

Overidentification in Shift-Share Designs

Jinyong Hahn
UCLA
hahn@econ.ucla.edu

Guido Kuersteiner
University of Maryland
gkuerte@umd.edu

Andres Santos
UCLA
andres@econ.ucla.edu

Wavid Willigrod
UCLA
dwwilligrod@gmail.com

April 25, 2024

Abstract

This paper studies the testability of identifying restrictions commonly employed to assign a causal interpretation to two stage least squares (TSLS) estimators based on Bartik instruments. For homogeneous effects models applied to short panels, our analysis yields testable implications previously noted in the literature for the two major available identification strategies. We propose overidentification tests for these restrictions that remain valid in high dimensional regimes and are robust to heteroskedasticity and clustering. We further show that homogeneous effect models in short panels, and their corresponding overidentification tests, are of central importance by establishing that: (i) In heterogeneous effects models, interpreting TSLS as a positively weighted average of treatment effects can impose implausible assumptions on the distribution of the data; and (ii) Alternative identifying strategies relying on long panels can prove uninformative in short panel applications. We highlight the empirical relevance of our results by examining the viability of Bartik instruments for identifying the effect of rising Chinese import competition on US local labor markets.

1 Introduction

Shift-share designs typically rely on linear combinations of unit specific variables (the “shares”) and aggregate level variables (the “shocks”) as instruments to obtain identification. Originally employed in the work of [Bartik \(1991\)](#) and [Blanchard and Katz \(1992\)](#), these instruments have become known in the literature as “Bartik” instruments. Bartik instruments have proven to be remarkably versatile yielding insights into, among others, the impact of immigration on labor markets ([Card, 2001](#)), the consequence of trade liberalization for poverty ([Topalova, 2010](#)), the effect of import competition on labor markets ([Autor et al., 2013](#)), and the welfare implications of geographic sorting by education ([Diamond, 2016](#)).

In this paper, we examine the testable implications of identifying restrictions commonly employed to assign a causal interpretation to two stage least squares (TSLS) estimators based on Bartik instruments. We largely focus our analysis on two recent complementary identification strategies. The first, studied by [Goldsmith-Pinkham et al. \(2020\)](#), attributes the exogeneity of the Bartik instrument to the exogeneity of the shares. The second, examined by [Adao et al. \(2019\)](#) and [Borusyak et al. \(2022\)](#), instead attributes the exogeneity of the Bartik instrument to the exogeneity of aggregate shocks. We systematically study the overidentifying content of both approaches by noting that their differing asymptotic frameworks either implicitly or explicitly require *conditional* moment restrictions to hold. Specifically, [Goldsmith-Pinkham et al. \(2020\)](#) necessitates the Bartik instrument to be exogenous conditional on the realization of the aggregate shocks, while [Adao et al. \(2019\)](#) and [Borusyak et al. \(2022\)](#) require the Bartik instrument to be exogenous conditional on the realization of the shares (among other variables).

The conditional moment restrictions implied by the different identification strategies yield previously noted overidentifying restrictions in homogeneous treatment effects models. For instance, when applied to the framework of [Goldsmith-Pinkham et al. \(2020\)](#), our analysis implies that the entire vector of shares must be exogenous. As noted by [Goldsmith-Pinkham et al. \(2020\)](#), the model is therefore overidentified because any deterministic linear combination of the shares yields a valid instrument. In contrast, when applied to the framework of [Adao et al. \(2019\)](#) and [Borusyak et al. \(2022\)](#), our results imply that the conditional mean of the Bartik instrument given controls, shares, and residuals must be a linear function of the controls only. As a result, the model is overidentified because any appropriately centered deterministic linear combination of the aggregate shocks yields a valid instrument. Our impression is that the latter overidentifying restrictions have not received the same level of attention by practitioners as the testable implications of [Goldsmith-Pinkham et al. \(2020\)](#).

Building on our overidentification analysis, we develop a framework for testing the overidentifying restrictions corresponding to both identification strategies. In many

of the applications that motivate us, the number of overidentifying restrictions can be “large” relative to the sample – e.g., in many applications the numbers of sectors exceeds the number of clusters. We therefore focus on developing tests that remain valid in such high dimensional settings by employing the high dimensional central limit theorem of [Chernozhukov et al. \(2022\)](#). Our tests utilize bootstrap based critical values and are robust to the presence of heteroskedasticity, clustering, and weighting.¹

A natural way to proceed, for instance after rejecting the validity of the homogeneous effects model, is to consider weaker assumptions that still enable us to attribute a causal interpretation to TSLS. To this end, we study a generalization of the model and an alternative identification strategy, but find both approaches to be potentially empirically limited in scope. As a generalization of the model, we consider linear heterogeneous treatment effects models previously employed in the literature ([Goldsmith-Pinkham et al., 2020](#); [Adao et al., 2019](#)). In such a setting, TSLS can readily be shown to estimate a weighted average of group specific treatment effects. As forcibly argued by [Kolesar \(2013\)](#) and [Blandhol et al. \(2022\)](#) among others, however, the averaging weights should be positive in order to attribute a causal interpretation to TSLS. Unfortunately, we find that in the case of Bartik instruments the weights can naturally be negative. In particular, within the identification framework of [Goldsmith-Pinkham et al. \(2020\)](#), we show that a necessary condition for the weights to be positive is that shares of different sectors be uncorrelated with each other – a requirement that a-fortiori fails when shares sum up to one. Conversely, within the identification framework of [Adao et al. \(2019\)](#) and [Borusyak et al. \(2022\)](#), we show that a necessary condition for the weights to be positive is that aggregate shocks to different sectors be (weakly) positively correlated with each other.

As an alternative identification strategy, we consider the possibility that the exogeneity of the instrument is not justified by the shares alone ([Goldsmith-Pinkham et al., 2020](#)) or the aggregate shocks alone ([Adao et al., 2019](#); [Borusyak et al., 2022](#)) but by the shares and shocks together. Formally, such an identification strategy corresponds to requiring that the Bartik instrument be uncorrelated with the error term when expectations are evaluated over both the time series distribution of aggregate shocks and the cross sectional distribution of shares. While this identification strategy renders the model “just identified,” it also necessitates a long panel in order to estimate the time series distribution of the aggregate shocks. We show that as a result, the effective number of observations for computing TSLS standard errors is governed by the number of time periods in the panel. Because the majority of empirical studies relying on Bartik instruments have relied on short panels, we expect TSLS to be statistically uninformative under this identification strategy. Our asymptotic analysis relies on a novel simulta-

¹In contrast, [Goldsmith-Pinkham et al. \(2020\)](#) employ overidentification tests based on LIML (which requires homoskedasticity) and [Chao et al. \(2014\)](#) (which precludes clustering). Neither test allows the number of restrictions to exceed the sample size.

neous time series and cross sectional study of long panels. These results significantly extend related work in [Hahn et al. \(2019, 2022\)](#) and may be of independent interest.

We highlight the empirical relevance of our analysis by revisiting the study by [Autor et al. \(2013\)](#) on the impact of rising Chinese import competition on local US labor markets. In this setting, our overidentification tests find evidence against the validity of the identification framework of [Goldsmith-Pinkham et al. \(2020\)](#) as well as the identification framework of [Adao et al. \(2019\)](#) and [Borusyak et al. \(2022\)](#). Moreover, since shares are empirically correlated across sectors, our analysis further suggests that TSLS may not have a causal interpretation in a heterogeneous effects model under the identification strategy of [Goldsmith-Pinkham et al. \(2020\)](#). Likewise, TSLS may not possess a causal interpretation in a heterogeneous effects model under the identification strategy of [Adao et al. \(2019\)](#) and [Borusyak et al. \(2022\)](#) when shocks within clusters are negatively correlated. Finally, we note that applying the proposed alternative just identified long panel identification strategy is not viable in this application because there are only two time periods.

The rest of the paper is organized as follows. Section 2 characterizes the overidentifying restrictions implied by the identification strategies of [Goldsmith-Pinkham et al. \(2020\)](#) as well [Borusyak et al. \(2022\)](#) and [Adao et al. \(2019\)](#). Section 3 examines the scope for attributing a causal interpretation to TSLS under heterogeneous effects models or long panel identification strategies. Our overidentification tests are developed in Section 4, while Section 5 illustrates the relevance of our analysis by revisiting [Autor et al. \(2013\)](#). Section 6 briefly concludes by providing recommendations for empirical practice. A series of appendices contain proofs of our results and a Monte Carlo study.

2 Overidentification

Following the main instrumental variables specifications of [Goldsmith-Pinkham et al. \(2020\)](#) and [Adao et al. \(2019\)](#), we begin by studying homogeneous treatment effects models in which the parameter of interest is common across all individuals. In the next section, we will examine heterogeneous treatment effects models instead. For ease of exposition, we focus on cross sectional applications though note that an extension to short panel data settings is immediate; see, e.g., [Remarks 2.1](#) and [2.2](#) below.

In what follows we let $Y_i \in \mathbb{R}$, $X_i \in \mathbb{R}$, $Z_i \in \mathbb{R}$, and $W_i \in \mathbb{R}^d$ denote the outcome variable, a scalar regressor of interest, a scalar instrument, and a vector of controls. We will refer to each “ i ” as an individual, though note that in applications i may represent, for example, a location. The variables (Y_i, X_i, W_i) are assumed to satisfy

$$Y_i = X_i\beta + W_i'\gamma_s + \varepsilon_i, \tag{1}$$

where the restrictions on ε_i will be stated shortly. We will further require that the instrument Z_i follow a Bartik structure in the sense that it satisfies

$$Z_i = \mathcal{Z}' S_i$$

with $\mathcal{Z} \in \mathbb{R}^p$ an aggregate variable and $S_i \in \mathbb{R}^p$ an individual specific variable.

For instance, in canonical applications \mathcal{Z} equals a vector of aggregate industry specific shocks and S_i equals the share of each industry in the economy of location i . For this reason, in what follows we refer to \mathcal{Z} as “shocks” and S_i as “shares.” Other variables, such as ε_i , may have a Bartik structure as well though we only make the Bartik structure explicit for Z_i . Finally, we complement equation (1) with the first stage equation

$$X_i = Z_i \lambda + W_i' \gamma_{\mathbf{f}} + \eta_i. \quad (2)$$

Here, the \mathbf{f} and \mathbf{s} subscripts on γ refer to the first and the second stage respectively. While we let the first stage coefficients in (2) be fixed for simplicity, it is worth emphasizing that all the results in this section continue to hold if the first stage coefficients are instead random as in [Adao et al. \(2019\)](#). The observable variables are the aggregate shocks \mathcal{Z} and the cross sectional variables $\{Y_i, X_i, W_i, S_i\}_{i=1}^n$.

Different asymptotic frameworks either implicitly or explicitly condition on different variables when delivering asymptotic promises. To reflect these differences, it will prove convenient to introduce the notation \mathcal{G}_n to denote the relevant conditioning variables. For instance, we will argue below that asymptotic approximations that rely on only the cross section being large implicitly condition on aggregate variables such as the shocks \mathcal{Z} . Hence, in such instances we would set \mathcal{G}_n to include all aggregate variables, including \mathcal{Z} . Alternatively, when specialized to equations (1) and (2), the asymptotic framework of [Adao et al. \(2019\)](#) and [Borusyak et al. \(2022\)](#) explicitly conditions on the cross-sectional variables $\{W_i, S_i, \varepsilon_i\}_{i=1}^n$ so we would in that case simply set \mathcal{G}_n to equal $\{W_i, S_i, \varepsilon_i\}_{i=1}^n$.

