then, for each cluster $c \in \mathcal{C}$, we construct the first-stage estimator as

$$\left(\hat{\theta}_c, \hat{\gamma}_c\right) = \underset{\theta_c, \gamma_c}{\operatorname{argmin}} \, \hat{Q}_c\left(\theta_c, \gamma_c\right), \tag{3.4}$$

where $\hat{Q}_c\left(\theta_c, \gamma_c\right)$ is the objective function for cluster c

$$\hat{Q}_c\left(\theta_c, \gamma_c\right) := \frac{1}{|\mathcal{J}_c|} \sum_{j \in \mathcal{J}_c} \bar{m}_{cj}\left(\theta_c, \gamma_c\right)' \, \hat{\Xi}_c \, \bar{m}_{cj}\left(\theta_c, \gamma_c\right),$$

$\bar{m}_{cj}\left(\theta_c, \gamma_c\right)$ is the sample moment for aggregated unit $i := \gamma_{cj} \in \mathcal{N}_{\gamma_c}$

$$\bar{m}_{cj}\left(\theta_c, \gamma_c\right) = \frac{1}{T} \sum_{t=1}^{T} b^K(Z_{ct}) \left(Y_{cit}^{\gamma_c} - h(X_{cit}^{\gamma_c}, \theta_c)\right),$$

and $\hat{\Xi}_c$ is the weight matrix

$$\hat{\Xi}_c = \left(\frac{1}{T} \sum_{t=1}^{T} b^K(Z_{ct}) \, b^K(Z_{ct})'\right)^{-}.$$

Put differently, we estimate $\hat{\gamma}_c$ by allowing that the structural parameter can depend on each cluster c. This is computationally feasible if Γ_c is not a huge space for each cluster c, that is, the size of each cluster is small enough. The consistency of $\hat{\gamma}_c$ can be shown as in Proposition 3.1. Then, given $\hat{\gamma} = \{\hat{\gamma}_c\}_{c \in \mathcal{C}}$, we estimate the second-stage estimator as

$$\tilde{\theta} = \arg\min_{\theta \in \Theta} \sum_{c \in \mathcal{C}} \sum_{i \in \mathcal{N}_{\hat{\gamma}_c}} \tilde{Q}_i(\theta; \hat{\gamma}_c),$$

where $\tilde{Q}_i(\theta; \hat{\gamma}_c)$ is the sum of squared residuals over T periods for aggregated unit $i \in \mathcal{N}_{\hat{\gamma}_c}$

$$\tilde{Q}_i(\theta; \hat{\gamma}_c) := \frac{1}{T} \sum_{t=1}^{T} \left(\hat{Y}_{it} - h(\hat{X}_{it}, \theta) \right)^2.$$

The consistency and asymptotic normality of the second-stage estimator $\tilde{\theta}$ can be easily shown as in Proposition 3.2.

3.4.2 Greedy algorithm

Another strategy is to use the following greedy algorithm.

Algorithm 3.1 (Greedy algorithm)**.**

(1) Let γ^s be the delineation in step s. Set the initial delineation γ^1 as the finest one, that is, $|\mathcal{N}^1| = |\mathcal{J}|$, where $\mathcal{N}^s := \{\gamma_j^s : j \in \mathcal{J}\}$ is the set of aggregated units in step s. We label each aggregated unit as the set of finer units in the aggregated unit, that is, $\gamma_j^1 = \{j\}$. Set $s = 1$ and go to Step (2).

(2) For each subset of aggregated units $\mathcal{N} \subseteq \mathcal{N}^s$, we consider the following delineation $\gamma^{\mathcal{N}}$ such that

$$\gamma_j^{\mathcal{N}} := \begin{cases} \bigcup_{i \in \mathcal{N}} i & \text{if } \gamma_j^s \in \mathcal{N} \\ \gamma_j^s & \text{if } \gamma_j^s \notin \mathcal{N} \end{cases}$$

(3) Update the delineation as

$$\gamma^{s+1} = \operatorname*{argmin}_{\gamma \in \Gamma^{s+1} \cup \{\gamma^s\}} \min_{\theta \in \Theta} \hat{Q}(\theta, \gamma),$$

where $\Gamma^{s+1} := \{\gamma^{\mathcal{N}} : \mathcal{N} \subseteq \mathcal{N}^s\}$.

(4) Set $s = s + 1$ and return to Step (2) until convergence (i.e., $\gamma^s = \gamma^{s-1}$).

(5) Let $\hat{\gamma} = \gamma^s$ and $\hat{\theta} = \arg\min_{\theta \in \Theta} \hat{Q}(\theta, \hat{\gamma})$.

We start the algorithm with the finest delineation, that is, $\gamma_j^1 = \{j\}$ for each $j \in \mathcal{J}$ and $|\mathcal{N}_{\gamma^1}| = |\mathcal{J}|$. Then, in each iteration s, we determine the subset of existing aggregated units (i.e., subset of $\mathcal{N}^s$) whose merger leads to the greatest reduction in the value of the objective function. Further, we can reduce the computational burden of each step by restricting the search in Step (2) to $\mathcal{N}$ with $|\mathcal{N}| = 2$, that is, by merging only two aggregated units at each step. The algorithm terminates in finite steps since $|\mathcal{N}^s|$ is weakly decreasing. A caveat is that it can converge to a local optimum.

4 Simulation

This section presents some simulation results.

4.1 Parametric regression model

First, we consider the following parametric regression model; $Y_{it} = \theta_1 + \theta_2 X_{it} + \theta_3 X_{it}^2 + \varepsilon_{it}$. The structural function is nonaffine if and only if $\theta_3 \neq 0$. Although the true delineation γ^0 is unknown and $\{(Y_{it}, X_{it})\}_{i \in \mathcal{N}^0}$ cannot be directly observed, block-level data $(W_{jt}, Z_{jt})_{j \in \mathcal{J}}$ is observed. We conduct simulations with various θ, $|\mathcal{J}|$, and T.

For each cluster, we generate the block-level data according to the following data generating process.

(i) The block-level exogenous variable is generated in the following way. First, for each block j, Z_{j1} is independently drawn from the normal distribution $N(5, 1)$. Then, for each $t \in \{2, \ldots, T\}$, we compute $Z_{jt} = 2.5 + 0.5 \cdot Z_{jt-1} + e_{jt}$, where e_{jt} is independently drawn from the normal distribution $N(0, 1)$.

(ii) The delineation γ^0 is independently drawn from Γ equally likely, where Γ is the set of all contiguous partitions of $\mathcal{J} = \{1, \ldots, J\}$ (i.e., $|\Gamma| = 2^{|\mathcal{J}|-1}$).

(iii) For each aggregated unit $i \in \mathcal{N}^0$ and $t \in \{1, \ldots, T\}$, we compute X_{it} based on γ^0 and $\{Z_{jt}\}_{j \in \mathcal{J}}$ according to (2.1). Then, we compute $Y_{it} = \theta_1 + \theta_2 X_{it} + \theta_3 X_{it}^2 + \varepsilon_{it}$, where ε_{it} is independently drawn from $N(0, 25)$ across i and t.

(iv) For each aggregated unit $i \in \mathcal{N}^0$ and $t \in \{1, \ldots, T\}$, we compute $(W_{j_1}, \ldots, W_{j_\ell}) = u_{it} \cdot Y_{it}$, where $\{j_1, ..., j_\ell\} = \{j \in \mathcal{J} : \gamma_j^0 = i\}$ and u_{it} is a random variable uniformly drawn from the $(\ell - 1)$-simplex. u_{it} is independently drawn across i and t.

Using the block-level data generated according to the process described above, we compute the estimator $(\hat{\theta}, \hat{\gamma})$ and $\tilde{\theta}$ based on the algorithms discussed in Section 3.4. We also compute the 95% confidence intervals of θ with the asymptotic distribution derived in Proposition 3.2. To be more specific, since ε_{it} is homoskedastic, we replace $\Lambda^{-1}V\Lambda^{-1}$ in Proposition 3.2 with the following estimator,

$$\widehat{\Lambda^{-1}V\Lambda^{-1}} = \left(\frac{1}{|\mathcal{N}_{\hat{\gamma}}|T} \sum_{i \in \mathcal{N}_{\hat{\gamma}}} \sum_{t=1}^{T} (1, \hat{X}_{it}, \hat{X}_{it}^2)'(1, \hat{X}_{it}, \hat{X}_{it}^2) \right)^{-1} \left(\frac{1}{|\mathcal{N}_{\hat{\gamma}}|T} \sum_{i \in \mathcal{N}_{\hat{\gamma}}} \sum_{t=1}^{T} (\hat{Y}_{it} - \tilde{\theta}_1 - \tilde{\theta}_2 \hat{X}_{it} - \tilde{\theta}_3 \hat{X}_{it}^2)^2 \right).$$

Table 1 presents the simulation results with $(\theta_1, \theta_2, \theta_3) \in \{(1.0, 1.0, 1.0), (1.0, 1.0, 0.0)\}$, $J = 10$, $C = 1$, and $T \in \{25, 50, 100, 500, 1000\}$. For the instrumental function b^K, we use a polynomial sieve of order 3 (i.e., $k = 3$) and include pairwise cross-products of the block-level variables. We conducted 500 simulations for each specification.

Table 1: Simulation Results: Parametric Regression Model

		Bias			Coverage (95%)			Delineation
θ_3	T	$\tilde{\theta}_1$	$\tilde{\theta}_2$	$\tilde{\theta}_3$	θ_1	θ_2	θ_3	$\Pr(\hat{\gamma} = \gamma^0)$
	25	0.118	-0.031	0.002	0.960	0.956	0.942	1.000
	50	0.351	-0.013	0.000	0.952	0.934	0.926	1.000
1.0	100	0.067	0.011	-0.001	0.960	0.948	0.956	1.000
	500	-0.078	0.004	0.000	0.952	0.956	0.954	1.000
	1000	0.045	-0.009	0.000	0.956	0.956	0.958	1.000
	25	1.903	-0.526	0.019	0.876	0.850	0.852	0.052
	50	1.642	-0.448	0.007	0.886	0.902	0.898	0.034
0.0	100	1.742	-0.413	0.000	0.814	0.834	0.870	0.034
	500	1.340	-0.354	0.004	0.560	0.694	0.806	0.190
	1000	0.705	-0.186	0.002	0.660	0.748	0.892	0.578

In summary, when $(\theta_1, \theta_2, \theta_3) = (1.0, 1.0, 1.0)$, we obtain the correct delineation with high probability, which leads to good performance of $\tilde{\theta}$, and the coverage probabilities of the confidence intervals. However, when $(\theta_1, \theta_2, \theta_3) = (1.0, 1.0, 0.0)$ (i.e., the structural function is affine), the probability that we obtain the correct delineation is lower than when $(\theta_1, \theta_2, \theta_3) = (1.0, 1.0, 1.0)$ (i.e., the structural function is nonaffine), which leads to lower coverage probability. Recall that we may not be able to pin down the delineation correctly if the structural function is affine.