We will first examine conditions under which the two stage least squares (TSLS) estimand equals β in the homogeneous effects model defined by (1) and (2). To this end, it is helpful to define the “population” residualized instrument $\dot{Z}_{i,n}$ according to

$$\dot{Z}_{i,n} \equiv Z_i - W_i' \pi_n \quad \pi_n \equiv \left(\frac{1}{n} \sum_{i=1}^n E [W_i W_i' | \mathcal{G}_n] \right)^{-1} \left(\frac{1}{n} \sum_{i=1}^n E [W_i Z_i | \mathcal{G}_n] \right), \quad (3)$$

which allows us to define the TSLS estimand $\beta_{0,n}$ as the solution to the equation

$$\frac{1}{n} \sum_{i=1}^n E \left[(Y_i - X_i \beta_{0,n}) \dot{Z}_{i,n} \middle| \mathcal{G}_n \right] = 0. \quad (4)$$

We note that in the notation we have let $\beta_{0,n}$ depend on n to reflect that the TSLS

estimand may depend on n through the conditioning on \mathcal{G}_n . Under an appropriate rank condition, plugging the outcome equation (1) into the moment condition in (4) readily yields that the estimand $\beta_{0,n}$ equals β if and only if we have

$$\frac{1}{n} \sum_{i=1}^n E [\varepsilon_i (Z_i - W_i' \pi_n) | \mathcal{G}_n] = 0. \quad (5)$$

We next examine the implications of this exogeneity condition under the different choices of conditioning set \mathcal{G}_n that correspond to different asymptotic approximations.

2.1 Conditioning on Shocks

Traditionally, empirical work employing shift-share designs has reported standard errors that are motivated by asymptotic approximations in which only the cross section becomes large (i.e. $n \rightarrow \infty$). Probabilistic statements associated with these asymptotic approximations correspond to a thought experiment in which we only re-sample the individual specific variables $\{Y_i, X_i, W_i, S_i\}_{i=1}^n$. For instance, in this context, the level of a confidence interval for β refers to the probability that a randomly drawn sample $\{Y_i, X_i, W_i, S_i\}_{i=1}^n$ yields a confidence interval that indeed includes β .

Probabilistic statements based on re-sampling only individual specific variables a fortiori keep aggregate variables such as the shocks \mathcal{Z} fixed. Consequently, in an asymptotic approximation in which only n becomes large we are only able to identify the distribution of individual specific variables *conditionally* on the realizations of aggregate variables such as \mathcal{Z} . When evaluating the exogeneity condition on our instrument (i.e. restriction (5)) under this asymptotic approximation, we should therefore view all aggregate variables as belonging to the conditioning set \mathcal{G}_n .

In order to illustrate the implications of these observations, we will for simplicity assume that the variables $\{S_i, W_i, \varepsilon_i\}_{i=1}^n$ are i.i.d. and independent of all aggregate variables.² Provided that the controls W_i are uncorrelated with the error ε_i , it then follows from a bit of algebra that the exogeneity condition in (5) is equivalent to

$$\mathcal{Z}' E [S_i \varepsilon_i] = 0. \quad (6)$$

Equation (6) highlights that the relevant exogeneity condition for this asymptotic framework depends on the realization of the aggregate shocks \mathcal{Z} and the correlation between the shares S_i and the error ε_i . Lacking a justification as to why (6) should hold at the *actual* realization of \mathcal{Z} and not others, we should in the interest of robustness demand that (6) hold at *all* possible realizations of \mathcal{Z} . Provided the aggregate shocks \mathcal{Z} exhibit

²This setting rules out, for example, that (W, ε) have a Bartik structure as in [Adao et al. \(2019\)](#). We note, however, that the main points made in this section carry over if we allow (W, ε) to have a Bartik structure as well, though such an extension requires additional notation and assumptions.

sufficient variation, however, it then follows that condition (6) can hold for all possible realizations of \mathcal{Z} if and only if in fact S_i itself is uncorrelated with the errors ε_i .

Our next simple proposition formalizes the preceding discussion.

Proposition 2.1. *Suppose that $\mathcal{Z} \in \mathcal{G}_n$ and that $\{S_i, W_i, \varepsilon_i\}_{i=1}^n$ are i.i.d. and independent of \mathcal{G}_n . If $E[W\varepsilon] = 0$ and the support of $\mathcal{Z} \in \mathbb{R}^p$ has dimension p , then*

$$\frac{1}{n} \sum_{i=1}^n E[\varepsilon_i(Z_i - W_i' \pi_n) | \mathcal{G}_n] = 0 \quad (7)$$

with probability one (over \mathcal{Z}) if and only if $E[S_i \varepsilon_i] = 0$.

Proposition 2.1 establishes that the entire vector of shares S_i must be a valid instrument in order for the scalar instrument Z_i itself to be valid. This conclusion echoes arguments in Goldsmith-Pinkham et al. (2020), who similarly argue that the exogeneity of the instrument Z_i should be understood in terms of the exogeneity of the shares S_i . The principal implication of Proposition 2.1 for our purposes is that the validity of the Bartik instrument renders the model overidentified. For instance, it follows that we may construct an overidentification test by employing the sample moments

$$\frac{1}{n} \sum_{i=1}^n S_i \hat{\varepsilon}_i,$$

where $\{\hat{\varepsilon}_i\}_{i=1}^n$ denotes the fitted residuals from the estimated model. One complication that arises in building overidentification tests from these moments is that the number of moments (i.e. the number of shares) is often too “large” for standard asymptotic approximations to remain accurate. In Section 4, we address this challenge by developing overidentification tests that rely on high dimensional asymptotics instead.

Remark 2.1. While we have focused on a cross-sectional setting for ease of exposition, our conclusions readily extends to panel data applications. Specifically, suppose there are $1 \leq t \leq T$ time periods, each with an aggregate shock \mathcal{Z}_t and a cross-section $\{Y_{it}, X_{it}, S_{it}, W_{it}\}_{i=1}^n$. Under a short panel asymptotic approximation (i.e. $n \rightarrow \infty$ and T fixed), conditions similar to those employed in Proposition 2.1 readily imply that as the relevant exogeneity condition we should require that

$$E[S_{it} \varepsilon_{it}] = 0 \quad \text{for } 1 \leq t \leq T.$$

Thus, in this context, the time dimension yields additional overidentifying restrictions to those available in the cross sectional setting. ■

2.2 Identification Through Shocks

As an alternative identification strategy, [Adao et al. \(2019\)](#) and [Borusyak et al. \(2022\)](#) advocate interpreting the exogeneity of the Bartik instrument Z_i as originating from the exogeneity of the aggregate shocks \mathcal{Z} . Within our context, the assumptions employed by their asymptotic approximations correspond to setting \mathcal{G}_n to equal $\{S_i, W_i, \varepsilon_i\}_{i=1}^n$ and letting the number of shocks p increase with the sample size n .³ Under this asymptotic framework, the relevant exogeneity condition (i.e. restriction (5)) is equivalent to

$$\frac{1}{n} \sum_{i=1}^n \varepsilon_i (S_i' E[\mathcal{Z} | \mathcal{G}_n] - W_i' \pi_n) = 0. \quad (8)$$

To gain intuition into this exogeneity requirement, it is instructive to consider the case in which there are no controls W_i . Letting \mathcal{Z}_j denote the j^{th} coordinate of the vector of aggregate shocks $\mathcal{Z} \in \mathbb{R}^p$ and S_{ij} the j^{th} coordinate of the shares $S_i \in \mathbb{R}^p$, it then follows that the exogeneity condition in (8) simplifies to the expression

$$\sum_{j=1}^p \left(E[\mathcal{Z}_j | \mathcal{G}_n] \times \left(\frac{1}{n} \sum_{i=1}^n \varepsilon_i S_{ij} \right) \right) = 0. \quad (9)$$

Moreover, since the moment restriction in (9) must hold for any possible realization of the sample $\{S_i, \varepsilon_i\}_{i=1}^n$, the exogeneity condition in this context implies, under appropriate assumptions, that the aggregate shocks $\{\mathcal{Z}_j\}_{j=1}^p$ must have mean zero and be mean independent of the individual specific variables $\{S_i, \varepsilon_i\}_{i=1}^n$. However, if all the aggregate shock $\{\mathcal{Z}_j\}_{j=1}^p$ are mean independent of the individual specific variables, then the instrument $Z_i = S_i' \mathcal{Z}$ must itself be mean independent of $\{S_i, \varepsilon_i\}_{i=1}^n$ because

$$E[Z_i | \{S_i, \varepsilon_i\}_{i=1}^n] = \sum_{j=1}^p E[\mathcal{Z}_j | \mathcal{G}_n] S_{ij} = 0.$$

Intuitively, if the aggregate shocks $\{\mathcal{Z}_j\}_{j=1}^p$ are exogenous in the sense that (9) holds, then any suitable linear combination of the shocks, such as $Z_i = S_i' \mathcal{Z}$, must be mean independent of $\{S_i, \varepsilon_i\}_{i=1}^n$ as well. In particular, we obtain the overidentifying restriction that Z_i must be uncorrelated with any functions of the shares S_i and the error ε_i .⁴

While the inclusion of controls W_i allows us to relax the requirements on the aggregate shocks \mathcal{Z} , the model remains overidentified through its implications on the conditional mean of the instrument Z_i . Our next proposition illustrates this conclusion in the simple setting in which \mathcal{Z} is independent of all individual specific variables.

³As noted by [Adao et al. \(2019\)](#), including $\{\varepsilon_i\}_{i=1}^n$ in the conditioning set is important for obtaining asymptotically valid standard errors in applications in which $\{\varepsilon_i\}_{i=1}^n$ possesses a Bartik structure.

⁴Equivalently, note that since any suitably linear combination of \mathcal{Z} is also a valid instrument, it follows that there is a surplus of instruments and hence that the model is overidentified.

Proposition 2.2. *Suppose $\mathcal{G}_n = \{S_i, W_i, \varepsilon_i\}_{i=1}^n$ and \mathcal{Z} is independent of \mathcal{G}_n . If the support of $\{\varepsilon_i\}_{i=1}^n$ conditional on $\{S_i, W_i\}_{i=1}^n$ has dimension n , then*

$$\frac{1}{n} \sum_{i=1}^n E[\varepsilon_i(Z_i - W_i' \pi_n) | \mathcal{G}_n] = 0 \quad (10)$$

with probability one (over $\{S_i, W_i, \varepsilon_i\}_{i=1}^n$) if and only if $E[Z_i | \{S_i, W_i, \varepsilon_i\}_{i=1}^n] = W_i' \pi_n$.

Proposition 2.2 establishes that the exogeneity requirement needed to obtain identification through the aggregate shocks \mathcal{Z} effectively imposes that the conditional mean of the instrument Z_i given $\{S_i, W_i, \varepsilon_i\}_{i=1}^n$ be equal to the linear projection of Z_i onto W_i . This conclusion can be shown to hold even if \mathcal{Z} is not independent of the individual specific variables $\{S_i, W_i, \varepsilon_i\}_{i=1}^n$ at the cost of additional notation and assumptions. For our purposes, the principal implication of Proposition 2.2 is that the residualized instrument $(Z_i - W_i' \pi_n)$ must be uncorrelated with any function of the shares S_i , controls W_i , and errors ε_i . This observation suggests employing the sample moments

$$\frac{1}{n} \sum_{i=1}^n g(\hat{\varepsilon}_i, W_i, S_i)(Z_i - W_i' \hat{\pi}_n)$$

to construct an overidentification test – here g is an arbitrary vector valued function and $\{Z_i - W_i' \hat{\pi}_n\}_{i=1}^n$ denotes the residuals from regressing the instrument on the controls. In Section 4 we will develop such an overidentification test while allowing the number of sample moments to potentially be high dimensional.

Remark 2.2. The overidentifying restrictions of Proposition 2.2 generalize to panel data applications. Specifically, suppose that at each time period $1 \leq t \leq T$ we observe a shock $\mathcal{Z}_t \in \mathbb{R}^p$ and individual specific variables $\{Y_{it}, X_{it}, S_{it}, W_{it}\}_{i=1}^n$. In this context, the asymptotic framework in Adao et al. (2019) and Borusyak et al. (2022) corresponds to setting \mathcal{G}_n to equal $\{S_{it}, W_{it}, \varepsilon_{it}\}_{i=1, t=1}^{n, T}$ and letting T be fixed while n and p become large. Under conditions similar to those imposed in Proposition 2.2, it follows that the exogeneity requirements imposed on the shocks $\{\mathcal{Z}_t\}_{t=1}^T$ to obtain identification imply

$$E[Z_{it} | \mathcal{G}_n] = W_{it}' \pi_n \quad \pi_n \equiv \left(\sum_{i=1}^n \sum_{t=1}^T W_{it} W_{it}' \right)^{-1} \left(\sum_{i=1}^n \sum_{t=1}^T W_{it} E[Z_{it} | \mathcal{G}_n] \right).$$

As a result, the model is now overidentified by the restriction that the residualized instrument $(Z_{it} - W_{it}' \pi_n)$ be uncorrelated with any function of $\{W_{it}, S_{it}, \varepsilon_{it}\}_{t=1}^T$. ■

Remark 2.3. Adao et al. (2019) and Borusyak et al. (2022) impose sufficient conditions for the overidentifying restrictions derived in Proposition 2.2 to hold. For instance, Adao et al. (2019) require that the controls satisfy $W_i = \mathcal{W}' S_i$, with \mathcal{W} a $p \times d$ matrix of aggregate shocks, and impose that $E[\mathcal{Z} | \{S_i, W_i, \varepsilon_i\}_{i=1}^n, \mathcal{W}] = \mathcal{W} \Gamma$ for some vector

$\Gamma \in \mathbb{R}^d$.⁵ Under this structure, it follows that π_n equals Γ and hence that

$$\begin{aligned} & E[Z_i | \{S_i, W_i, \varepsilon_i\}_{i=1}^n] \\ &= E[S_i' E[Z | \{S_i, W_i, \varepsilon_i\}_{i=1}^n, \mathcal{W}] | \{S_i, W_i, \varepsilon_i\}_{i=1}^n] = E[S_i' \mathcal{W} \Gamma | \{S_i, W_i, \varepsilon_i\}_{i=1}^n] = W_i' \pi_n, \end{aligned}$$

which implies the overidentifying restrictions of Proposition 2.2 are indeed satisfied. ■

3 Just Identification?

We have so far shown that the homogeneous effects model is overidentified in shift-share designs. One possible way to proceed, for instance after statistically rejecting the model, is to weaken assumptions in a manner that renders the model potentially just identified (e.g., in the sense of (Chen and Santos, 2018)). In this section, we show that two natural relaxations to the model of Section 2 are severely limited in the set of empirical contexts to which they may be successfully applied. As a result, we conclude that the homogeneous effects model of Section 2 is of central empirical importance in shift-share designs relying on Bartik instruments.

First, we examine the possibility of relaxing the model in Section 2 to a linear heterogeneous effects model. In this context, the TSLS estimand can be interpreted as a weighted average of causal effects for different population subgroups. We find, however, that ensuring the estimand equals a *positively* weighted average of causal effects requires strong, often unrealistic, assumptions on the distribution of the data.

Second, we examine the possibility of relaxing the model in Section 2 by only requiring that the Bartik instrument be exogenous when expectations are evaluated over *both* the shocks and the shares. We show that, under this exogeneity requirement, identification crucially relies on the time series variation of the shocks and, as a result, that the standard errors of TSLS depend on the time dimension. Because the majority of empirical studies relying on shift-share designs have employed short panels, we expect TSLS estimates to be highly imprecise under this identification strategy.

3.1 Heterogeneous Effects

We first re-examine the shift-share design of Section 2 in the presence of heterogeneous treatment effects. To this end, we generalize the second stage equation by setting

$$Y_i = X_i \beta_i + W_i' \gamma_{\mathbf{s}} + \varepsilon_i; \tag{11}$$

⁵More precisely, Adao et al. (2019) impose that $W_i = \mathcal{W}' S_i + U_i$ and require U_i to be asymptotically negligible in a suitable sense; see their Assumptions 3(iii)(iv).

i.e. the “treatment effect” β_i is allowed to depend on the individual but, for simplicity, we keep the coefficient for the controls fixed. We complement the second stage by following [Adao et al. \(2019\)](#) and [Goldsmith-Pinkham et al. \(2020\)](#) in imposing a linear first stage

$$X_i = \mathcal{Z}'\Lambda_i S_i + W_i'\gamma_f + \eta_i \quad \Lambda_i = \text{diag}\{(\lambda_{i1}, \dots, \lambda_{ip})\}, \quad (12)$$

where “ $\text{diag}\{(a_1, \dots, a_p)\}$ ” denotes a diagonal matrix with diagonal entries (a_1, \dots, a_p) and again, for simplicity, we set the coefficient for the controls W_i to be fixed. In analogy to [Imbens and Angrist \(1994\)](#), we refer to Λ_i as the “type” of the individual.

Within this context, we examine conditions under which the TSLS estimand retains a causal interpretation. Formally, we remain interested in the parameter $\beta_{0,n}$ solving

$$\frac{1}{n} \sum_{i=1}^n E \left[(Y_i - X_i \beta_{0,n}) \dot{Z}_{i,n} \mid \mathcal{G}_n \right] = 0,$$

but now study conditions under which $\beta_{0,n}$ equals a positively weighted average of the average treatment effects for different subgroups. Under suitable exogeneity and rank conditions on the instrument Z_i , it is possible to show that the estimand $\beta_{0,n}$ satisfies

$$\beta_{0,n} = E \left[\left(\sum_{i=1}^n \omega_{i,n} E[\beta_i \mid \Lambda_i, W_i, \mathcal{G}_n] \right) \mid \mathcal{G}_n \right] \quad \omega_{i,n} \equiv \frac{E[X_i \dot{Z}_{i,n} \mid \Lambda_i, W_i, \mathcal{G}_n]}{\sum_{i=1}^n E[X_i \dot{Z}_{i,n} \mid \mathcal{G}_n]}; \quad (13)$$

i.e. $\beta_{0,n}$ equals the expectation of a $\{\omega_{i,n}\}_{i=1}^n$ -weighted average of the average treatment effects for subgroups determined by Λ_i and W_i ; see [Lemma A.2.1](#) for a formal statement. As forcibly argued in the literature, a minimal requirement for $\beta_{0,n}$ to possess a causal interpretation is that the weights $\{\omega_{i,n}\}_{i=1}^n$ be positive ([Kolesar, 2013](#); [Blandhol et al., 2022](#)). In what follows, we study the implications of requiring that such a positivity condition hold under different asymptotic frameworks – i.e. under different choices of \mathcal{G}_n . For succinctness, we will refer to TSLS as having a causal interpretation whenever the weights $\{\omega_{i,n}\}_{i=1}^n$ are positive with probability one.

3.1.1 Conditioning on Shocks

We begin by studying the conditions under which TSLS has a causal interpretation in an asymptotic setting in which only the cross section grows (i.e. $n \rightarrow \infty$). Recall that in [Section 2.1](#) we showed that this setting implicitly conditions on the shocks \mathcal{Z} and requires the shares S_i to be a valid instrument.

The analysis in this section relies on the following assumption and additional regularity conditions that we formally state in the Appendix; see [Assumption A.2.1](#).

Assumption 3.1. (i) \mathcal{Z} and $\{Y_i, X_i, W_i, S_i\}_{i=1}^n$ satisfy equations (11) and (12); (ii) $(\mathcal{Z}, \Lambda_i, \beta_i, \varepsilon_i, \eta_i) \perp\!\!\!\perp S_i$ conditionally on W_i ; (iii) $E[Z_i|W_i, \mathcal{G}_n] = W_i' \pi_n$ with $\mathcal{G}_n = \{\mathcal{Z}\}$.

Assumption 3.1(i) formally imposes the structure of the model, while Assumption 3.1(ii) requires the shares S_i to be suitably exogenous. In Assumption 3.1(iii) we further demand that the conditional mean of the instrument given the covariates be linear.⁶ Blandhol et al. (2022) showed in a related context that Assumption 3.1(iii) is a necessary condition for TSLS to have a causal interpretation. We therefore impose Assumption 3.1(iii) not because it is innocuous, but because it enables us to examine what additional conditions are necessary for TSLS to have a causal interpretation. Assumption 3.1(iii) is of course testable, and practitioners may wish to examine its validity in assessing whether the TSLS estimator can be assigned a causal interpretation.

Our next proposition characterizes the weights $\{\omega_{i,n}\}_{i=1}^n$ and obtains necessary and sufficient conditions for them to be positive with probability one.

Proposition 3.1. *If Assumptions 3.1 holds and $\mathcal{G}_n = \mathcal{Z}$, then it follows that*

$$\omega_{i,n} = \frac{\mathcal{Z}'(\Lambda_i \text{Var}\{S_i|W_i\})\mathcal{Z}}{\sum_{j=1}^n \mathcal{Z}'(\Lambda_j \text{Var}\{S_j|W_j\})\mathcal{Z}}. \quad (14)$$

Moreover, if in addition Assumption A.2.1 holds, then for n large $\omega_{i,n}$ is positive for all i with probability one (over $\Lambda_i, W_i, \mathcal{Z}$) if and only if the matrix $\Lambda_i \text{Var}\{S_i|W_i\}$ is either positive semi-definite with probability one, or negative semi-definite with probability one.

The first part of Proposition 3.1 shows that the weights may be expressed as a quadratic form in the shocks \mathcal{Z} (see (14)). Intuitively, provided that the support of the shocks \mathcal{Z} is sufficiently rich, it therefore follows that the weights can only be positive for all possible realizations of the shocks \mathcal{Z} if in fact the matrices $\Lambda_i \text{Var}\{S_i|W_i\}$ are positive semi-definite (or negative semi-definite) with probability one. The second part of Proposition 3.1 formalizes this intuition under the requirement that the support of the shocks \mathcal{Z} contain a neighborhood of zero.⁷

The necessary and sufficient conditions derived in Proposition 3.1 for TSLS to have a causal interpretation are both highly restrictive and testable. Our next corollary shows that, under mild conditions on the support of Λ_i , these conditions in fact necessarily fail whenever shares are correlated with each other.

Corollary 3.1. *Let the conditions of Proposition 3.1 hold and suppose that the support of $\lambda_{ij}/\lambda_{ik}$ is unbounded for any $j \neq k$. If for some $j \neq k$ we have that*

$$P(\text{Cov}\{S_{ij}, S_{ik}|W_i\} \neq 0) > 0, \quad (15)$$

⁶Under regularity conditions, Assumption 3.1(iii) is equivalent to $E[S_i|W_i]$ being linear in W_i .

⁷This requirement is formally stated in Assumption A.2.1 in the Appendix.

then the weights $\{\omega_{i,n}\}_{i=1}^n$ are negative with positive probability (over $(\Lambda_i, W_i, \mathcal{Z})$).

Corollary 3.1 implies that, whenever shares are correlated, TSLS necessarily lacks a causal interpretation under certain realizations of the shocks \mathcal{Z} . Because shares must be correlated whenever they sum up to one, we view Corollary 3.1 as a warning that TSLS based on the Bartik instrument can easily fail to have a causal interpretation under a heterogenous effects model. However, it may be worth emphasizing that, as argued by Goldsmith-Pinkham et al. (2020), empirical researchers may still be able to estimate causal parameters by carefully employing the shares as separate instruments instead of combining them into a single scalar Z_i by employing the shocks \mathcal{Z} .

Remark 3.1. Corollary 3.1 yields testable implications beyond restricting the correlation between shares. Notably, because $\text{Var}\{S_i|W_i\}$ must be diagonal and Λ_i must have constant sign in order for TSLS to have a causal interpretation, it follows that the sign of the first stage must be constant under certain choices of instrument. In particular, letting $\tilde{Z}_{ij} = f_j(W_i)\mathcal{Z}_j$ for any positive f_j and $\tilde{\mathbf{Z}}_i = (\tilde{Z}_{i1}, \dots, \tilde{Z}_{ip})'$, it follows that

$$\text{sign} \left\{ E[X_i \dot{Z}_{i,n} | \mathcal{Z}] \right\} = \text{sign} \left\{ E[X_i (S_i - E[S_i|W_i])' \tilde{\mathbf{Z}}_i | \mathcal{Z}] \right\}. \quad (16)$$

Since the conditional mean of the shares is linear in the controls in our setting, we obtain a simple diagnostic: Compare the sign of the covariances between X_i and $\dot{S}'_{i,n} \tilde{\mathbf{Z}}_i$ for different choices of $\tilde{\mathbf{Z}}_i$ – here, $\dot{S}_{i,n}$ denotes the residual from regressing S_i on W_i . ■

3.1.2 Identification Through Shocks

We next examine the conditions under which TSLS retains a causal interpretation in an asymptotic framework in which identification is driven by the exogeneity of the aggregate shocks. To this end, we follow Adao et al. (2019) and Borusyak et al. (2022) in augmenting the conditioning set \mathcal{G}_n of Section 2.2 to include all unobserved heterogeneity – i.e., we set \mathcal{G}_n to equal $\{S_i, W_i, \Lambda_i, \beta_i, \varepsilon_i, \eta_i\}_{i=1}^n$.