Figure 1: Delineation accuracy vs. degree of nonaffinity

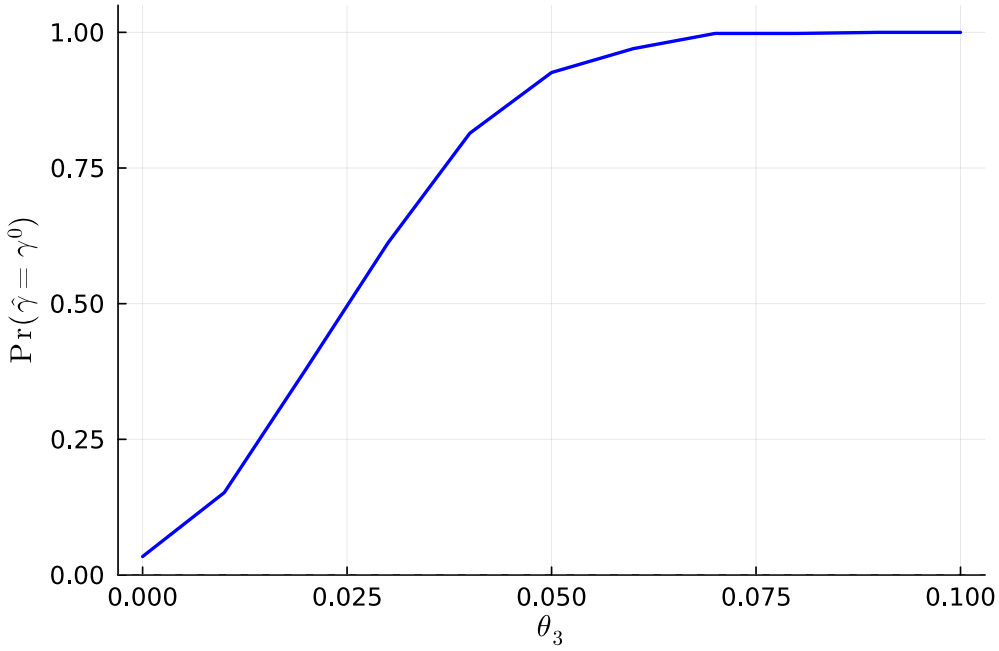


Figure 1 showcases how the probability that $\hat{\gamma} = \gamma^0$ changes as θ_3 changes with $T = 100$. These results imply that the true delineation becomes more likely to be correctly estimated as θ_3 increases, which is consistent with the identification result in Theorem 2.1, since a larger value of θ_3 corresponds to a greater degree of nonaffinity.

4.2 Spatial entry competition

We consider a static entry game with Q symmetric firms competing in each market and time period independently. The payoff from entry for each firm $q \in \{1, \ldots, Q\}$ in market i (i.e., spatial unit i) at time t is given by

$$\pi_{qit} = \alpha + \beta X_{it} + \delta \sum_{q' \neq q} Y_{q'it} + \varepsilon_{qit},$$

where $Y_{q'it} \in \{0, 1\}$ is the entry decision of firm q', X_{it} is a market characteristic (e.g., population), and ε_{qit} follows i.i.d. Gumbel distribution. The structural parameters are $\theta = (\alpha, \beta, \delta)$, where α is a constant, β captures the effect of market characteristics on entry profitability, and $\delta < 0$ represents the competitive effect of rival entry. The payoff from non-entry is normalized to 0.

In a symmetric Bayesian Nash equilibrium, all firms use the same entry probability $p(X_{it})$, which

satisfies the fixed point equation

$$p = \sigma\left(\alpha + \beta X_{it} + \delta(Q-1)p\right),$$

where $\sigma(v) = \exp(v)/(1+\exp(v))$ is the logistic function. The equilibrium conditional choice probability (CCP) is obtained by iterating this equation until convergence.

To connect with our general framework in Section 2, let Y_{it} be the number of firms that enter market i in period t. The structural function is defined as the expected number of entrants $h(X_{it}) := Q\,p(X_{it})$, and we have $\mathbb{E}[Y_{it} \mid X_{it}] = h(X_{it})$. This falls within the scope of Assumption 2.1 since both $Y_{it} = \sum_{j:\gamma_j^0=i} W_{jt}$ and $X_{it} = \sum_{j:\gamma_j^0=i} Z_{jt}$ satisfy the aggregation formulae, where Z_{jt} is the block-level counterpart of the market characteristic X_{it} and W_{jt} is the number of firms that entered block j. Unless $\beta = 0$, the structural function h is nonaffine in X_{it} due to the logistic transformation and the equilibrium fixed point, which makes the model suitable for our identification framework.

The data generating process is as follows.

(i) The block-level exogenous variable Z_{jt} is generated as in the parametric regression model.

(ii) The delineation γ^0 is independently drawn from contiguous partitions of $\mathcal{J}$.

(iii) For each aggregated unit $i \in \mathcal{N}^0$ and $t \in \{1, \ldots, T\}$, we compute X_{it} based on γ^0 and $\{Z_{jt}\}_{j\in\mathcal{J}}$ according to (2.1). We then solve for the equilibrium $p(X_{it})$.

(iv) Each firm $q \in \{1, \ldots, Q\}$ independently decides to enter with probability $p(X_{it})$. Conditional on entry, firm q randomly selects one block j within the aggregated unit i (i.e., from $\{j \in \mathcal{J} : \gamma_j^0 = i\}$). The block-level outcome W_{jt} counts the number of firms that entered and selected block j.

We estimate the model in two stages. In the first stage, we estimate the delineation $\hat{\gamma}$ by minimizing the GMM objective function, where the structural function $\mathbb{E}[Y_{it} \mid X_{it}]$ is approximated by a polynomial sieve. In the second stage, given $\hat{\gamma}$, we recover the structural parameters (α, β, δ) by nested fixed point (NFXP) estimation. For each candidate parameter vector, we solve the equilibrium fixed point for $p(X_{it})$, and we minimize the sum of squared deviations of the observed number of entrants Y_{it} from its model-implied mean $Q\,p(X_{it})$.

Table 2 presents the simulation results for the entry game with $(\alpha, \beta, \delta) = (-2.5, 0.5, -0.2)$, $J = 10$, $T = 50$, $C \in \{5, 10\}$, and $Q \in \{5, 10\}$. We use a polynomial sieve of order 4 for CCP estimation

and include pairwise cross-products as instrumental variables. We conducted 500 simulations for each specification.

Table 2: Simulation Results: Spatial Entry Competition

		Bias			RMSE			Delineation
C	Q	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\delta}$	α	β	δ	Share correct
5	5	-0.006	0.034	-0.097	0.122	0.110	0.324	0.995
5	10	0.012	-0.011	0.015	0.081	0.089	0.135	0.940
10	5	-0.009	0.021	-0.056	0.073	0.085	0.246	0.997
10	10	0.009	0.009	-0.015	0.059	0.068	0.102	0.940

The results show that α and β are estimated with small biases. The mean bias of δ is larger, reaching -0.097 against the true value -0.2 when $C = 5$ and $Q = 5$, but it is driven by a small number of replications with large estimates: the median estimate of δ is -0.210 in the same specification, so that the median bias is below 0.01. The last column of Table 2 reports the share of clusters c for which $\hat{\gamma}_c = \gamma_c^0$, averaged over the simulations. The delineation of each cluster is correctly recovered with high probability (94.0–99.7%). The RMSE decreases as the number of clusters increases.

5 Empirical Application

In this section, we apply the methodology to the economics of agglomeration with U.S. data. In particular, we consider the following regression

$$Y_{it} = \theta_1 X_{1it} + \theta_2 \frac{X_{1it}^2}{X_{2i}} + \varepsilon_{it},$$

where Y_{it}, X_{1it}, and X_{2i} are aggregated payroll, employment, and the area (km^2) of the aggregated unit i in period t. We rewrite the structural equation as in Example 2.1 so that Assumption 2.1 holds. Note that the structural function is nonaffine with respect to the regressors X_{it} unless $\theta_2 = 0$. As discussed in Example 2.1, θ_2 represents the economy of agglomeration.[7]

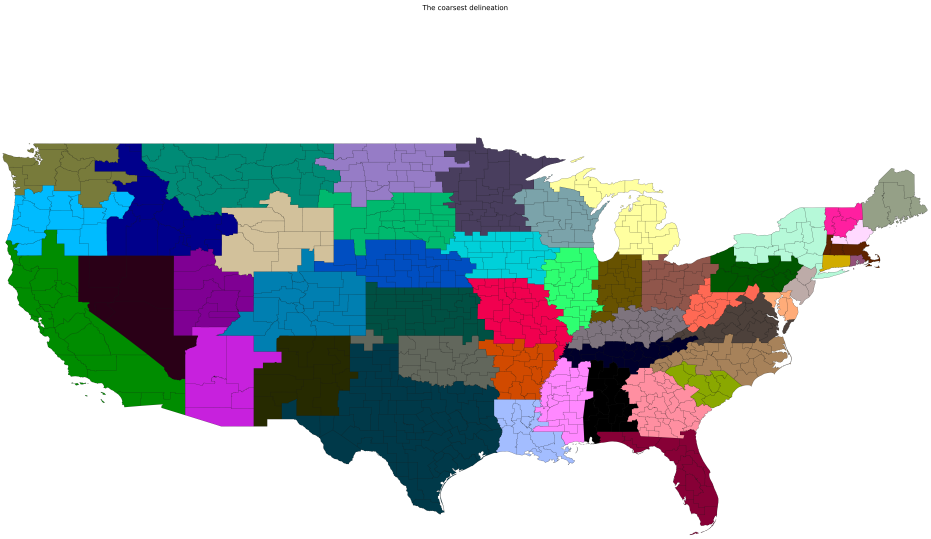
5.1 Data

We use Commuting Zone (CZ) level panel data from the United States. The data includes annual payroll, employment, and land area for 722 CZs across 47 states (excluding Alaska, Hawaii, Delaware,

[7] As mentioned in Section 2.4, the coefficients can depend on each aggregated unit i. Thus we can see that the identification result can be applied by interpreting θ_2/X_{2i} as a structural parameter of aggregated unit i.

and the District of Columbia). Each state constitutes a cluster c, resulting in $C = 47$ clusters. We consider 26 years (1998–2023) and 76 NAICS 3-digit industry sectors, and hence the number of year-sector combinations is $T = 1{,}976$. Since CZs are defined based on commuting patterns and can span multiple states, we assign each CZ to a single state using a majority rule. More specifically, each CZ is assigned to the state that contains the largest number of its constituent counties. The finest delineation γ^{fine} treats each CZ as a separate unit, yielding 722 aggregated units. The coarsest delineation γ^{coarse} aggregates all CZs within each state into a single unit, resulting in 47 aggregated units corresponding to states. Figure 2 shows the state and the CZ boundaries, representing the coarsest delineation structure.