Our analysis in this section relies on the following assumption and a set of regularity conditions that we formally state in the Appendix; see Assumption A.3.1.

Assumption 3.2. (i) \mathcal{Z} and $\{Y_i, X_i, W_i, S_i\}_{i=1}^n$ satisfy equations (11) and (12); (ii) $E[Z_i|W_i, \mathcal{G}_n] = W_i' \pi_n$ with $\mathcal{G}_n = \{S_i, W_i, \Lambda_i, \beta_i, \varepsilon_i, \eta_i\}_{i=1}^n$.

Assumption 3.2 imposes the structure of our model and requires that the conditional mean of the instrument given the controls be linear.⁸ As previously noted in Section 3.1.1, linearity has been shown to be a necessary condition for TSLS to preserve a causal

⁸Sufficient conditions for Assumption 3.2(ii) were introduced by Adao et al. (2019); see Remark 2.3.

interpretation in related contexts. We reiterate that we therefore impose linearity to examine what additional conditions are needed for TSLS to have a causal interpretation, and not because we view linearity as an innocuous assumption. It is also worth emphasizing the connection between Assumption 3.2(ii) and Proposition 2.2, which established that linearity was a necessary condition for the instrument to be exogenous in a homogeneous effects model. As a result, the overidentification test for the homogeneous effects model that we develop in Section 4 may be readily adapted to test Assumption 3.2(ii).

Our next proposition obtains a characterization of the weights $\{\omega_{i,n}\}_{i=1}^n$ and derives a necessary and sufficient condition for TSLS to have a causal interpretation.

Proposition 3.2. *If Assumption 3.2 holds, and $\mathcal{G}_n = \{S_i, W_i, \Lambda_i, \beta_i, \varepsilon_i, \eta_i\}_{i=1}^n$, then*

$$\omega_{i,n} = \frac{S_i'(\Lambda_i \text{Var}\{\mathcal{Z}|\mathcal{G}_n\})S_i}{\sum_{j=1}^n S_j'(\Lambda_j \text{Var}\{\mathcal{Z}|\mathcal{G}_n\})S_j}. \quad (17)$$

Moreover, if in addition Assumption A.3.1 holds, then for n large $\omega_{i,n}$ is positive for all i with probability one (over \mathcal{G}_n) if and only if the distribution of $\Lambda_i \text{Var}\{\mathcal{Z}|\mathcal{G}_n\}$ satisfies

$$P\left(\sup_{s \in \mathbb{R}_+^p} s'(\Lambda_i \text{Var}\{\mathcal{Z}|\mathcal{G}_n\})s \leq 0\right) = 1 \text{ or } P\left(\inf_{s \in \mathbb{R}_+^p} s'(\Lambda_i \text{Var}\{\mathcal{Z}|\mathcal{G}_n\})s \geq 0\right) = 1. \quad (18)$$

The first part of Proposition 3.2 establishes that the weights $\{\omega_{i,n}\}_{i=1}^n$ equal a quadratic form in the shares S_i . Intuitively, the expression for the weights as a quadratic form in the shares implies that the support of the random matrix $\Lambda_i \text{Var}\{\mathcal{Z}|\mathcal{G}_n\}$ must be restricted in order for the weights to be positive for any possible realization of the shares. The second part of Proposition 3.2 formalizes this intuition under a requirement that the support of the shares is suitably rich; see Assumption A.3.1 for a formal statement.

Because, unlike the shocks, the shares are always positive, the necessary and sufficient condition for TSLS to have causal interpretation (i.e. (18)) is weaker than requiring the random matrix $\Lambda_i \text{Var}\{\mathcal{Z}|\mathcal{G}_n\}$ to be positive (or negative) semidefinite with probability one. The conditions derived by Proposition 3.2 are nonetheless restrictive and testable. Our next corollary, for instance, employs Proposition 3.2 to establish that TSLS fails to have a causal interpretation whenever shocks are negatively correlated.

Corollary 3.2. *Let the conditions of Proposition 3.2 hold and suppose $P(\lambda_{ij} = 0) = 0$ for all j and the support of $\lambda_{ij}/\lambda_{ik}$ is unbounded for any $j \neq k$. If in addition*

$$P(\text{Cov}\{\mathcal{Z}_j, \mathcal{Z}_k|\mathcal{G}_n\} < 0) > 0 \quad (19)$$

for some $j \neq k$, then the weights $\{\omega_{i,n}\}_{i=1}^n$ are negative with positive probability.

Traditionally, the literature has assumed shocks to either be uncorrelated or clustered for asymptotic purposes (Adao et al., 2019; Borusyak et al., 2022). Corollary 3.2

highlights that the covariance structure of the shocks is important not only for delivering asymptotic approximations and computing standard errors, but also for assigning TSLS a causal interpretation. Corollary 3.2 additionally has important implications for applications in which standard errors are clustered. In particular, since clustered standard errors reflect a concern that shocks within a cluster are correlated, Corollary 3.2 implies that TSLS can only retain a causal interpretation if in fact all correlations within a cluster are positive conditionally on \mathcal{G}_n . Empirical researchers clustering standard errors should therefore argue that all correlations within a cluster are positive or, alternatively, rely on a homogeneous effects model.

Remark 3.2. Because all entries of $\text{Var}\{\mathcal{Z}|\mathcal{G}_n\}$ must be (weakly) positive and the sign of Λ_i must be constant for TSLS to possess a causal interpretation, it follows that

$$\text{sign} \left\{ \sum_{i=1}^n E[X_i \dot{Z}_{i,n} | \mathcal{G}_n] \right\} = \text{sign} \left\{ \sum_{i=1}^n E[X_i (\mathcal{Z} - E[\mathcal{Z} | \mathcal{G}_n])' f(S_i, W_i) | \mathcal{G}_n] \right\} \quad (20)$$

for any positive $f(S_i, W_i) \in \mathbb{R}^p$ and \mathcal{G}_n as in Corollary 3.2. In parallel to Remark 3.1, we may interpret the restriction in (20) as demanding that the sign of the first stage be constant under certain alternative choices of instrument. Empirically evaluating this restriction can require an estimator for the conditional mean of the shocks. For this purpose, we note that multiple such estimators have been proposed in the literature as they are also needed to compute the standard errors of the TSLS estimator when conditioning on \mathcal{G}_n (Adao et al., 2019; Borusyak et al., 2022). ■

3.2 Long Panel

Our cross sectional analysis has so far conditioned on either the aggregate shocks or the shares to obtain identification. Both approaches yield overidentifying restrictions for homogeneous effects models that carry over to short panel settings (i.e. T fixed); see Remarks 2.1 and 2.2. In this section, we conclude by exploring the implications of employing *unconditional* moment restrictions for identification instead.

For illustrative purposes, we consider a homogeneous effects models with no controls and a scalar endogenous variable. We assume that at each time period $1 \leq t \leq T$ we observe an aggregate shock Z_t and variables $\{Y_{it}, X_{it}, Z_{it}, S_{it}\}_{i=1}^n$ satisfying

$$Y_{it} = X_{it}\beta + \varepsilon_{it} \quad Z_{it} = S_{it}' Z_t.$$

As an identifying assumption, we now employ the just identified moment restriction

$$\frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T E[\varepsilon_{it} Z_{it}] = 0. \quad (21)$$

Crucially, the expectation in (21) is taken over both the cross section and the time series – e.g., over both \mathcal{Z}_t and $(S_{it}, \varepsilon_{it})$. In particular, the moment restriction in (21) contrasts with Goldsmith-Pinkham et al. (2020) and Adao et al. (2019) who instead consider conditional expectations given \mathcal{Z}_t and $(S_{it}, \varepsilon_{it})$ respectively; see Section 2.

The natural estimator for β continues to be the same TSLS estimator that we have considered so far. However, relying on the unconditional moment restriction in (21) for identification now requires us to let both n and T grow so that we may approximate expectations over both the cross section and the time series. As a result, when employing the just identified moment restriction in (21) to obtain identification, we need to rely on “long panel” asymptotic approximations and their corresponding standard errors. The main message of this section is that the long panel standard errors for TSLS decrease at a rate of $1/\sqrt{T}$. Therefore, long panel standard errors, and hence the identification restriction in (21), are unlikely to prove informative in applications in which T is small.

In the rest of the section, we provide a summary of the long panel asymptotic properties of the TSLS estimator. Due to the technical nature of the analysis, we relegate formal statements to the appendix and focus instead on providing intuition for the results. To this end, we begin by noting that the TSLS estimator $\hat{\beta}_n$ satisfies

$$\hat{\beta}_n - \beta = \left(\frac{1}{nT} \sum_{t=1}^T \sum_{i=1}^n X_{it} Z_{it} \right)^{-1} \left(\frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T \mathcal{Z}'_t S_{it} \varepsilon_{it} \right). \quad (22)$$

As the cross section and time series grow, the denominator in (22) converge in probability to some constant $D \neq 0$ under standard conditions; i.e., the denominator satisfies

$$\frac{1}{nT} \sum_{t=1}^T \sum_{i=1}^n X_{it} Z_{it} \xrightarrow{p} D. \quad (23)$$

The asymptotic distribution of the TSLS estimator is therefore governed by the numerator in (22). In order to derive this asymptotic distribution, we let \mathcal{C}_t denote all aggregate shocks at time t (which includes \mathcal{Z}_t) and assume that $\{S_{it}\varepsilon_{it}\}_{i=1}^n$ are i.i.d. across i (but not t) conditionally on the aggregate shocks \mathcal{C}_t . Defining the variables

$$\zeta_t \equiv E[S_{it}\varepsilon_{it}|\mathcal{C}_t], \quad \nu_{it} \equiv S_{it}\varepsilon_{it} - \zeta_t,$$

we then obtain a decomposition into a time series and a panel process by noting that

$$\frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T \mathcal{Z}'_t S_{it} \varepsilon_{it} = \frac{1}{T} \sum_{t=1}^T \mathcal{Z}'_t \zeta_t + \frac{1}{nT} \sum_{t=1}^T \sum_{i=1}^n \mathcal{Z}'_t \nu_{it}. \quad (24)$$

Crucially, the time series process $\{\mathcal{Z}'_t \zeta_t\}_{t=1}^T$ has mean zero because of the moment restriction in (21), while the panel process $\{\mathcal{Z}'_t \nu_{it}\}_{i,t=1}^{n,T}$ has mean zero by construction.

Thus, under appropriate restrictions, we should expect both the time series and panel processes to be asymptotically normally distributed.

The preceding discussion is formalized in the appendix, where we establish:

Proposition 3.3. *Under the technical conditions presented in Appendix A.4 we have*

$$\left(\frac{1}{\sqrt{T}} \sum_{t=1}^T \mathbf{Z}'_t \zeta_t, \frac{1}{\sqrt{nT}} \sum_{t=1}^T \sum_{i=1}^n \mathbf{Z}'_t \nu_{i,t} \right) \xrightarrow{d} (\mathbb{G}_\zeta, \mathbb{G}_\nu) \quad (25)$$

as $n, T \rightarrow \infty$, with \mathbb{G}_ζ and \mathbb{G}_ν independent Gaussian random variables.