Figure 2: Coarsest delineation



5.2 Estimation Results

We apply our proposed estimator to estimate the optimal delineation $\hat{\gamma}$ and the agglomeration parameter θ_2. Table 3 reports the point estimates and 95% confidence intervals for θ_2 under three delineations: finest, coarsest, and estimated.

The estimated agglomeration effect under the finest delineation is $\hat{\theta}_2 = 0.5800$ with a 95% confidence interval of $[0.5763, 0.5837]$. Under the coarsest delineation, the estimate increases to $\hat{\theta}_2 = 1.2639$ with interval $[1.1850, 1.3429]$. Our proposed estimator using $\hat{\gamma}$ yields $\hat{\theta}_2 = 1.9097$ with

Table 3: Parameter estimates

Delineation: γ	Estimate: θ_2
Finest	0.5800
	[0.5763, 0.5837]
Coarsest	1.2639
	[1.1850, 1.3429]
$\hat{\gamma}$	1.9097
	[1.8932, 1.9262]

interval $[1.8932, 1.9262]$. This estimate is not between the estimates under the finest and coarsest delineations, and hence the effect of the delineation on the estimate is not monotone in the coarseness of the delineation. The empirical results demonstrate the practical relevance of our methodology. The difference between estimates under finest and coarsest delineations (0.5800 vs. 1.2639) highlights the importance of choosing appropriate spatial units. Compared with the estimate under $\hat{\gamma}$, the CZ-level analysis underestimates the agglomeration effect by about 70%, and the state-level analysis underestimates it by about 34%.

Under the estimated delineation $\hat{\gamma}$, there are 441 aggregated units from the original 722 CZs. The aggregation ratio (i.e., the number of aggregated units divided by the number of original units) is 0.611. This indicates that the algorithm finds substantial aggregation of CZs, although the estimated delineation is still much finer than the coarsest one. Figure 3 displays the estimated delineation structure. CZs sharing the same color belong to the same aggregated unit. The figure shows that the algorithm aggregates adjacent CZs in most states. Table 4 provides detailed delineation information by state.

Table 5 presents summary statistics for the variables under the three delineations. Under the finest delineation, the average employment per CZ is 990 with substantial variation (standard deviation of 3,020). The coarsest delineation dramatically increases the average to 15,240 per state. Under the estimated delineation, the average is 1,620, which is closer to the finest than to the coarsest. Mean density (employees per km^2) under the finest delineation is 0.090, which increases to 0.177 under the coarsest delineation. Under our estimated delineation, the mean density is 0.089, which is close to the finest.

Figure 3: Estimated delineation $\hat{\gamma}$

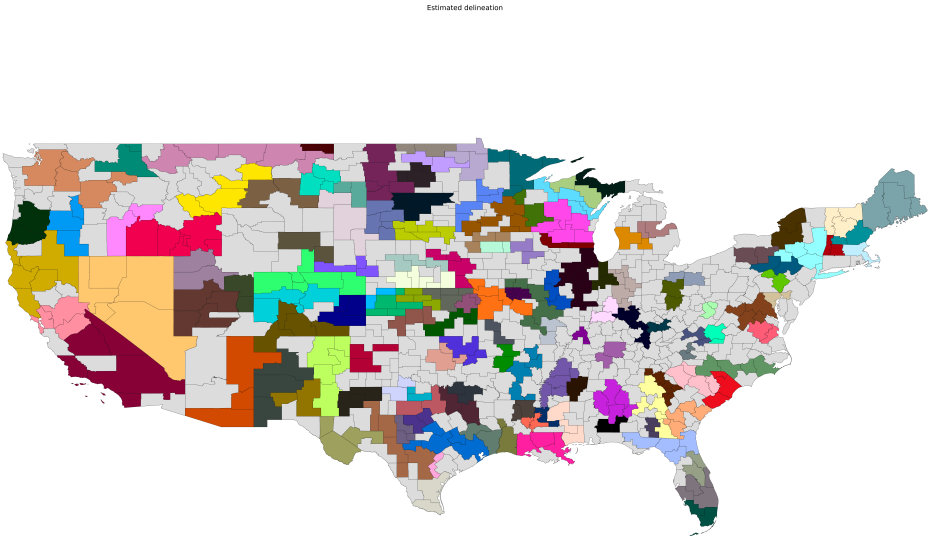


Table 4: Delineation information by state

State	CZs	Units	Rate	State	CZs	Units	Rate
Alabama	14	9	0.643	Nevada	5	1	0.200
Arizona	5	3	0.600	New Hampshire	2	1	0.500
Arkansas	16	11	0.688	New Jersey	3	3	1.000
California	18	8	0.444	New Mexico	12	5	0.417
Colorado	19	7	0.368	New York	12	6	0.500
Connecticut	1	1	1.000	North Carolina	22	18	0.818
Florida	16	8	0.500	North Dakota	20	10	0.500
Georgia	28	15	0.536	Ohio	16	14	0.875
Idaho	12	6	0.500	Oklahoma	17	14	0.824
Illinois	18	12	0.667	Oregon	13	10	0.769
Indiana	15	9	0.600	Pennsylvania	11	9	0.818
Iowa	22	18	0.818	Rhode Island	1	1	1.000
Kansas	33	17	0.515	South Carolina	8	3	0.375
Kentucky	24	18	0.750	South Dakota	23	9	0.391
Louisiana	13	8	0.615	Tennessee	16	15	0.938
Maine	4	1	0.250	Texas	65	41	0.631
Maryland	4	3	0.750	Utah	9	4	0.444
Massachusetts	5	3	0.600	Vermont	3	1	0.333
Michigan	18	14	0.778	Virginia	17	11	0.647
Minnesota	24	12	0.500	Washington	11	7	0.636
Mississippi	19	11	0.579	West Virginia	13	12	0.923
Missouri	25	14	0.560	Wisconsin	15	6	0.400
Montana	22	11	0.500	Wyoming	10	9	0.900
Nebraska	23	12	0.522				

Note: Units represents the number of aggregated units in $\hat{\gamma}$. Rate represents the aggregation rate, i.e., the number of aggregated units divided by the number of CZs.

Table 5: Summary statistics by delineation

	Finest	Coarsest	Estimated
Number of units	722	47	441
Payroll (million $)			
Mean	44.53	684.02	72.90
Std. Dev.	162.02	847.75	230.55
Employment (thousands)			
Mean	0.99	15.24	1.62
Std. Dev.	3.02	16.46	4.43
Area (thousand km^2)			
Mean	10.82	166.16	17.71
Std. Dev.	10.65	125.11	25.30
Density (employees per km^2)			
Mean	0.090	0.177	0.089
Std. Dev.	0.262	0.265	0.184

Note: For each aggregated unit, payroll and employment are averaged over the year-sector combinations. Density is the averaged employment divided by the area. Mean and Std. Dev. are taken across aggregated units.

6 Conclusion

This paper considers panel regression models in which each unit corresponds to an unknown aggregated spatial unit. Instead of observing data at the correct aggregation level, the researcher only has access to data on finer disaggregated spatial units. This setting is relevant to a wide range of applications, especially when researchers are interested in aggregated or equilibrium outcomes in local markets. Under a nonaffinity assumption on the structural function, we provide nonparametric identification results on both the structural function and the correct delineation. When the structural function is affine, we show that its derivatives can be still identified, which is crucial for obtaining marginal effects. Furthermore, we propose practical estimation and inference procedures based on GMM-type objective functions. A possible extension for future research is to consider a model with spatial spillovers across aggregated units. Another direction is investigating asymptotic properties of the estimators when the number of finer units also increases. Developing inference procedures for the delineation itself is also important.

References

Acemoglu, D., & Restrepo, P. (2020). Robots and jobs: Evidence from US labor markets. *Journal of Political Economy*, *128*(6), 2188–2244.

Aguirregabiria, V., & Mira, P. (2002). Swapping the nested fixed point algorithm: A class of estimators for discrete markov decision models. *Econometrica*, *70*(4), 1519–1543.

Aguirregabiria, V., & Mira, P. (2007). Sequential estimation of dynamic discrete games. *Econometrica*, *75*(1), 1–53.

Autor, D. H., Dorn, D., & Hanson, G. H. (2013). The China syndrome: Local labor market effects of import competition in the United States. *American Economic Review*, *103*(6), 2121–2168.

Berry, S., Gaynor, M., & Morton, F. S. (2019). Do increasing markups matter? Lessons from empirical industrial organization. *Journal of Economic Perspectives*, *33*(3), 44–68.

Bierens, H. J. (1982). Consistent model specification tests. *Journal of Econometrics*, *20*(1), 105–134.

Bierens, H. J. (1990). A consistent conditional moment test of functional form. *Econometrica: Journal of the Econometric Society*, 1443–1458.

Bierens, H. J., & Ploberger, W. (1997). Asymptotic theory of integrated conditional moment tests. *Econometrica: Journal of the Econometric Society*, 1129–1151.

Blundell, R., & Stoker, T. M. (2005). Heterogeneity and aggregation. *Journal of Economic Literature*, *43*(2), 347–391.

Bresnahan, T. F., & Reiss, P. C. (1991). Entry and competition in concentrated markets. *Journal of Political Economy*, *99*(5), 977–1009.

Briant, A., Combes, P.-P., & Lafourcade, M. (2010). Dots to boxes: Do the size and shape of spatial units jeopardize economic geography estimations? *Journal of Urban Economics*, *67*(3), 287–302.

Buchinsky, M., Li, F., & Liao, Z. (2022). Estimation and inference of semiparametric models using data from several sources. *Journal of Econometrics*, *226*(1), 80–103.

Chen, X., Christensen, T., & Kankanala, S. (2025). Adaptive estimation and uniform confidence bands for nonparametric structural functions and elasticities. *Review of Economic Studies*, *92*(1), 162–196.

Chen, X., & Christensen, T. M. (2018). Optimal sup-norm rates and uniform inference on nonlinear functionals of nonparametric IV regression. *Quantitative Economics*, *9*(1), 39–84.

Collard-Wexler, A. (2013). Demand fluctuations in the ready-mix concrete industry. *Econometrica*, *81*(3), 1003–1037.

Conley, T. G. (1999). GMM estimation with cross sectional dependence. *Journal of Econometrics*, *92*(1), 1–45.

Cross, P. J., & Manski, C. F. (2002). Regressions, short and long. *Econometrica*, *70*(1), 357–368.

de Paula, A., Rasul, I., & Souza, P. C. (2024). Identifying network ties from panel data: Theory and an application to tax competition. *Review of Economic Studies*, rdae088.

Domínguez, M. A., & Lobato, I. N. (2004). Consistent estimation of models defined by conditional moment restrictions. *Econometrica*, *72*(5), 1601–1615.

Duncan, O. D., & Davis, B. (1953). An alternative to ecological correlation. *American sociological review*, *18*(6).

Duranton, G. (2021). Classifying locations and delineating space: An introduction. *Journal of Urban Economics*, *125*, 103353.

Fan, Y., Sherman, R., & Shum, M. (2014). Identifying treatment effects under data combination. *Econometrica*, *82*(2), 811–822.

Fan, Y., Sherman, R., & Shum, M. (2016). Estimation and inference in an ecological inference model. *Journal of Econometric Methods*, *5*(1), 17–48.

Gehlke, C. E., & Biehl, K. (1934). Certain effects of grouping upon the size of the correlation coefficient in census tract material. *Journal of the American Statistical Association*, *29*(185A), 169–170.

Gorman, W. M. (1953). Community preference fields. *Econometrica*, *21*(1), 63–80.

Hong, Y., & White, H. (1995). Consistent specification testing via nonparametric series regression. *Econometrica: Journal of the Econometric Society*, 1133–1159.

Horowitz, J. L. (2011). Applied nonparametric instrumental variables estimation. *Econometrica*, *79*(2), 347–394.

Hotz, V. J., & Miller, R. A. (1993). Conditional choice probabilities and the estimation of dynamic models. *The Review of Economic Studies*, *60*(3), 497–529.

King, G. (1997). *A solution to the ecological inference problem: Reconstructing individual behavior from aggregate data*. Princeton University Press.

Kuczma, M. (2009). *An introduction to the theory of functional equations and inequalities: Cauchy's equation and Jensen's inequality*. Springer.

Lam, C., & Souza, P. C. (2016). Detection and estimation of block structure in spatial weight matrix. *Econometric Reviews*, *35*(8-10), 1347–1376.

Lam, C., & Souza, P. C. (2020). Estimation and selection of spatial weight matrix in a spatial lag model. *Journal of Business & Economic Statistics*, *38*(3), 693–710.

Manresa, E. (2013). Estimating the structure of social interactions using panel data. *Unpublished Manuscript*.

Molinari, F., & Peski, M. (2006). Generalization of a result on "regressions, short and long". *Econometric Theory*, *22*(1), 159–163.

Moon, S. (2024). Partial identification of individual-level parameters using aggregate data in a nonparametric binary outcome model. *arXiv preprint arXiv:2403.07236*.

Newey, W. K. (1991). Uniform convergence in probability and stochastic equicontinuity. *Econometrica: Journal of the Econometric Society*, 1161–1167.

Newey, W. K., & McFadden, D. (1994). Large sample estimation and hypothesis testing. *Handbook of Econometrics*, *4*, 2111–2245.

Newey, W. K., & Powell, J. L. (2003). Instrumental variable estimation of nonparametric models. *Econometrica*, *71*(5), 1565–1578.

Pesendorfer, M., & Schmidt-Dengler, P. (2008). Asymptotic least squares estimators for dynamic games. *The Review of Economic Studies*, *75*(3), 901–928.

Ridder, G., & Moffitt, R. (2007). The econometrics of data combination. *Handbook of Econometrics*, *6*, 5469–5547.

Robinson, W. (1950). Ecological correlations and the behavior of individuals. *American Sociological Review*, *15*(3), 351–357.

Ryan, S. P. (2012). The costs of environmental regulation in a concentrated industry. *Econometrica*, *80*(3), 1019–1061.

Sarafidis, V., & Wansbeek, T. (2012). Cross-sectional dependence in panel data analysis. *Econometric Reviews*, *31*(5), 483–531.

Theil, H. (1954). *Linear aggregation of economic relations*. North-Holland.

Tolbert, C. M., & Killian, M. S. (1987). Labor market areas for the United States. washington, dc: US department of agriculture. *Economic Research Service*.

Tolbert, C. M., & Sizer, M. (1996). US commuting zones and labor market areas: A 1990 update.

A Proofs in Section 2

A.1 Auxiliary lemmas

Definition A.1. *A real-valued function $f : \mathbb{R}^{d_x} \to \mathbb{R}$ is called additive if it satisfies Cauchy's functional equation*

$$f(x_1 + x_2) = f(x_1) + f(x_2)$$

for all $x_1, x_2 \in \mathbb{R}^{d_x}$.

Lemma A.1. *If a real-valued function $f : \mathbb{R}^{d_x} \to \mathbb{R}$ is a continuous additive function, then there exists $\beta \in \mathbb{R}^{d_x}$ such that*

$$f(x) = \beta x.$$

Proof of Lemma A.1. See Theorem 5.5.2 in Kuczma, 2009. □

Lemma A.2. *Let $G \subset \mathbb{R}^{d_x} \times \mathbb{R}^{d_x}$ be an open and connected set such that $G_1 \cap G_2 \cap G_3 \neq \emptyset$, where*

$$\begin{aligned} G_1 &= \left\{ x_1 \in \mathbb{R}^{d_x} \mid \text{there exists } x_2 \text{ such that } (x_1, x_2) \in G \right\}, \\ G_2 &= \left\{ x_2 \in \mathbb{R}^{d_x} \mid \text{there exists } x_1 \text{ such that } (x_1, x_2) \in G \right\} \\ G_3 &= \left\{ z \in \mathbb{R}^{d_x} \mid z = x_1 + x_2 \text{ for some } (x_1, x_2) \in G \right\}. \end{aligned}$$

Also, let $D \subset \mathbb{R}^N$ be a set fulfilling $G_1 \cup G_2 \cup G_3 \subset D$. If a function $f : D \to \mathbb{R}$ satisfies $f(x_1 + x_2) = f(x_1) + f(x_2)$ on $(x_1, x_2) \in G$, then there exist a unique additive function $g : \mathbb{R}^{d_x} \to \mathbb{R}$ such that $f(x) = g(x)$ for $x \in G_1 \cup G_2 \cup G_3$.

Proof of Lemma A.2. See Corollary 13.6.2 in Kuczma, 2009. □

A.2 Proof of Theorem 2.1

We show the following lemma, which immediately implies Theorem 2.1.

Lemma A.3. *Suppose Assumption 2.1, 2.2, 2.3, and 2.4 hold. We also assume that h^0 is nonaffine on $\mathcal{X}_i^0$ for $i \in \mathcal{N}^0$. If $(h, \gamma) \in \mathcal{H} \times \Gamma$ satisfies the conditional moment restrictions;*

$$\mathbb{E}\left[Y_{kt}^{\gamma} - h(X_{kt}^{\gamma}) \mid Z_t \right] = 0 \quad \textit{(almost surely)}, \quad \textit{for every } k \in \mathcal{N}_{\gamma}, \tag{A.1}$$

then $\{j \in \mathcal{J} : \gamma_j^0 = i\} = \{j \in \mathcal{J} : \gamma_j = k\}$ for some $k \in \mathcal{N}_{\gamma}$ and $h(x) = h^0(x)$ for every $x \in \mathcal{X}_i^0$.

Proof of Lemma A.3. It suffices to show that the conditional moment restriction (A.1) is satisfied only with γ such that $\{j \in \mathcal{J} : \gamma_j^0 = i\} = \{j \in \mathcal{J} : \gamma_j = k\}$ for some $k \in \mathcal{N}_\gamma$. Take any γ such that $\{j \in \mathcal{J} : \gamma_j^0 = i\} \neq \{j \in \mathcal{J} : \gamma_j = k\}$ for any $k \in \mathcal{N}_\gamma$. We assume, without loss of generality, $\mathcal{N}^0 = \{1, 2, ..., N\}$ and $\mathcal{N}_\gamma = \{1, 2, ..., N'\}$. We suppress the subscript t hereafter. Suppose, by contradiction,

$$\mathbb{E}\left[Y_k^\gamma - h(X_k^\gamma) \mid Z\right] = 0 \quad \text{(almost surely)}, \quad \text{for every } k \in \mathcal{N}_\gamma \tag{A.2}$$

for some function $h \in \mathcal{H}$.

For each $i \in \mathcal{N}^0$ and $k \in \mathcal{N}_\gamma$, we define $Z_{ik} := \sum_{j \in \mathcal{J}} \mathbb{1}(\gamma_j^0 = i \text{ and } \gamma_j = k) \cdot Z_j \in \mathcal{Z}_{ik}$, where $\mathcal{Z}_{ik}$ is the support of Z_{ik}. Let z_{ik}^* be an arbitrary point in the relative interior of $\mathcal{Z}_{ik}$. We also define $x_i^* := \sum_{k \in \mathcal{N}_\gamma} z_{ik}^*$, which belongs to the interior of $\mathcal{X}_i^0$ since $\mathcal{X}_i^0 = \sum_{k \in \mathcal{N}_\gamma} \mathcal{Z}_{ik}$ and at least one $\mathcal{Z}_{ik}$ is nondegenerate. Notice that the conditional moment restriction (A.2) implies

$$\begin{aligned}
\mathbb{E}\left[Y_k^\gamma \mid Z\right] &= \mathbb{E}\left[h(X_k^\gamma) \mid Z\right] \\
&= \mathbb{E}\left[h(Z_{1k} + \cdots + Z_{ik} + \cdots + Z_{Nk}) \mid Z\right] \\
&= h(Z_{1k} + \cdots + Z_{ik} + \cdots + Z_{Nk})
\end{aligned}$$

for each $k \in \mathcal{N}_\gamma$. Then we have

$$\begin{aligned}
\mathbb{E}\left[\sum_{j \in \mathcal{J}} W_j \mid Z\right] &= \mathbb{E}\left[\sum_{k \in \mathcal{N}_\gamma} Y_k \mid Z\right] \\
&= \sum_{k \in \mathcal{N}_\gamma} h(Z_{1k} + \cdots + Z_{ik} + \cdots + Z_{Nk}) \quad \text{(almost surely)}
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{E}\left[\sum_{j \in \mathcal{J}} W_j \mid Z\right] &= \mathbb{E}\left[\sum_{i \in \mathcal{N}^0} Y_i \mid Z\right] \\
&= \sum_{i \in \mathcal{N}^0} h^0(Z_{i1} + \cdots + Z_{ik} + \cdots + Z_{iN'}),
\end{aligned}$$

which imply

$$\sum_{k \in \mathcal{N}_\gamma} h(Z_{1k} + \cdots + Z_{ik} + \cdots + Z_{Nk}) = \sum_{i \in \mathcal{N}^0} h^0(Z_{i1} + \cdots + Z_{ik} + \cdots + Z_{iN'}) \quad \text{(almost surely)}. \tag{A.3}$$

Since h^0 and h are continuous, the equality (A.3) holds on $\text{supp}(Z_t)$. Recall that $\{j \in \mathcal{J} : \gamma_j^0 = i\} \neq \{j \in \mathcal{J} : \gamma_j = k\}$ for any $k \in \mathcal{N}_\gamma$. Thus, it suffices to consider the following Case 1 and Case 2.