The long panel asymptotic properties of the TSLS estimator immediately follow from Proposition 3.3. For instance, combining Proposition 3.3 with results (22), (23), and (24) allows us to approximate the variance of $\hat{\beta}_n$ by the expression

$$\text{Var} \left\{ \hat{\beta}_n - \beta \right\} \approx \left(\frac{1}{D} \right)^2 \left\{ \frac{1}{T} \text{Var} \{ \mathbb{G}_\zeta \} + \frac{1}{nT} \text{Var} \{ \mathbb{G}_\nu \} \right\}. \quad (26)$$

In particular, it follows that the long panel standard errors for TSLS decrease at a rate of $1/\sqrt{T}$. Intuitively, this dependence on the time dimension results from the dependence of the moment restriction on the distribution of the time series. One important exception to this phenomenon arises when the time series process has zero variance (i.e. $\text{Var} \{ \mathbb{G}_\zeta \} = 0$). This exception corresponds to the case in which $\zeta_t = 0$ or, equivalently,

$$E[S_{it} \varepsilon_{it} | C_t] = 0. \quad (27)$$

However, as we discussed in Section 2.1, the conditional moment restriction in (27) is exactly the identifying assumption made by Goldsmith-Pinkham et al. (2020) that renders the model overidentified. In summary, the long panel analysis suggests that practitioners should either report standard errors based on (26) or rely on the short panel identification strategies discussed in Section 2 and report the corresponding overidentification tests instead.

4 Overidentification Tests

We next build on our results by developing overidentification tests for the homogeneous effects model studied in Section 2. While we focus on homogeneous effects models due to their demonstrated importance in shift-share designs, we note that it is also possible to test the overidentifying restrictions for heterogeneous effects models derived in Remarks 3.1 and 3.2 by adapting the insights in, e.g., Bai et al. (2022).

4.1 The Approach

The identification strategies commonly employed for homogeneous effects models yield testable implications in the form of moment equality restrictions. In this section, we present a general inference approach that is portable across the moment restrictions yielded by different identification strategies.

In what follows, we let $V_i \equiv (Y_i, X_i, W_i, S_i, \mathcal{Z})$ for notational simplicity. The over-identifying moment restrictions we derived in Section 2 have the structure

$$E[f(V_i, \theta) | \mathcal{G}_n] = 0,$$

where θ represent some unknown parameter (e.g., β) and $f(V_i, \theta) \in \mathbb{R}^q$ is a vector valued function. In many applications, the number of moment equality restrictions is high dimensional in the sense that it is “large” relative to the sample size. Because the distributional approximations implicit in chi-squared overidentification tests can be unreliable in high dimensional problems, we instead employ test statistics based on the high dimensional central limit theorem of Chernozhukov et al. (2022). Specifically, for an estimator $\hat{\theta}_n$ of θ and $f_j(V_i, \hat{\theta}_n)$ denoting the j^{th} coordinate of $f(V_i, \hat{\theta}_n) \in \mathbb{R}^q$ we set

$$T_n \equiv \max_{1 \leq j \leq q} \left| \sum_{i=1}^n f_j(V_i, \hat{\theta}_n) \right| \quad (28)$$

as the test statistic for the null hypothesis that the moment restrictions hold.

The main assumption we impose to construct a test is that the sample moments be asymptotically linear; see Appendix A.5 for a formal statement of the assumptions and results for this section. Intuitively, we require that for some collection of independent mean zero random variables $\{\psi_i\}_{i=1}^{b_n} \subset \mathbb{R}^q$ the test statistic approximately equals

$$T_n \approx \max_{1 \leq j \leq q} \left| \sum_{i=1}^{b_n} \psi_{ij} \right|,$$

where ψ_{ij} denotes the j^{th} coordinate of ψ_i . We refer to the number b_n of random variables ψ_i as the “effective number of observations,” and note that it need not equal the sample size. Such a distinction is useful, for example, when the data is clustered (in which case b_n equals the number of clusters) or when relying in the asymptotic framework in Adao et al. (2019) (in which case b_n equals the number of shocks p).

We obtain critical values for our test by relying on a bootstrap approximation. Specifically, given estimates $\{\hat{\psi}_i\}_{i=1}^{b_n}$ for $\{\psi_i\}_{i=1}^{b_n}$, we define a bootstrap statistic

$$T_n^* \equiv \max_{1 \leq j \leq q} \left| \sum_{i=1}^{b_n} \omega_i \left(\hat{\psi}_{ij} - \frac{1}{b_n} \sum_{k=1}^{b_n} \hat{\psi}_{kj} \right) \right|,$$

where $\hat{\psi}_{ij}$ denotes the j^{th} coordinate of $\hat{\psi}_i \in \mathbb{R}^q$ and the random weights $\{\omega_i\}_{i=1}^{b_n}$ are drawn independently of the data $\{V_i\}_{i=1}^n$. For instance, we may set the bootstrap weights $\{\omega_i\}_{i=1}^{b_n}$ to be i.i.d. and drawn from a standard normal or Rademacher distribution.⁹ The bootstrap critical value for a level α test is then given by the bootstrap quantile

$$\hat{c}_n \equiv \inf\{c : P(T_n^* \leq c | \{V_i\}_{i=1}^n) \geq 1 - \alpha\};$$

i.e. we employ the $1 - \alpha$ quantile of T_n^* conditional on the data $\{V_i\}_{i=1}^n$ but unconditionally on the bootstrap weights $\{\omega_i\}_{i=1}^{b_n}$. The critical value \hat{c}_n can as usual be obtained through simulation by employing multiple draws of the bootstrap weights $\{\omega_i\}_{i=1}^{b_n}$ to approximate the distribution of T_n^* conditionally on the data.

Our next result establishes that a test that rejects whenever T_n exceeds the critical value \hat{c}_n has the correct asymptotic size. We defer the statement of the relevant regularity conditions to Appendix A.5, though note that they allow for the number of moment restrictions q to grow together with the effective sample size b_n .

Proposition 4.1. *Under the regularity conditions discussed in Appendix A.5, we have*

$$\lim_{n \rightarrow \infty} P(T_n > \hat{c}_n) = \alpha.$$

4.2 Conditioning on Shocks

The general approach discussed in the preceding section can readily be specialized to develop an overidentification test for application that implicitly condition on aggregate shocks in their analysis. Recall that such applications require the restriction

$$E[S_i \varepsilon_i | \mathcal{G}_n] = 0,$$

where \mathcal{G}_n is understood to contain all aggregate variables, including \mathcal{Z} (Goldsmith-Pinkham et al., 2020). As our test statistic we therefore employ

$$T_n = \max_{1 \leq j \leq p} \left| \frac{1}{\hat{\sigma}_j} \sum_{i=1}^n S_{ij} \hat{\varepsilon}_i \right|, \quad (29)$$

where S_{ij} denotes the j^{th} coordinate of $S_i \in \mathbb{R}^p$, $\{\hat{\varepsilon}_i\}_{i=1}^n$ are the residuals from the estimated model, and $\hat{\sigma}_j$ is an estimated weight that allows us to ensure that all p moments are on a comparable scale.

In order to map this setting into the overidentification test of the preceding section, we simply set $f_j(V_i, \hat{\theta}_n) = S_{ij} \hat{\varepsilon}_i / \hat{\sigma}_j$ so that the test statistic in (29) becomes a special

⁹The Rademacher distribution corresponds to setting ω_i to satisfy $P(\omega_i = 1) = P(\omega_i = -1) = 1/2$.

case of (28). Setting $A_i \equiv (Z_i, W_i)'$, σ_j to be the probability limit of $\hat{\sigma}_j$, and defining

$$\psi_{ij} = \frac{U_{ij}}{\sigma_j} \quad U_{ij} \equiv \left(S_{ij}\varepsilon_i - E[S_{ij}(X_i, W_i') | \mathcal{G}_n] \left(E[A_i(X_i, W_i') | \mathcal{G}_n] \right)^{-1} A_i \varepsilon_i \right),$$

it is then possible to show that, under the null hypothesis, T_n approximately equals

$$T_n \approx \max_{1 \leq j \leq p} \left| \sum_{i=1}^n \psi_{ij} \right|$$

provided the observations $\{Y_i, S_i, X_i\}_{i=1}^n$ are i.i.d. conditionally on \mathcal{G}_n .¹⁰ Our inference approach requires an estimator for the influence function ψ_i and to this end we set

$$\hat{\psi}_{ij} \equiv \frac{\hat{U}_{ij}}{\hat{\sigma}_j} \quad \hat{U}_{ij} \equiv \left(S_{ij}\hat{\varepsilon}_i - \left(\frac{1}{n} \sum_{k=1}^n S_{kj}(X_k, W_k') \right) \left(\frac{1}{n} \sum_{k=1}^n A_k(X_k, W_k') \right)^{-1} A_i \hat{\varepsilon}_i \right).$$

We also let $\hat{\sigma}_j^2$ be an estimator of the asymptotic variance of the j^{th} moment by setting

$$\hat{\sigma}_j^2 = \frac{1}{n} \sum_{i=1}^n \left(\hat{U}_{ij} - \frac{1}{n} \sum_{k=1}^n \hat{U}_{kj} \right)^2.$$

Following the approach in the preceding section, we obtain critical values for our test statistic by: (i) Drawing $b \in \{1, \dots, B\}$ samples $\{\omega_i^{(b)}\}_{i=1}^n$ of standard normal random variables independent of the data; (ii) For each drawn sample $\{\omega_i^{(b)}\}_{i=1}^n$ computing

$$T_n^{*(b)} \equiv \max_{1 \leq j \leq p} \left| \sum_{i=1}^n \omega_i^{(b)} \left(\hat{\psi}_{ij} - \frac{1}{n} \sum_{k=1}^n \hat{\psi}_{kj} \right) \right|;$$

and (iii) For a level α test, letting $\hat{c}_{1-\alpha}$ denote the $1 - \alpha$ quantile of $\{T_n^{*(b)}\}_{b=1}^B$. Our overidentification test then rejects whenever T_n exceeds the critical value $\hat{c}_{1-\alpha}$.

Remark 4.1. The proposed test can easily be adapted to account for clustered data. Specifically, suppose observations are divided into clusters $\{c_1, \dots, c_{|C|}\} \equiv C$ – e.g., C may represent all states, $|C|$ the total number of states, and each cluster $c \in C$ a specific state. While the test statistic remains the same, for our bootstrap implementation we now set $\hat{\psi}_{cj} \equiv \sum_{i \in c} \hat{U}_{ij} / \hat{\sigma}_j$ for $\hat{\sigma}_j^2$ the cluster robust sample variance of $\{\hat{U}_{ij}\}_{i=1}^n$, and let

$$T_n^* \equiv \max_{1 \leq j \leq p} \left| \sum_{c \in C} \omega_c \left(\hat{\psi}_{cj} - \frac{1}{|C|} \sum_{\tilde{c} \in C} \hat{\psi}_{\tilde{c}j} \right) \right|,$$

with $\{\omega_i\}_{i \in C}$ an i.i.d. sample of standard normal random variables drawn independently of the data – i.e., the bootstrap procedure is modified by simply employing the same

¹⁰See the supplemental appendix for calculations supporting the claims in this section.

draw of ω_c for all observations in the same cluster. We note that this procedure maps into the notation of Section 4.1 by letting b_n equal the number of clusters. ■

4.3 Identification Through Shocks

We conclude our discussion of overidentification tests by specializing our approach to applications in which the exogeneity of the Bartik instrument originates from the exogeneity of the shocks. Recall from Section 2.2 that the asymptotic framework designed for such applications relies on setting $\mathcal{G}_n \equiv \{S_i, W_i, \varepsilon_i\}_{i=1}^n$, which yields the restriction

$$E[Z_i | \{S_i, W_i, \varepsilon_i\}_{i=1}^n] = W_i' \pi_n$$

for π_n the population regression coefficient from regressing $\{Z_i\}_{i=1}^n$ on $\{W_i\}_{i=1}^n$ conditionally on \mathcal{G}_n (as in (3)). As our overidentification test statistic we therefore employ

$$T_n = \max_{1 \leq j \leq q} \left| \frac{1}{\hat{\sigma}_j} \sum_{i=1}^n g_j(\hat{\varepsilon}_i, W_i, S_i) (Z_i - W_i' \hat{\pi}_n) \right|, \quad (30)$$

where $\{\hat{\varepsilon}_i\}_{i=1}^n$ are the residuals from the estimated model, $\hat{\pi}_n$ is the regression coefficient from regressing $\{Z_i\}_{i=1}^n$ on $\{W_i\}_{i=1}^n$, $\hat{\sigma}_j$ is again an estimated weight, and $g_j(\hat{\varepsilon}_i, W_i, S_i)$ denotes the j^{th} coordinate of the vector valued function $g(\hat{\varepsilon}_i, W_i, S_i) \in \mathbb{R}^q$.