Case 1: There exist two distinct $k_1, k_2 \in \mathcal{N}_\gamma$ such that $\{j \in \mathcal{J} : \gamma^0 = i \text{ and } \gamma = k_1\} \neq \emptyset$ and $\{j \in \mathcal{J} : \gamma^0 = i \text{ and } \gamma = k_2\} \neq \emptyset$.

First, we construct $m_1, m_2 \subset \mathcal{N}_\gamma$ such that $m_1 \cup m_2 = \mathcal{N}_\gamma$, $m_1 \cap m_2 = \emptyset$, $k_1 \in m_1$, and $k_2 \in m_2$. Define $Z_{m_1} := \sum_{k \in m_1} Z_{ik}$ and $Z_{m_2} := \sum_{k \in m_2} Z_{ik}$. We use $\mathcal{Z}_{m_1}$ and $\mathcal{Z}_{m_2}$ to denote the supports of Z_{m_1} and Z_{m_2} respectively. Notice that each of these supports is a nondegenerate d_x-dimensional interval. Also, let $z^*_{m_1} := \sum_{k \in m_1} z^*_{ik}$ and $z^*_{m_2} := \sum_{k \in m_2} z^*_{ik}$. Note that $\mathcal{X}_i^0 = \mathcal{Z}_{m_1} + \mathcal{Z}_{m_2}$ and $z^*_{m_1} + z^*_{m_2} = x_i^*$, where $z^*_{m_1} \in \text{int}\mathcal{Z}_{m_1}$ and $z^*_{m_2} \in \text{int}\mathcal{Z}_{m_2}$.

Fix Z_j arbitrary for each $j \in \mathcal{J}$ such that $\gamma_j^0 \neq i$. Then, by the equation (A.3), we have

$$h^0(Z_{m_1} + Z_{m_2}) + C = \sum_{k \in m_1} h(Z_{ik} + C_k) + \sum_{k \in m_2} h(Z_{ik} + C_k), \tag{A.4}$$

where C and C_k are constants which does not depend on Z_j for any j such that $\gamma_j^0 = i$. Then, by (A.4), we have the following four equations

$$\begin{aligned}
h^0(z^*_{m_1} + z^*_{m_2}) + C &= \sum_{k \in m_1} h(z^*_{ik} + C_k) + \sum_{k \in m_2} h(z^*_{ik} + C_k) \\
h^0(z^*_{m_1} + Z_{m_2}) + C &= \sum_{k \in m_1} h(z^*_{ik} + C_k) + \sum_{k \in m_2} h(Z_{ik} + C_k) \\
h^0(Z_{m_1} + z^*_{m_2}) + C &= \sum_{k \in m_1} h(Z_{ik} + C_k) + \sum_{k \in m_2} h(z^*_{ik} + C_k) \\
h^0(Z_{m_1} + Z_{m_2}) + C &= \sum_{k \in m_1} h(Z_{ik} + C_k) + \sum_{k \in m_2} h(Z_{ik} + C_k)
\end{aligned}$$

for every $Z_{m_1} \in \mathcal{Z}_{m_1}$, $Z_{m_2} \in \mathcal{Z}_{m_2}$, and $\{Z_{ik}\}_{k \in \mathcal{N}_\gamma}$ such that $Z_{m_1} := \sum_{k \in m_1} Z_{ik}$ and $Z_{m_2} := \sum_{k \in m_2} Z_{ik}$. Then, by combining the four equations above, we have

$$h^0(z^*_{m_1} + z^*_{m_2}) + h^0(Z_{m_1} + Z_{m_2}) = h^0(Z_{m_1} + z^*_{m_2}) + h^0(z^*_{m_1} + Z_{m_2}) \tag{A.5}$$

for every $Z_{m_1} \in \mathcal{Z}_{m_1}$ and $Z_{m_2} \in \mathcal{Z}_{m_2}$.

Let $\mathcal{X}_\Delta := \mathcal{X}_i^0 - \{x_i^*\}$. Define $f : \mathcal{X}_\Delta \to \mathbb{R}$ as

$$f(x_\Delta) := h^0(x_\Delta + x_i^*) - h^0(x_i^*). \tag{A.6}$$

Then, by equation (A.5) and (A.6), we have

$$f(Z_{m_1} - z^*_{m_1} + Z_{m_2} - z^*_{m_2}) = f(Z_{m_1} - z^*_{m_1}) + f(Z_{m_2} - z^*_{m_2}),$$

for every $Z_{m_1} \in \mathcal{Z}_{m_1}$ and $Z_{m_2} \in \mathcal{Z}_{m_2}$, which implies

$$f(Z_{\Delta_1} + Z_{\Delta_2}) = f(Z_{\Delta_1}) + f(Z_{\Delta_2}),$$

for every $(Z_{\Delta_1}, Z_{\Delta_2}) \in G := \text{int}\mathcal{Z}_{\Delta_1} \times \text{int}\mathcal{Z}_{\Delta_2} := \text{int}(\mathcal{Z}_{m_1} - \{z^*_{m_1}\}) \times \text{int}(\mathcal{Z}_{m_2} - \{z^*_{m_2}\})$. Let

$$\begin{aligned}
G_1 &:= \{z_{\Delta_1} \mid \text{ there exists } z_{\Delta_2} \text{ such that } (z_{\Delta_1}, z_{\Delta_2}) \in G\} = \text{int}\mathcal{Z}_{\Delta_1} \\
G_2 &:= \{z_{\Delta_2} \mid \text{ there exists } z_{\Delta_1} \text{ such that } (z_{\Delta_1}, z_{\Delta_2}) \in G\} = \text{int}\mathcal{Z}_{\Delta_2} \\
G_3 &:= \{x_\Delta \mid x_\Delta = z_{\Delta_1} + z_{\Delta_2} \text{ for some } (z_{\Delta_1}, z_{\Delta_2}) \in G\} = \text{int}\mathcal{X}_\Delta = \text{int}\mathcal{Z}_{\Delta_1} + \text{int}\mathcal{Z}_{\Delta_2}.
\end{aligned}$$

Then, $G_1 \cap G_2 \cap G_3 \neq \emptyset$. Notice that Lemma A.2 provides an additive function g which coincides with f on $G_3 = \text{int}\mathcal{X}_\Delta$, and that g is continuous on $\mathbb{R}^{d_x}$ since f is continuous by Assumption 2.3 and $g(x+h) - g(x) = g(h) = g(x_0 + h) - g(x_0)$ for every $x, h \in \mathbb{R}^{d_x}$ and $x_0 \in G_3$. Thus, by Lemma A.1 and A.2, there exists $\beta \in \mathbb{R}^{d_x}$ such that $f(x_\Delta) = \beta x_\Delta$ for every $x_\Delta \in \text{int}\mathcal{X}_\Delta$. Then, the construction of f (i.e., (A.6)) and the continuity of h^0 imply that $h^0(x) = (h^0(x^*_i) - \beta x^*_i) + \beta x$ for every $x \in \mathcal{X}^0_i$, which is a contradiction to the assumption that h^0 is nonaffine on $\mathcal{X}^0_i$.

Case 2: There exist $\ell \geq 1$, distinct $j_1, ..., j_\ell \in \mathcal{J} \setminus \{j \in \mathcal{J} : \gamma^0_j = i\}$ and $k \in \mathcal{N}_\gamma$ such that $\{j \in \mathcal{J} : \gamma^0_j = i\} \cup \{j_1, j_2, ..., j_\ell\} = \{j \in \mathcal{J} : \gamma_j = k\}$.

Fix Z_j arbitrary for each $j \in \mathcal{J}$ such that $\gamma_j \neq k$. Define

$$\tilde{h}(Z_{j_1}, ..., Z_{j_\ell}) := \sum_{i' \in \mathcal{N}^0 \setminus \{i\}} h^0\left(Z_{i'1} + \cdots + Z_{i'k} + \cdots + Z_{i'N'}\right) \tag{A.7}$$

Then, by (A.3) and (A.7), we have

$$h^0(X_i) + \tilde{h}(Z_{j_1}, ..., Z_{j_\ell}) = h(X_i + Z_{j_1} + \cdots Z_{j_\ell}) + C, \tag{A.8}$$

where $X_i = X^0_i = \sum_{j \in \mathcal{J}} \mathbb{1}(\gamma^0_j = i) Z_j$ and C is a constant which does not depend on $X_i, Z_{j_1}, ..., Z_{j_\ell}$.

Now, by (A.8), we have the following four equations

$$\begin{aligned}
h^0(x_i^*) + \tilde{h}(z_{j_1}^*, ..., z_{j_\ell}^*) &= h(x_i^* + z_{j_1}^* + \cdots + z_{j_\ell}^*) + C \\
h^0(X_\Delta + x_i^*) + \tilde{h}(z_{j_1}^*, ..., z_{j_\ell}^*) &= h(X_\Delta + x_i^* + z_{j_1}^* + \cdots + z_{j_\ell}^*) + C \\
h^0(X'_\Delta + x_i^*) + \tilde{h}(Z_{j_1}, ..., Z_{j_\ell}) &= h(X'_\Delta + x_i^* + Z_{j_1} + \cdots + Z_{j_\ell}) + C \\
h^0(X_\Delta + X'_\Delta + x_i^*) + \tilde{h}(Z_{j_1}, ..., Z_{j_\ell}) &= h(X_\Delta + X'_\Delta + x_i^* + Z_{j_1} + \cdots + Z_{j_\ell}) + C
\end{aligned} \tag{A.9}$$

for every $X_\Delta, X'_\Delta \in \mathcal{X}_\Delta := \mathcal{X}_i^0 - \{x_i^*\}$, and $Z_{j_1} \in \mathcal{Z}_{j_1}, ..., Z_{j_\ell} \in \mathcal{Z}_{j_\ell}$ such that $X_\Delta + X'_\Delta \in \mathcal{X}_\Delta$ and $X'_\Delta = (z_{j_1}^* + \cdots + z_{j_\ell}^*) - (Z_{j_1} + \cdots + Z_{j_\ell})$, where $z_{j_1}^* \in \text{int}\mathcal{Z}_{j_1}, \ldots, z_{j_\ell}^* \in \text{int}\mathcal{Z}_{j_\ell}$ are arbitrarily fixed points.

Define $f : \mathcal{X}_\Delta \to \mathbb{R}$ as

$$f(x_\Delta) := h^0(x_\Delta + x_i^*) - h^0(x_i^*). \tag{A.10}$$

Then, equations (A.9) and (A.10) imply that

$$f(X_\Delta + X'_\Delta) = f(X_\Delta) + f(X'_\Delta) \tag{A.11}$$

for every $(X_\Delta, X'_\Delta) \in G$, where

$$\begin{aligned}
G := \text{int}\{(X_\Delta, X'_\Delta) : X_\Delta, X'_\Delta \in \mathcal{X}_\Delta, & X_\Delta + X'_\Delta \in \mathcal{X}_\Delta, \\
& \text{and } X'_\Delta \in \mathcal{Z}_\Delta\}
\end{aligned} \tag{A.12}$$

Here $\mathcal{Z}_\Delta := \{z_{j_1}^* + \cdots + z_{j_\ell}^*\} - (\mathcal{Z}_{j_1} + \cdots + \mathcal{Z}_{j_\ell})$ is the set of the values that X'_Δ can take, and $0 \in \text{int}\mathcal{Z}_\Delta$ since $z_{j_m}^* \in \text{int}\mathcal{Z}_{j_m}$ for each m. Let

$$\begin{aligned}
G_1 :&= \left\{x_\Delta \mid \text{ there exists } x'_\Delta \text{ such that } (x_\Delta, x'_\Delta) \in G\right\} = \text{int}\mathcal{X}_\Delta \\
G_2 :&= \left\{x'_\Delta \mid \text{ there exists } x_\Delta \text{ such that } (x_\Delta, x'_\Delta) \in G\right\} \\
&= \text{int}\mathcal{X}_\Delta \cap \text{int}\mathcal{Z}_\Delta \\
G_3 :&= \left\{x''_\Delta \mid x''_\Delta = x_\Delta + x'_\Delta \text{ for some } (x_\Delta, x'_\Delta) \in G\right\} = \text{int}\mathcal{X}_\Delta.
\end{aligned}$$

Notice that $G_1 \cap G_2 \cap G_3 \neq \emptyset$. Thus, by Lemma A.1 and A.2 as in Case 1, there exists $\beta \in \mathbb{R}^{d_x}$ such that $f(x_\Delta) = \beta x_\Delta$ for every $x_\Delta \in \text{int}\mathcal{X}_\Delta$. Then, the construction of f (i.e., (A.10)) and the continuity of h^0 imply that $h^0(x) = (h^0(x_i^*) - \beta x_i^*) + \beta x$ for every $x \in \mathcal{X}_i^0$, which is a contradiction

to the assumption that h^0 is nonaffine on $\mathcal{X}_i^0$.

□

A.3 Proof of Theorem 2.2

We show the following lemma, which immediately implies Theorem 2.2.

Lemma A.4. *Suppose that Assumption 2.1, 2.2, 2.3, and 2.4 hold. We also assume that h^0 is affine on $\mathcal{X}_i^0$ for some $i \in \mathcal{N}^0$, that is, there exists $(\alpha_i, \beta_i) \in \mathbb{R}^{d_x+1}$ such that $h^0(x) = \alpha_i + \beta_i x$ for every $x \in \mathcal{X}_i^0$. If $(h, \gamma) \in \mathcal{H} \times \Gamma$ satisfies the conditional moment restrictions (2.5), then $\partial h(x) = \beta_i$ for every $x \in \mathcal{X}_k^\gamma$ with $k \in \mathcal{N}_\gamma$ such that $\gamma_j^0 = i$ and $\gamma_j = k$ for some $j \in \mathcal{J}$.*

Note that Assumption 2.6 implies $\beta_i = \beta^0$ for every $i \in \mathcal{N}^0$, since h^0 is affine on the entire $\mathcal{X}^0$ in that case. Note also that every $k \in \mathcal{N}_\gamma$ satisfies the condition in Lemma A.4 for some $i \in \mathcal{N}^0$, because $\{j \in \mathcal{J} : \gamma_j = k\} \neq \emptyset$. Thus, Lemma A.4 gives $\partial h(x) = \beta^0$ for every $x \in \bigcup_{k \in \mathcal{N}_\gamma} \mathcal{X}_k^\gamma = \mathcal{X}^\gamma$ under Assumption 2.6, which is the claim of Theorem 2.2.

Proof of Lemma A.4. We suppress the subscript t. Suppose that

$$\mathbb{E}\left[Y_k^\gamma - h(X_k^\gamma) \mid Z\right] = 0 \quad \text{(almost surely)}, \quad \text{for every } k \in \mathcal{N}_\gamma$$

holds for some $h \in \mathcal{H}$. Fix $k \in \mathcal{N}_\gamma$ such that $\gamma_j^0 = i$ and $\gamma_j = k$ for some $j \in \mathcal{J}$.

First, suppose $\gamma = \gamma^0$. Then $\{j \in \mathcal{J} : \gamma_j = k\} = \{j \in \mathcal{J} : \gamma_j^0 = i\}$ and hence $\mathcal{X}_k^\gamma = \mathcal{X}_i^0$. The conditional moment restriction above and Assumption 2.2 give $h(X_i^0) = \mathbb{E}[Y_i^0 \mid Z] = h^0(X_i^0)$ almost surely, so that $h = h^0$ on $\mathcal{X}_i^0$ by the continuity of h and h^0. Therefore, $\partial h(x) = \beta_i$ for every $x \in \mathcal{X}_k^\gamma = \mathcal{X}_i^0$.

Next, take any $\gamma \neq \gamma^0$. As in the proof of Lemma A.3, we have

$$\begin{aligned} &\sum_{k' \in \mathcal{N}_\gamma} h(Z_{1k'} + \cdots + Z_{ik'} + \cdots + Z_{Nk'}) \\ &= \{\alpha_i + \beta_i(Z_{i1} + \cdots + Z_{ik'} + \cdots + Z_{iN'})\} + \sum_{i' \in \mathcal{N}^0 \setminus \{i\}} h^0(Z_{i'1} + \cdots + Z_{i'k'} + \cdots + Z_{i'N'}), \end{aligned} \tag{A.13}$$

on $\mathrm{supp}(Z)$. Fix Z_j arbitrarily for each $j \in \mathcal{J}$ such that $\gamma_j^0 \neq i$. Then, the second term on the right hand side of (A.13) is a constant, and $X_{k'}^\gamma = Z_{ik'} + C_{k'}$ for each $k' \in \mathcal{N}_\gamma$, where $C_{k'} := \sum_{i' \in \mathcal{N}^0 \setminus \{i\}} Z_{i'k'}$

is a constant. Hence, (A.13) reduces to

$$\sum_{k' \in \mathcal{N}_\gamma} h(Z_{ik'} + C_{k'}) = \beta_i \sum_{k' \in \mathcal{N}_\gamma} Z_{ik'} + C, \tag{A.14}$$

where C does not depend on Z_j for any j such that $\gamma_j^0 = i$. Notice that $\{Z_{ik'}\}_{k' \in \mathcal{N}_\gamma}$ are sums of Z_j over mutually disjoint sets of blocks, and hence they vary independently over their supports by Assumption 2.4. Thus, differentiating both sides of (A.14) with respect to Z_{ik} gives $\partial h(Z_{ik}+C_k) = \beta_i$ for every $Z_{ik} \in \mathcal{Z}_{ik}$. Finally, C_k ranges over the support of $\sum_{j \in \mathcal{J}} \mathbb{1}(\gamma_j^0 \neq i \text{ and } \gamma_j = k) Z_j$ as we vary the blocks fixed above, and hence $Z_{ik} + C_k$ sweeps out the entire $\mathcal{X}_k^\gamma$ by Assumption 2.4. Therefore, $\partial h(x) = \beta_i$ for every $x \in \mathcal{X}_k^\gamma$. □

A.4 Proof of Theorem 2.3

Proof of Theorem 2.3. We suppress the subscript t. Suppose that

$$\mathbb{E}\left[Y_k^\gamma - h(X_k^\gamma) \mid Z\right] = 0 \quad \text{(almost surely)}, \quad \text{for every } k \in \mathcal{N}_\gamma$$

holds for some $h \in \mathcal{H}$ and $\gamma \in \Gamma$. For each $i \in \mathcal{N}^0$ such that h^0 is affine on $\mathcal{X}_i^0$, let $\beta_i \in \mathbb{R}^{d_x}$ denote the slope of h^0 on $\mathcal{X}_i^0$.

Take any $x \in \mathcal{X}^0 \cap \mathcal{X}^\gamma$. Then, there exist $i \in \mathcal{N}^0$ and $k \in \mathcal{N}_\gamma$ such that $x \in \mathcal{X}_i^0$ and $x \in \mathcal{X}_k^\gamma$. Since $\{j \in \mathcal{J} : \gamma_j = k\} \neq \emptyset$, we can take $j^* \in \mathcal{J}$ such that $\gamma_{j^*} = k$. Let $i' := \gamma_{j^*}^0 \in \mathcal{N}^0$. It suffices to consider the following three cases, which are exhaustive.

Case 1: h^0 is nonaffine on $\mathcal{X}_i^0$. Lemma A.3 implies that $h(x') = h^0(x')$ for every $x' \in \mathcal{X}_i^0$. Since $x \in \mathcal{X}_i^0$, we obtain $\partial h(x) = \partial h^0(x)$.