In order to obtain a suitable critical value, we first need a characterization of the influence function for each sample moment. To this end, we define

$$\begin{aligned} \delta_j &\equiv \left(\sum_{i=1}^n W_i W_i' \right)^{-1} \sum_{i=1}^n W_i g_j(\varepsilon_i, W_i, S_i) \\ \kappa_j &\equiv \left(\sum_{i=1}^n E[S_i' \mathcal{E} X_i | \mathcal{G}_n] \right)^{-1} \left(\sum_{i=1}^n E[S_i' \mathcal{E} X_i | \mathcal{G}_n] \frac{\partial}{\partial \varepsilon} g_j(\varepsilon_i, W_i, S_i) \right) \end{aligned}$$

with $\mathcal{E} \equiv Z - E[Z | \mathcal{G}_n]$, and for σ_j the probability limit of $\hat{\sigma}_j$ we let ψ_{ij} be given by

$$\psi_{ij} \equiv \frac{U_{ij}}{\sigma_j} \quad U_{ij} \equiv \mathcal{E}_i \times \sum_{k=1}^n S_{ki} (g_j(\varepsilon_k, W_k, S_k) - W_k' \delta_j - \varepsilon_k \kappa_j).$$

Under conditions similar to those employed by [Adao et al. \(2019\)](#), it is then possible to show that under the null hypothesis the statistic T_n satisfies the approximation¹¹

$$T_n \approx \max_{1 \leq j \leq q} \left| \sum_{i=1}^p \psi_{ij} \right|.$$

We note that, because identification is driven by the exogeneity of the shocks, the

¹¹See the supplemental appendix for calculations supporting the claims in this section.

effective number of observations equals p and not n – i.e. the sample $\{\psi_i\}$ is of size p .

Our bootstrap critical value also relies on an estimate of ψ_i , and to this end we define

$$\begin{aligned}\hat{\delta}_j &\equiv \left(\sum_{i=1}^n W_i W_i' \right)^{-1} \sum_{i=1}^n W_i g_j(\hat{\varepsilon}_i, W_i, S_i) \\ \hat{\kappa}_j &\equiv \left(\sum_{i=1}^n (Z_i - W_i' \hat{\pi}_n) X_i \right)^{-1} \left(\sum_{i=1}^n (Z_i - W_i' \hat{\pi}_n) X_i \frac{\partial}{\partial \varepsilon} g_j(\hat{\varepsilon}_i, W_i, S_i) \right),\end{aligned}$$

and let $\hat{\mathcal{E}} \in \mathbb{R}^p$ denote a suitable estimator for $\mathcal{E} \equiv \mathcal{Z} - E[\mathcal{Z}|\mathcal{G}_n]$; see Remark 4.2 for possible choices of $\hat{\mathcal{E}}$. As our estimator for the influence function ψ_{ij} we then set

$$\hat{\psi}_{ij} \equiv \frac{\hat{U}_{ij}}{\hat{\sigma}_j} \quad \hat{U}_{ij} \equiv \hat{\mathcal{E}}_i \times \sum_{k=1}^n S_{ki} (g_j(\hat{\varepsilon}_k, W_i, S_i) - W_k' \hat{\delta}_j - \hat{\varepsilon}_k \hat{\kappa}_j).$$

Finally, for the purposes of studentizing the sample moments we set $\hat{\sigma}_j^2$ to be given by

$$\hat{\sigma}_j^2 \equiv \frac{1}{p} \sum_{i=1}^p \left(\hat{U}_{ij} - \frac{1}{p} \sum_{k=1}^p \hat{U}_{kj} \right)^2.$$

Given the estimates $\{\hat{\psi}_i\}$, we can obtain asymptotically valid critical values by following the same procedure as in the preceding section. Specifically, we: (i) Draw $b \in \{1, \dots, B\}$ samples $\{\omega_i^{(b)}\}_{i=1}^p$ of standard normal random variables independent of the data; (ii) For each drawn sample $\{\omega_i^{(b)}\}_{i=1}^p$ we compute the bootstrap statistic

$$T_n^{*(b)} \equiv \max_{1 \leq j \leq q} \left| \sum_{i=1}^p \omega_i^{(b)} \left(\hat{\psi}_{ij} - \frac{1}{p} \sum_{k=1}^p \hat{\psi}_{kj} \right) \right|;$$

and (iii) For a level α test, we let $\hat{c}_{1-\alpha}$ denote the $1 - \alpha$ quantile of $\{T_n^{*(b)}\}_{b=1}^B$. Our overidentification test then rejects whenever T_n exceeds the critical value $\hat{c}_{1-\alpha}$.

Remark 4.2. Multiple estimates $\hat{\mathcal{E}}$ for $\mathcal{E} = \mathcal{Z} - E[\mathcal{Z}|\mathcal{G}_n]$ have been previously studied in the literature. [Adao et al. \(2019\)](#), for instance, propose the estimator

$$\hat{\mathcal{E}} \equiv \left(\sum_{i=1}^n S_i S_i' \right)^{-1} \sum_{i=1}^n S_{it} (Z_{it} - W_{it}' \hat{\pi}_n). \quad (31)$$

Alternatively, in applications in which W_i contains variables equal to $\mathcal{Q}' S_i$ for some $p \times k$ observable matrix of aggregate variables \mathcal{Q} , [Borusyak et al. \(2022\)](#) advocate setting

$$\hat{\mathcal{E}} \equiv \mathcal{Z} - \mathcal{Q}(\mathcal{Q}'\mathcal{Q})^{-1}\mathcal{Q}'\mathcal{Z} \quad (32)$$

instead. In our empirical analysis of [Autor et al. \(2013\)](#) we employ a ridge regression

version of (31) because: (i) The matrix $\sum_i S_i S_i'$ is ill-conditioned in that application,¹² which prevents the use of (31); and (ii) We focus on the specification employed in Autor et al. (2013), which does not contain controls with the structure required by (32). ■

5 Empirical Application: The China Syndrome

We illustrate the empirical relevance of our analysis by revisiting the seminal work of Autor et al. (2013) examining the impact of rising Chinese import competition on local US markets. We focus on the main specification of Autor et al. (2013), in which

$$Y_{it} = X_{it}\beta + W_{it}'\gamma_s + \varepsilon_{it} \quad (33)$$

with i indexing a commuting zone (CZ), t indexing one of two time periods, Y_{it} and X_{it} denoting the change in employment manufacturing and a measure of import exposure, and W_{it} denoting a set of CZ level characteristics.¹³ Autor et al. (2013) estimate the model in (33) by TSLS based on a Bartik instrument $Z_{it} = Z_t' S_{it}$ in which S_{it} and Z_t consist of a vector of lagged industry shares and a vector of Chinese imports growth to other countries for different industries. Their analysis also employs four digit SIC codes to define sectors, leading to a total of $p = 397$ sectors.

The standard errors in Autor et al. (2013) rely on an asymptotic approximation that implicitly conditions on the shocks. As we argued in Section 2.1, such standard errors corresponds to an identification strategy in which shares are viewed as exogenous; i.e.

$$E[S_{it}\varepsilon_{it}] = 0 \text{ for all } t. \quad (34)$$

We therefore begin by examining the viability of a constant treatment effects model in this application by employing the overidentification test of Section 4.2. Following Autor et al. (2013), we conduct inference by clustering at the state level leading to only 48 effective observations – a number much smaller than the $p \times T = 794$ number of moment restrictions represented by (34). In a Monte Carlo study based on this dataset we found the finite sample rejection probabilities of our test to be below the nominal level when employing all 794 restrictions; see the Supplemental Appendix. We therefore also test the validity of moment restrictions implied by (34) by aggregating shares to higher level SIC codes and focusing on particular time periods.

Table 1 reports the p-values for our overidentification tests under different choices of moment restrictions. Overall, our analysis yields evidence against the validity of the constant effects model under an identification strategy that implicitly conditions on shocks.

¹²Adao et al. (2019) instead address this ill-conditioning by dropping sectors from the shares vector.

¹³Specifically, we employ the specification in column (6) of Table 3 in Autor et al. (2013).

Table 1: Overidentification tests corresponding to conditioning on shocks

Moment Restrictions	# Moments	p-value
Four Digit SIC & all time periods	794	0.127
Four Digit SIC & $t = 1$	397	0.115
Four Digit SIC & $t = 2$	397	0.098
Three Digit SIC & all time periods	272	0.049
Three Digit SIC & $t = 1$	136	0.009
Three Digit SIC & $t = 2$	136	0.072
Two Digit SIC & all time periods	40	0.005
Two Digit SIC & $t = 1$	20	0.002
Two Digit SIC & $t = 2$	20	0.366

Table reports p-values for overidentification tests of constant treatment effects models identified by implicitly conditioning on the realization of the shocks; see Section 4.2. All tests are implemented by clustering at the state level and weighting commuting zones by their start of period population as in [Autor et al. \(2013\)](#).

[Goldsmith-Pinkham et al. \(2020\)](#) reach a similar conclusion, though we note that the overidentification tests they employ do not cluster at the state level and are not robust to a “large” number of moment restrictions. Given the implausibility of the constant treatment effects model, it is natural to ask whether TSLS can be attributed a causal interpretation through a heterogeneous treatment effects model instead. Unfortunately, Corollary 3.1 establishes that such an interpretation necessitates that shares be uncorrelated across sectors – a requirement readily contradicted by the data, which exhibits correlations for some sectors in excess of 0.9. We note, however, that as advocated by [Goldsmith-Pinkham et al. \(2020\)](#) it may still be possible to proceed by employing the shares as instruments without combining them into a Bartik instrument.

We next examine the viability of a constant effects model under an identification strategy in which the exogeneity of the Bartik instrument is due to the exogeneity of the shocks ([Adao et al., 2019](#); [Borusyak et al., 2022](#)). To this end, we implement the overidentification test of Section 4.3, which requires us to select a vector of moments ($\{g_j\}$ in (30)) and an estimator for $\mathcal{E} \equiv \mathcal{Z} - E[\mathcal{Z}|\mathcal{G}_n]$ (as in Remark 4.2). For the former, we select a total of twenty moments and let the functions $\{g_j\}$ depend on ε only. In particular, we select one function g_j to simply equal ε^2 , and the other nineteen to equal the Logit pdf centered at equispaced points in the support of ε ; see the Supplemental Appendix for details. Finally, as discussed in Remark 4.2, we estimate \mathcal{E} through Ridge regression and follow [Borusyak et al. \(2022\)](#) in clustering at the three level SIC code.

Table 2 reports the p-values for our test corresponding to different choice of Ridge penalties. Our results show some sensitivity of the p-value to the choice of Ridge penalty. This conclusion echoes [Adao et al. \(2019\)](#) who similarly find that the manner in which

Table 2: Overidentification tests corresponding to identification through shocks

	Ridge Penalty			
	1e-3	1e-4	1e-5	1e-6
p-value	0.0012	0.0074	0.0368	0.065

Table reports p-values for overidentification tests of constant treatment effects models identified through the exogeneity of shocks; see Section 4.3. All tests are implemented by clustering at the three digit SIC level and weighting commuting zones by their start of period population as in Autor et al. (2013).

they address the singularity of the shares design matrix affects their standard errors.¹⁴ Overall, we interpret Table 2 as providing evidence against the plausibility of the constant effects model under the identification strategy that attributes the exogeneity of the Bartik instruments to the shocks. Our preferred choice of penalty for this application is 1e-5, yielding a p-value of 0.0368, which in a Monte Carlo study based on this dataset yielded finite sample rejection probabilities closest to a 5% nominal level; see the Supplemental Appendix. It is also worth emphasizing that clustering at the three digit SIC level has important implications for whether TSLS can be attributed a causal interpretation in a heterogeneous effects model. For example, Corollary 3.2 implies that such a causal interpretation necessitates all industries within the same 3-digit SIC code to be (weakly) positively correlated. It is important to note, however, that the latter requirement is necessary but not sufficient for TSLS to have a causal interpretation; see Remark 3.2 for additional restrictions.