Case 2: h^0 is affine on $\mathcal{X}_i^0$ and nonaffine on $\mathcal{X}_{i'}^0$. Lemma A.3 implies that $\{j \in \mathcal{J} : \gamma_j^0 = i'\} = \{j \in \mathcal{J} : \gamma_j = k'\}$ for some $k' \in \mathcal{N}_\gamma$ and $h(x') = h^0(x')$ for every $x' \in \mathcal{X}_{i'}^0$. Since $\gamma_{j^*}^0 = i'$, we have $j^* \in \{j \in \mathcal{J} : \gamma_j = k'\}$, which gives $k' = \gamma_{j^*} = k$ and hence $\mathcal{X}_k^\gamma = \mathcal{X}_{i'}^0$. Since $x \in \mathcal{X}_k^\gamma = \mathcal{X}_{i'}^0$, we obtain $\partial h(x) = \partial h^0(x)$.

Case 3: h^0 is affine on both $\mathcal{X}_i^0$ and $\mathcal{X}_{i'}^0$. Since $\gamma_{j^*}^0 = i'$ and $\gamma_{j^*} = k$, Lemma A.4 applied to the true aggregated unit i' implies that $\partial h(x') = \beta_{i'}$ for every $x' \in \mathcal{X}_k^\gamma$, and in particular $\partial h(x) = \beta_{i'}$. By Assumption 2.7, we can take $\tilde{x} \in \operatorname{int}\left(\mathcal{X}_i^0 \cap \mathcal{X}_{i'}^0\right)$. Since h^0 is affine with slope β_i on $\mathcal{X}_i^0$ and with slope $\beta_{i'}$ on $\mathcal{X}_{i'}^0$, we have $\beta_i = \partial h^0(\tilde{x}) = \beta_{i'}$. Therefore, $\partial h(x) = \beta_{i'} = \beta_i = \partial h^0(x)$, where the last equality holds since $x \in \mathcal{X}_i^0$. □

A.5 Implication without the common support assumption

Recall that Theorem 2.1 relies on Assumption 2.5, that is, h^0 is nonaffine on $\mathcal{X}_i^0$ for every $i \in \mathcal{N}^0$. On the other hand, Theorem 2.2 relies on Assumption 2.6, that is, h^0 is affine on $\mathcal{X}^0$. Section 2.3 discusses implications of the two identification results under the common support assumption (i.e., $\mathcal{X}^0 = \mathcal{X}_i^0$ for every $i \in \mathcal{N}^0$). Note that, under the common support assumption, either Assumption 2.5 or Assumption 2.6 holds. Thus, Theorem 2.1 and 2.2 jointly imply that, if (h, γ) satisfies the conditional moment restrictions (2.5) and h is nonaffine on $\mathcal{X}_k^\gamma$ for some $k \in \mathcal{N}_\gamma$, then $\gamma = \gamma^0$ and $h(x) = h^0(x)$ for every $x \in \mathcal{X}^0$.

We can consider more general cases in which the support of the regressors vary across aggregated units. h^0 is (non)affine on the support of some true aggregated unit, but may not on the support of other units. In fact, Lemma A.3 and A.4 deal with such general cases. Lemma A.3 states that a true aggregated unit $i \in \mathcal{N}^0$ can be correctly delineated and h and h^0 coincide on $\mathcal{X}_i^0$ if h^0 is nonaffine on $\mathcal{X}_i^0$. Lemma A.4 claims that if h^0 is affine on $\mathcal{X}_i^0$ and its slope is β_i, then the derivative of h is also β_i on $\mathcal{X}_k^\gamma$ for every $k \in \mathcal{N}_\gamma$ such that k contains a block belonging to $i \in \mathcal{N}^0$ under the true delineation. As a result, Lemma A.3 and A.4 jointly imply that, if (h, γ) satisfies the conditional moment restrictions (2.5) and h is nonaffine on $\mathcal{X}_k^\gamma$ for some $k \in \mathcal{N}_\gamma$, then we find that the aggregated unit k is correctly delineated under γ and h coincides with h^0 on $\mathcal{X}_k^\gamma$. Note that this generalizes the discussion in Section 2.3.

B Proofs in Section 3

B.1 Proof of Proposition 3.1

Proof of Proposition 3.1. The proof is analogous to Newey, 1991 and Newey and Powell, 2003. First, we define the population objective function

$$Q(\theta,\gamma) = \frac{1}{|\mathcal{J}|}\sum_{j\in\mathcal{J}} \mathbb{E}\left[\bar{\rho}_j^{\gamma}(Z_t,\theta)^2\right],$$

where $\bar{\rho}_j^{\gamma}(Z_t,\theta) = \mathbb{E}\left[\rho_j^{\gamma}(W_t,Z_t,\theta) \mid Z_t\right]$, $\rho_j^{\gamma}(W_t,Z_t,\theta) = Y_{it}^{\gamma} - h(X_{it}^{\gamma},\theta)$, and $i := \gamma_j \in \mathcal{N}_{\gamma}$ for each $\gamma \in \Gamma$.

The sample objective function $\hat{Q}(\theta,\gamma)$ can be rewritten as

$$\hat{Q}(\theta,\gamma) = \frac{1}{|\mathcal{J}|}\sum_{j\in\mathcal{J}} \frac{1}{T}\left\{\hat{\mathbb{E}}\left[\rho_j^{\gamma}(W_t,Z_t,\theta) \mid Z_t\right]\right\}^2,$$

where

$$\hat{\mathbb{E}}\left[\rho_j^{\gamma}(W_t,Z_t,\theta) \mid Z_t\right] := b^K(Z_t)'(B'B)^{-}B'\rho_j^{\gamma}(\theta),$$

$B := (b^K(Z_1),\ldots,b^K(Z_T))'$ and $\rho_j^{\gamma}(\theta) = (\rho_j^{\gamma}(W_1,Z_1,\theta),\ldots,\rho_j^{\gamma}(W_T,Z_T,\theta))'$.

For each $\gamma \in \Gamma$, we define

$$Q_j(\theta;\gamma) := \mathbb{E}\left[\bar{\rho}_j^{\gamma}(Z_t,\theta)^2\right]$$

and

$$\hat{Q}_j(\theta;\gamma) := \frac{1}{T}\left\{\hat{\mathbb{E}}\left[\rho_j^{\gamma}(W_t,Z_t,\theta) \mid Z_t\right]\right\}^2.$$

so that

$$Q(\theta,\gamma) = \frac{1}{|\mathcal{J}|}\sum_{j\in\mathcal{J}} Q_j(\theta;\gamma)$$

and

$$\hat{Q}(\theta,\gamma) = \frac{1}{|\mathcal{J}|}\sum_{j\in\mathcal{J}} \hat{Q}_j(\theta;\gamma)$$

Notice that the conditional moment restrictions

$$\bar{\rho}_j^{\gamma}(Z_t,\theta) = 0 \quad \text{for every } j \in \mathcal{J} \tag{B.1}$$

are satisfied if and only if $\gamma = \gamma^0$ and $\theta = \theta^0$. To see why, take any $\gamma \neq \gamma^0$. Then, the conditional moment restrictions (B.1) cannot be satisfied for any $\theta \in \Theta$ by Theorem 2.1. Also, if $\gamma = \gamma^0$, the restrictions (B.1) can be satisfied only for $\theta = \theta^0$ by Assumption 3.1 (iii). Thus, $Q(\theta, \gamma) = (1/|\mathcal{J}|) \sum_{j \in \mathcal{J}} Q_j(\theta; \gamma) \geq 0$ and the equality holds if and only if $(\theta, \gamma) = (\theta^0, \gamma^0)$.

Fix $\theta \in \Theta$ and $\gamma \in \Gamma$ arbitrarily. Take any $j \in \mathcal{J}$. For notational simplicity, Let $u = \rho_j^\gamma(\theta)$ and $v = (\bar{\rho}_j^\gamma(Z_1, \theta), \ldots \bar{\rho}_j^\gamma(Z_T, \theta))$. We also define $\Omega = B(B'B)^- B'$. Then we have

$$\begin{aligned}\left|\hat{Q}_j(\theta; \gamma) - \left(v'v/T\right)\right| &= \left|\|\Omega u\|^2 - \|v\|^2\right| / T \\ &\leq \left(\|\Omega u - v\|^2 + 2\|v\| \cdot \|\Omega u - v\|\right) / T.\end{aligned}$$

Assumption 3.1 (iv), (v), and (vii) imply that there exists a sequence π_K such that

$$\begin{aligned}\mathbb{E}\left[\|\Omega u - v\|^2\right] =& \mathbb{E}\left[\mathbb{E}\left[\|\Omega u - v\|^2 \mid Z\right]\right] \\ =& \mathbb{E}\left[\mathbb{E}\left[\|\Omega(u-v) - (I-\Omega)v\|^2 \mid Z\right]\right] \\ =& \mathbb{E}\left[\operatorname{trace}\left\{\mathbb{E}\left[\Omega(u-v)(u-v)'\Omega \mid Z\right]\right\} + v'(I-\Omega)v\right] \\ =& \mathbb{E}\left[\operatorname{trace}\{\Omega\mathbb{E}\left[(u-v)(u-v)' \mid Z\right]\Omega\} + \{v - B\pi_K\}'(I-\Omega)\{v - B\pi_K\}\right] \\ \leq& C\,\mathbb{E}[\operatorname{trace}(\Omega)] + T\mathbb{E}\left[\left\{\bar{\rho}_j^\gamma(Z_t, \theta) - b^K(Z_t)'\pi_K\right\}^2\right] \\ =& O(K) + o(T),\end{aligned}$$

where $Z = (Z_1, \ldots, Z_T)$ and C is the constant in Assumption 3.1 (v). The inequality follows from $\operatorname{trace}\{\Omega\mathbb{E}[(u-v)(u-v)' \mid Z]\Omega\} \leq \lambda_{\max}(\mathbb{E}[(u-v)(u-v)' \mid Z])\operatorname{trace}(\Omega) \leq C\operatorname{trace}(\Omega)$ and $\operatorname{trace}(\Omega) \leq K$. Thus, we have $\|\Omega u - v\|^2/T = o_p(1)$ by the Markov inequality. Then, the weak law of large numbers gives $v'v/T = Q_j(\theta; \gamma) + o_p(1)$ under Assumption 3.1 (i) and (vii). Thus, $\hat{Q}_j(\theta; \gamma) \xrightarrow{p} Q_j(\theta; \gamma)$ holds by the triangle inequality, which implies the pointwise convergence of the objective function $\hat{Q}(\theta, \gamma) \xrightarrow{p} Q(\theta, \gamma)$.

Next, fix $\gamma \in \Gamma$ and $j \in \mathcal{J}$ arbitrarily. Define $M = \left(M\left(X_{i1}^\gamma\right), \ldots, M\left(X_{iT}^\gamma\right)\right)'$ where $i := \gamma_j$. By Assumption 3.1 (v), (vi) and (vii), it holds that $\|M\|^2/T = O_p(1)$ and $\left\|\rho_j^\gamma\left(\theta^0\right)\right\|^2/T = O_p(1)$ by Markov inequality. Thus, $\tilde{B}_T := \left(\|M\|^2 + 2\left\|\rho_j^\gamma\left(\theta^0\right)\right\| \cdot \|M\|\right)/T = O_p(1)$. Then, by Assumption

3.1 (ii) and (vi),

$$
\begin{aligned}
&\left|\hat{Q}_j(\tilde{\theta};\gamma) - \hat{Q}_j(\theta;\gamma)\right| \\
=&\mid \|\Omega\rho_j^\gamma(\tilde{\theta})\|^2 - \|\Omega\rho_j^\gamma(\theta)\|^2 \mid /T \\
=&\mid \|\Omega\rho_j^\gamma(\tilde{\theta})\| - \|\Omega\rho_j^\gamma(\theta)\| \mid \cdot \left(\|\Omega\rho_j^\gamma(\tilde{\theta})\| + \|\Omega\rho_j^\gamma(\theta)\|\right) /T \\
\leq&\|\rho_j^\gamma(\tilde{\theta}) - \rho_j^\gamma(\theta)\| \cdot \left(\|\rho_j^\gamma(\tilde{\theta})\| + \|\rho_j^\gamma(\theta)\|\right) /T \\
\leq&\left(\|\rho_j^\gamma(\tilde{\theta}) - \rho_j^\gamma(\theta)\|^2 + 2\|\rho_j^\gamma(\theta)\| \cdot \|\rho_j^\gamma(\tilde{\theta}) - \rho_j^\gamma(\theta)\|\right) /T \\
=&\left(\|\rho_j^\gamma(\tilde{\theta}) - \rho_j^\gamma(\theta)\|^2 + 2\|\rho_j^\gamma(\theta^0) + \rho_j^\gamma(\theta) - \rho_j^\gamma(\theta^0)\| \cdot \|\rho_j^\gamma(\tilde{\theta}) - \rho_j^\gamma(\theta)\|\right) /T \\
\leq&\left(\|\rho_j^\gamma(\tilde{\theta}) - \rho_j^\gamma(\theta)\|^2 + 2\left(\|\rho_j^\gamma(\theta^0)\| + \|\rho_j^\gamma(\theta) - \rho_j^\gamma(\theta^0)\|\right) \cdot \|\rho_j^\gamma(\tilde{\theta}) - \rho_j^\gamma(\theta)\|\right) /T \\
\leq&\left\{\|M\|^2 g(\|\tilde{\theta} - \theta\|) + 2\left\|\rho_j^\gamma\left(\theta^0\right)\right\| \cdot \|M\| + 2\|M\|^2 g(\|\theta - \theta^0\|)\right\} \times g(\|\tilde{\theta} - \theta\|)/T \\
\leq&C\tilde{B}_T g(\|\tilde{\theta} - \theta\|) = B_T g(\|\tilde{\theta} - \theta\|),
\end{aligned}
$$

where C is some constant and $B_T := C\tilde{B}_T = O_p(1)$. The second inequality above uses the reverse triangle inequality and the fact that Ω is an orthogonal projection matrix so that $\|\Omega a\| \leq \|a\|$ for any a.

Thus, Corollary 2.2 in Newey, 1991 gives $\sup_{\theta\in\Theta} |Q_j(\theta;\gamma) - \hat{Q}_j(\theta;\gamma)| = o_p(1)$ and $Q_j(\theta;\gamma)$ is continuous, which implies the uniform convergence $\sup_{(\theta,\gamma)\in\Theta\times\Gamma} |Q(\theta,\gamma) - \hat{Q}(\theta,\gamma)| = o_p(1)$ and continuity of $Q(\theta,\gamma)$ since Γ is finite. Then, by Theorem 2.1 in Newey and McFadden, 1994, the consistency result follows. □

B.2 Proof of Proposition 3.2

Proof of Proposition 3.2. **Step 1:** First, define

$$\bar{Q}(\theta;\gamma) := \frac{1}{|\mathcal{N}_\gamma|} \sum_{i\in\mathcal{N}_\gamma} \bar{Q}_i(\theta;\gamma),$$

where

$$\bar{Q}_i(\theta;\gamma) := \mathbb{E}\left[\{Y_{it}^\gamma - h\left(X_{it}^\gamma, \theta\right)\}^2\right].$$

Note that $\bar{Q}(\theta;\gamma^0)$ is uniquely minimized at θ^0 by Assumption 3.1 (iii). Next we have

$$\begin{aligned}
&\mathbb{E}\left[\{Y_{it}^0 - h\left(X_{it}^0,\theta\right)\}^2\right] \\
=&\mathbb{E}\left[\{Y_{it}^0 - h\left(X_{it}^0,\theta^0\right)\}^2 + 2\{h(X_{it}^0,\theta^0) - h(X_{it}^0,\theta)\}\{Y_{it}^0 - h\left(X_{it}^0,\theta^0\right)\} + \{h(X_{it}^0,\theta^0) - h(X_{it}^0,\theta)\}^2\right] \\
\leq&\mathbb{E}\left[\{Y_{it}^0 - h\left(X_{it}^0,\theta^0\right)\}^2 + 2M(X_{it}^0)g(\|\theta-\theta^0\|)\left|Y_{it}^0 - h\left(X_{it}^0,\theta^0\right)\right| + M(X_{it}^0)^2 g(\|\theta-\theta^0\|)^2\right],
\end{aligned}$$

which implies that, by Assumption 3.1 (ii), (v) and (vi), there exists $d(Y_{it}^0, X_{it}^0)$ such that $\{Y_{it}^0 - h\left(X_{it}^0,\theta\right)\}^2 \leq d(Y_{it}^0, X_{it}^0)$ for all $\theta \in \Theta$ and $\mathbb{E}[d(Y_{it}^0, X_{it}^0)] < \infty$. Then, under Assumption 3.1 (i), (ii), and Assumption 3.2 (ii), Lemma 2.4 in Newey and McFadden, 1994 (and its extension to strictly stationary and ergodic data) implies that $\sup_{\theta\in\Theta}\|\tilde{Q}(\theta;\gamma^0) - \bar{Q}(\theta;\gamma^0)\| \xrightarrow{p} 0$ and $\bar{Q}(\theta;\gamma^0)$ is continuous. Since $\mathbb{P}(\hat{\gamma} = \gamma^0) \to 1$, we have

$$\sup_{\theta\in\Theta}\|\tilde{Q}(\theta;\hat{\gamma}) - \bar{Q}(\theta;\gamma^0)\| = \sup_{\theta\in\Theta}\|\tilde{Q}(\theta;\gamma^0) - \bar{Q}(\theta;\gamma^0)\| + o_p(1) \xrightarrow{p} 0.$$

Hence, by Theorem 2.1 in Newey and McFadden, 1994, $\tilde{\theta} \xrightarrow{p} \theta^0$ as $T \to \infty$.

Step 2: Define

$$\Lambda(\theta) = \frac{1}{|\mathcal{N}^0|}\sum_{i\in\mathcal{N}^0}\Lambda_i(\theta)$$

where

$$\Lambda_i(\theta) = \mathbb{E}\left[\partial_\theta h\left(X_{it}^0,\theta\right)\partial_\theta h\left(X_{it}^0,\theta\right)' - \left\{Y_{it}^0 - h\left(X_{it}^0,\theta\right)\right\}\partial_{\theta\theta} h\left(X_{it}^0,\theta\right)\right].$$

Similarly, define

$$\tilde{\Lambda}(\theta) = \frac{1}{|\mathcal{N}^0|}\sum_{i\in\mathcal{N}^0}\tilde{\Lambda}_i(\theta)$$

where

$$\tilde{\Lambda}_i(\theta) = \frac{1}{T}\sum_{t=1}^{T}\left[\partial_\theta h\left(X_{it}^0,\theta\right)\partial_\theta h\left(X_{it}^0,\theta\right)' - \left\{Y_{it}^0 - h\left(X_{it}^0,\theta\right)\right\}\partial_{\theta\theta} h\left(X_{it}^0,\theta\right)\right].$$

Notice that $\partial_{\theta\theta}\tilde{Q}(\theta;\gamma^0) = 2\tilde{\Lambda}(\theta)$ and $\Lambda(\theta^0) = \Lambda$, where the latter follows since $\mathbb{E}[\varepsilon_{it}\partial_{\theta\theta}h(X_{it}^0,\theta^0)] = 0$ under Assumption 2.2. Then, under Assumption 3.1 (i) and Assumption 3.2 (iii), we have

$$\sup_{\theta\in\Theta(\theta^0)}\left\|\partial_{\theta\theta}\tilde{Q}(\theta;\gamma^0) - 2\Lambda(\theta)\right\| \xrightarrow{p} 0$$

and $\Lambda(\theta)$ is continuous in $\Theta(\theta^0)$. Since $\mathbb{P}(\hat{\gamma} = \gamma^0) \to 1$, we have

$$\sup_{\theta \in \Theta(\theta^0)} \left\| \partial_{\theta\theta} \tilde{Q}\left(\theta; \hat{\gamma}\right) - 2\Lambda(\theta) \right\| = \sup_{\theta \in \Theta(\theta^0)} \left\| \partial_{\theta\theta} \tilde{Q}\left(\theta; \gamma^0\right) - 2\Lambda(\theta) \right\| + o_p(1) \xrightarrow{p} 0.$$

Moreover, Assumption 3.2 (iv) gives

$$\sqrt{|\mathcal{N}^0|T}\, \partial_\theta \tilde{Q}(\theta^0; \gamma^0) = -\frac{2}{\sqrt{|\mathcal{N}^0|T}} \sum_{i \in \mathcal{N}^0} \sum_{t=1}^{T} \partial_\theta h(X_{it}^0, \theta^0) \varepsilon_{it} \xrightarrow{d} N(0, 4V).$$

Then, by Theorem 3.1 in Newey and McFadden, 1994, we have

$$\sqrt{|\mathcal{N}_{\hat{\gamma}}|T} \left(\tilde{\theta} - \theta^0\right) = \sqrt{|\mathcal{N}^0|T} \left(\tilde{\theta} - \theta^0\right) + o_p(1) \xrightarrow{d} N\left(0, \Lambda^{-1} V \Lambda^{-1}\right).$$

□