Finally, we note that the “just identified” long panel identification strategy of Section 3.2 is not viable in this application. Specifically, because there are only two time periods in Autor et al. (2013), it is not possible to learn the time series distribution of the shocks as required by such an approach. In other words, it would be foolhardy to invoke an asymptotic approximation based on T diverging to infinity when T equals two.

6 Summary and Recommendations

In this paper, we examined the testable implications of identifying restrictions employed to assign a causal interpretation to TSLS based on Bartik instruments. We developed specification tests for homogeneous effects models that are robust to heteroskedasticity, clustering, and weighting. Because our tests are based on the high dimensional central limit theorem, we expect them to be more robust than their alternatives to the “high” degree of overidentification present in shift-share designs. Finally, we argued that our overidentification tests are of central importance due to the potentially limited empirical scope of the natural alternatives to the homogeneous effects models.

¹⁴Adao et al. (2019) addressed the singularity of $\sum_i S_i S_i'$ by dropping sectors from their analysis. As stated by the authors, their standard errors are affected by what sectors are dropped; see page 3 in <https://github.com/kolesarm/ShiftShareSE/blob/master/doc/ShiftShareSE.pdf>

Our analysis has a number of important implications for applied work:

- We recommend that practitioners report overidentification tests corresponding to the identification strategy of [Goldsmith-Pinkham et al. \(2020\)](#) as well as that of [Adao et al. \(2019\)](#) and [Borusyak et al. \(2022\)](#). These tests can provide evidence as to which identification strategy is more compatible with the data.
- Practitioners relying on the identification strategy of [Goldsmith-Pinkham et al. \(2020\)](#) should be wary of assigning a causal interpretation to TSLS by relying on a heterogeneous effects model. Such an interpretation necessitates that shares be uncorrelated across sectors, which is empirically untenable. On the other hand, as advocated by [Goldsmith-Pinkham et al. \(2020\)](#), it may still be possible to employ the shares for causal inference by not combining them into a Bartik instrument.
- Practitioners should likewise be careful in assigning TSLS a causal interpretation by relying on a heterogeneous effects model under the identification strategy of [Adao et al. \(2019\)](#) and [Borusyak et al. \(2022\)](#). A necessary condition for TSLS to be causal is that shocks to different sectors not be negatively correlated. Reasoning through the correlation structure of shocks is thus important not only for the purposes of obtaining standard errors, but also for interpreting the TSLS estimand.
- In applications with long panels, it is possible to rely on an alternative identification strategy in which the randomness of the instrument is understood jointly over the cross sectional distribution of shares and the time series distribution of shocks. Practitioners relying on this approach, however, need to employ different standard errors that are governed by the length of the panel.

Appendix

This Appendix contains the proofs of all the results in the main text.

A.1 Proofs for Section 2

Proof of Proposition 2.1. First note that since $\{S_i, W_i, \varepsilon_i\}_{i=1}^n$ are i.i.d. and independent of \mathcal{G}_n and \mathcal{Z} is measurable with respect to \mathcal{G}_n it follows $E[W_i \varepsilon_i] = 0$ that

$$\frac{1}{n} \sum_{i=1}^n E[\varepsilon_i (Z_i - W_i' \pi_n) | \mathcal{G}_n] = \mathcal{Z}' [S_i \varepsilon_i].$$

It is therefore immediate that $E[S_i \varepsilon_i] = 0$ is a sufficient condition for (7) to hold with probability one over \mathcal{Z} . To see that this condition is also necessary, suppose by way of contradiction that $E[S_i \varepsilon_i] \neq 0$. By Lemma A.1.1, it then follows that there is a z^* in the support of \mathcal{Z} satisfying $(z^*)' E[S_i \varepsilon_i] \neq 0$. Assuming without loss of generality that $(z^*)' E[S_i \varepsilon_i] > 0$ and setting $B_\delta(z^*) \equiv \{z \in \mathbb{R}^p : \|z - z^*\|_2^2 < \delta\}$, we then obtain that $(E[S_i \varepsilon_i])' z > 0$ for all $z \in B_\delta(z^*)$ provided that $\delta > 0$ is chosen to be sufficiently small. Hence, employing that z^* is in the support of \mathcal{Z} yields that

$$P(\mathcal{Z}' E[S_i \varepsilon_i] \neq 0) \geq P(\mathcal{Z} \in B_\delta(z^*)) > 0,$$

which implies that for (7) to hold with probability one we must have $E[S_i \varepsilon_i] = 0$. ■

Proof of Proposition 2.2. First note that since $\mathcal{G}_n = \{S_i, W_i, \varepsilon_i\}_{i=1}^n$ and \mathcal{Z} is independent of \mathcal{G}_n by hypothesis, it follows from $Z_i = S_i' \mathcal{Z}$ by definition that we have

$$\frac{1}{n} \sum_{i=1}^n E[\varepsilon_i (Z_i - W_i' \pi_n) | \mathcal{G}_n] = \frac{1}{n} \sum_{i=1}^n \varepsilon_i (S_i' E[\mathcal{Z}] - W_i' \pi_n).$$

Since $E[Z_i | \{S_i, W_i, \varepsilon_i\}_{i=1}^n] = S_i' E[\mathcal{Z}]$, we can conclude that $E[Z_i | \{S_i, W_i, \varepsilon_i\}_{i=1}^n] = W_i' \pi_n$ is a sufficient condition for (10) to hold with probability one. In order to establish the converse direction, we next define the event A_n to be given by

$$A_n \equiv \{S_i' E[\mathcal{Z}] \neq W_i' \pi_n \text{ for some } 1 \leq i \leq n\}$$

and note that A_n is measurable with respect to $\{S_i, W_i\}$ by definition of π_n . For any realization of $\{S_i, W_i\}_{i=1}^n$ in the event A_n , Lemma A.1.1 allows us to conclude that there exists a $\{e_i^*\}_{i=1}^n$ in the support of $\{\varepsilon_i\}_{i=1}^n$ conditional on $\{S_i, W_i\}$ satisfying

$$\frac{1}{n} \sum_{i=1}^n e_i^* (S_i' E[\mathcal{Z}] - W_i' \pi_n) \neq 0.$$

Hence, defining $B_\delta(\{e_i^*\}_{i=1}^n) \equiv \{\{e_i\}_{i=1}^n : \sum_i (e_i - e_i^*)^2 < \delta^2\}$ we obtain from $\{e_i^*\}_{i=1}^n$ being in the support of $\{\varepsilon_i\}_{i=1}^n$ conditional on $\{S_i, W_i\}_{i=1}^n$ that for δ sufficiently small

$$\begin{aligned} P\left(\frac{1}{n} \sum_{i=1}^n \varepsilon_i (S_i' E[\mathcal{Z}] - W_i' \pi_n) \neq 0 \mid \{S_i, W_i\}_{i=1}^n\right) \\ \geq P(\{\varepsilon_i\}_{i=1}^n \in B_\delta(\{e_i^*\}_{i=1}^n) \mid \{S_i, W_i\}_{i=1}^n) > 0. \quad (\text{A.1}) \end{aligned}$$

In particular, since the conclusion in (A.1) applies to any $\{S_i, W_i\}_{i=1}^n$ in the event A_n , it follows that A_n must have probability zero in order for condition (10) to hold with probability one. Since $S_i' E[\mathcal{Z}] = E[Z_i \mid \{S_i, W_i, \varepsilon_n\}_{i=1}^n]$, the lemma follows. ■

Lemma A.1.1. *If the linear span of a set $V \subseteq \mathbb{R}^k$ has dimension k , then it follows that for any $0 \neq x \in \mathbb{R}^k$ there is a $v^* \in V$ such that $x'v^* \neq 0$.*

Proof. Since the dimension of the linear span of V equals k , it follows that there are vectors $\{v_j\}_{j=1}^k \subset \mathbb{R}^k$ and scalars $\{\alpha_j\}_{j=1}^k$ satisfying $x = \sum_j \alpha_j v_j$. In particular, since $x \neq 0$, we have $(\sum_j \alpha_j v_j)'x = \|x\|_2^2 > 0$ and hence that $v_{j^*}'x \neq 0$ for some j^* . ■

A.2 Proofs for Section 3.1.1

The results in Section 3.1.1 rely on the next set of regularity conditions. In the statement below, the notation $\text{supp}\{V\}$ refers to the support of a random variable V .

Assumption A.2.1. (i) $\{\Lambda_i, W_i\}_{i=1}^n$ are i.i.d. conditional on \mathcal{Z} ; (ii) $\text{supp}\{(\Lambda_i, W_i, \mathcal{Z})\} = \text{supp}\{\Lambda_i\} \times \text{supp}\{W_i\} \times \text{supp}\{\mathcal{Z}\}$; (iii) $\text{rank}\{\text{Var}\{S_i \mid W_i\}\} \geq p - 1$ and $\text{Var}\{S_{ij} \mid W_i\} > 0$ for all $1 \leq j \leq p$ with probability one; (iv) $(\lambda_{i1}, \dots, \lambda_{ip})$ has a continuous density w.r.t. Lebesgue measure; (v) $\text{supp}\{\mathcal{Z}\}$ includes a neighborhood of zero.

Assumption A.2.1(i) allows for dependence across observations though it implies that the marginal distribution of $(\Lambda_i, W_i, \mathcal{Z})$ is the same for all observations. Assumptions A.2.1(ii)(iv) further impose support restrictions on $(\Lambda_i, W_i, \mathcal{Z})$, while Assumption A.2.1(iii) states our restrictions on the covariance matrix of $S_i \in \mathbb{R}^p$ conditionally on W_i (denoted $\text{Var}\{S_i \mid W_i\}$). Assumption A.2.1(iii) allows for $\text{Var}\{S_i \mid W_i\}$ to be rank-deficient to recognize that the shares may sum up to one. Finally, Assumptions A.2.1(iv) simply demands that the “types” vector $(\lambda_{i1}, \dots, \lambda_{ip})$ be continuously distributed.

Proof of Proposition 3.1. First note that Assumption 3.1(ii) implies $(\Lambda_i, \beta_i, \varepsilon_i, \eta_i)$ is independent of (\mathcal{Z}, S_i) conditionally on (\mathcal{Z}, W_i) . Therefore, by Assumptions 3.1(i)(iii), we may apply Lemma A.2.1 with $\mathcal{G}_n = \{\mathcal{Z}\}$ to conclude that

$$\begin{aligned} E\left[X_i \dot{Z}_{i,n} \mid \Lambda_i, W_i, \mathcal{Z}\right] &= E\left[(\mathcal{Z}' \Lambda_i S_i) \dot{Z}_{i,n} \mid \Lambda_i, W_i, \mathcal{Z}\right] \\ &= \mathcal{Z}' \Lambda_i E\left[S_i (S_i' - E[S_i' \mid W_i, \mathcal{Z}]) \mid \Lambda_i, W_i, \mathcal{Z}\right] \mathcal{Z} = \mathcal{Z}' (\Lambda_i \text{Var}\{S_i \mid W_i\}) \mathcal{Z}, \quad (\text{A.2}) \end{aligned}$$

where the second equality holds by $Z_i = S_i' \mathcal{Z}$ and Assumption 3.1(iii), and the final equality holds by Assumption 3.1(ii). The claim of the proposition therefore follows from result (A.2), the definition of $\omega_{i,n}$ in (13), and Lemma A.2.2. ■

Proof of Corollary 3.1. Suppose that condition (15) holds for some $j \neq k$ and define

$$\sigma_l^2(W_i) \equiv \text{Var}\{S_{il}|W_i\} \quad \sigma_{jk}(W_i) = \text{Cov}\{S_{ij}, S_{ik}|W_i\}.$$

Next, note that by Proposition 3.1 a necessary condition for $\omega_{i,n}$ to be positive for all $1 \leq i \leq n$ with probability one is that the corresponding 2×2 matrix

$$\begin{pmatrix} \lambda_{ij} & 0 \\ 0 & \lambda_{ik} \end{pmatrix} \begin{pmatrix} \sigma_j^2(W_i) & \sigma_{jk}(W_i) \\ \sigma_{jk}(W_i) & \sigma_k^2(W_i) \end{pmatrix} = \begin{pmatrix} \lambda_{ij}\sigma_j^2(W_i) & \lambda_{ij}\sigma_{jk}(W_i) \\ \lambda_{ik}\sigma_{jk}(W_i) & \lambda_{ik}\sigma_k^2(W_i) \end{pmatrix}$$

be either positive semi-definite or negative semi-definite with probability one. Hence, setting $\bar{\lambda}_i = (\lambda_{ij} + \lambda_{ik})/2$ and noting that for any 2×2 matrix A and $a \in \mathbb{R}^2$ we have $a' A a = a'(A + A')a/2$, it follows that a necessary condition for $\omega_{i,n}$ to be positive for all $1 \leq i \leq n$ with probability one is that the matrices Ω_i defined by

$$\Omega_i \equiv \begin{pmatrix} \lambda_{ij}\sigma_j^2(W_i) & \bar{\lambda}_i\sigma_{jk}(W_i) \\ \bar{\lambda}_i\sigma_{jk}(W_i) & \lambda_{ik}\sigma_k^2(W_i) \end{pmatrix}$$

be either positive semi-definite with probability one or negative semi-definite with probability one. However, in order for Ω_i to be positive semi-definite or negative-semidefinite with probability one we must have $\lambda_{ij}\lambda_{ik} \geq 0$ due to $\min\{\sigma_j^2(W_i), \sigma_k^2(W_i)\} > 0$ by Assumption A.2.1(iii). Hence, since the determinant of Ω_i equals the product of its eigenvalues, a necessary condition for $\omega_{i,n}$ to be positive for all i with probability one is

$$1 = P(\det\{\Omega_i\} \geq 0 \text{ and } \lambda_{ij}\lambda_{ik} \geq 0).$$

However, setting $\rho_{jk}(W_i) \equiv \sigma_{jk}(W_i)/\sigma_j(W_i)\sigma_k(W_i)$ we obtain by direct calculation that

$$\begin{aligned} P(\det\{\Omega_i\} \geq 0 \text{ and } \lambda_{ij}\lambda_{ik} \geq 0) &= P(\lambda_{ij}\lambda_{ik} - \bar{\lambda}_i^2 \rho_{jk}^2(W_i) \geq 0 \text{ and } \lambda_{ij}\lambda_{ik} \geq 0) \\ &\leq P\left(1 \geq \frac{\rho_{jk}^2(W_i)}{4} \left(\frac{\lambda_{ij}}{\lambda_{ik}} + \frac{\lambda_{ik}}{\lambda_{ij}} + 2\right)\right) < 1, \end{aligned}$$

where the first inequality follows from $\lambda_{ij}\lambda_{ik} \neq 0$ with probability one by Assumption A.2.1(iv), and the final from Assumption A.2.1(ii), the support of $\lambda_{ij}/\lambda_{ik}$ being unbounded, and $\rho_{jk}^2(W_i) \neq 0$ with positive probability due to condition (15). ■

Lemma A.2.1. *Suppose that \mathcal{Z} and $\{Y_i, X_i, W_i, S_i\}_{i=1}^n$ satisfy equations (11) and (12), $E[Z_i|W_i, \mathcal{G}_n] = W_i' \pi_n$, and $(\Lambda_i, \beta_i, \varepsilon_i, \eta_i) \perp\!\!\!\perp (\mathcal{Z}, S_i)$ conditionally on (W_i, \mathcal{G}_n) . Then:*

$$E\left[X_i \dot{Z}_{i,n} \middle| \Lambda_i, W_i, \mathcal{G}_n\right] = E[(\mathcal{Z}' \Lambda_i S_i) \dot{Z}_{i,n} | \Lambda_i, W_i, \mathcal{G}_n]. \quad (\text{A.3})$$

If in addition $\sum_{i=1}^n E[X_i \dot{Z}_{i,n} | \mathcal{G}_n] \neq 0$, then it also follows that equation (13) holds.

Proof. First note that for any function f of $(\Lambda_i, \beta_i, \varepsilon_i, \eta_i, W_i)$ it follows from the requirements that $(\Lambda_i, \beta_i, \varepsilon_i, \eta_i) \perp\!\!\!\perp (\mathcal{Z}, S_i)$ conditionally on (\mathcal{G}_n, W_i) and $Z_i = \mathcal{Z}' S_i$ that

$$\begin{aligned} E \left[\dot{Z}_{i,n} f(\Lambda_i, \beta_i, \varepsilon_i, \eta_i, W_i) \middle| \mathcal{G}_n \right] \\ = E \left[E \left[\dot{Z}_{i,n} | W_i, \mathcal{G}_n \right] E \left[f(\Lambda_i, \beta_i, \varepsilon_i, \eta_i, W_i) | W_i, \mathcal{G}_n \right] \middle| \mathcal{G}_n \right] = 0, \end{aligned} \quad (\text{A.4})$$

where the final equality follows from $E[Z_i | W_i, \mathcal{G}_n] = W_i' \pi_n$. Combining result (A.4) with the specifications of the first and second stages in (11) and (12) then yields that

$$E \left[Y_i \dot{Z}_{i,n} | \mathcal{G}_n \right] = E \left[X_i \dot{Z}_{i,n} \beta_i | \mathcal{G}_n \right] = E \left[(\mathcal{Z}' \Lambda_i S_i) \dot{Z}_{i,n} \beta_i | \mathcal{G}_n \right]. \quad (\text{A.5})$$

Letting \mathcal{Z}_j denote the j^{th} coordinate of \mathcal{Z} , S_{ij} the j^{th} coordinate of S_i , and recalling that $\Lambda_i = \text{diag}\{\lambda_{i1}, \dots, \lambda_{ip}\}$ we may also employ (12) and (A.4) to obtain

$$\begin{aligned} E \left[X_i \dot{Z}_{i,n} | \Lambda_i, W_i, \mathcal{G}_n \right] &= E \left[(\mathcal{Z}' \Lambda_i S_i) \dot{Z}_{i,n} | \Lambda_i, W_i, \mathcal{G}_n \right] \\ &= \sum_{j=1}^p E \left[(\mathcal{Z}_j \lambda_{ij} S_{ij}) \dot{Z}_{i,n} | \Lambda_i, W_i, \mathcal{G}_n \right], \end{aligned} \quad (\text{A.6})$$

which verifies that (A.3) indeed holds. Finally, observe that (Λ_i, β_i) being independent of (\mathcal{Z}, S_i) conditionally on (W_i, \mathcal{G}_n) further allows us to conclude that

$$\begin{aligned} E[(\mathcal{Z}' \Lambda_i S_i) \dot{Z}_{i,n} \beta_i | \mathcal{G}_n] &= \sum_{j=1}^p E[E[\mathcal{Z}_j S_{ij} \dot{Z}_{i,n} | W_i, \mathcal{G}_n] E[\lambda_{ij} \beta_i | W_i, \mathcal{G}_n] | \mathcal{G}_n] \\ &= \sum_{j=1}^p E[E[\mathcal{Z}_j S_{ij} \dot{Z}_{i,n} | W_i, \mathcal{G}_n] \lambda_{ij} E[\beta_i | \Lambda_i, W_i, \mathcal{G}_n] | \mathcal{G}_n] \\ &= E[X_i \dot{Z}_{i,n} | \Lambda_i, W_i, \mathcal{G}_n] E[\beta_i | \Lambda_i, W_i, \mathcal{G}_n] | \mathcal{G}_n, \end{aligned} \quad (\text{A.7})$$

where the final equality follows from result (A.6) and $\Lambda_i \perp\!\!\!\perp (\mathcal{Z}, S_i)$ conditionally on (W_i, \mathcal{G}_n) . The claim that (13) holds therefore follows from the definition of $\beta_{0,n}$, the rank condition $\sum_{i=1}^n E[X_i \dot{Z}_{i,n} | \mathcal{G}_n] \neq 0$ and results (A.5), (A.6), and (A.7). ■

Lemma A.2.2. *Let Assumption A.2.1 hold. Then, for n sufficiently large, it follows*

$$P \left(\min_{1 \leq i \leq n} \mathcal{Z}' (\Lambda_i \text{Var}\{S_i | W_i\}) \mathcal{Z} \geq 0 \text{ or } \max_{1 \leq i \leq n} \mathcal{Z}' (\Lambda_i \text{Var}\{S_i | W_i\}) \mathcal{Z} \leq 0 \right) = 1 \quad (\text{A.8})$$

if and only if either $\Lambda_i \text{Var}\{S_i | W_i\}$ is positive semi-definite with probability one for all i , or $\Lambda_i \text{Var}\{S_i | W_i\}$ is negative semi-definite with probability one for all i .

Proof. For notational simplicity we first set $\Gamma_i \equiv \Lambda_i \text{Var}\{S_i | W_i\}$ and use “ $\Gamma_i \geq 0$ ” and

“ $\Gamma_i \leq 0$ ” to denote that Γ_i is positive semi-definite and negative-semidefinite respectively. Then note that if either $P(\Gamma_i \geq 0) = 1$ for all i or $P(\Gamma_i \leq 0) = 1$ for all i , then (A.8) immediately holds. For the opposite direction, we next observe that

$$\begin{aligned} & P\left(\min_{1 \leq i \leq n} \mathcal{Z}'\Gamma_i \mathcal{Z} \geq 0 \text{ or } \max_{1 \leq i \leq n} \mathcal{Z}'\Gamma_i \mathcal{Z} \leq 0\right) \\ &= P\left(\min_{1 \leq i \leq n} \mathcal{Z}'\Gamma_i \mathcal{Z} \geq 0\right) + P\left(\max_{1 \leq i \leq n} \mathcal{Z}'\Gamma_i \mathcal{Z} \leq 0\right) - P(\mathcal{Z}'\Gamma_i \mathcal{Z} = 0 \text{ for all } 1 \leq i \leq n) \\ &= E\left[P^n(\mathcal{Z}'\Gamma_1 \mathcal{Z} \geq 0 | \mathcal{Z}) + P^n(\mathcal{Z}'\Gamma_1 \mathcal{Z} \leq 0 | \mathcal{Z}) - P^n(\mathcal{Z}'\Gamma_1 \mathcal{Z} = 0 | \mathcal{Z})\right], \end{aligned} \quad (\text{A.9})$$

where the second equality follows from Assumption A.2.1(i). However, since $a^n = o(1)$ for any $a \in [0, 1)$ as n tends to infinity, result (A.9) implies that in order for the claim in (A.8) to hold for n sufficiently large we must have with probability one over \mathcal{Z} that

$$\max\{P(\mathcal{Z}'\Gamma_1 \mathcal{Z} \geq 0 | \mathcal{Z}), P(\mathcal{Z}'\Gamma_1 \mathcal{Z} \leq 0 | \mathcal{Z})\} = 1. \quad (\text{A.10})$$

By Assumption A.2.1(ii), $\text{supp}\{(\Gamma_1, \mathcal{Z})\} = \text{supp}\{\Gamma_1\} \times \text{supp}\{\mathcal{Z}\}$ and hence the distribution of Γ_1 is absolutely continuous with respect to the distribution of Γ_1 conditionally on \mathcal{Z} . Defining the sets S_+ and S_- to be given by

$$\begin{aligned} S_+ &\equiv \{z \in \mathbb{R}^p : P(z'\Gamma_1 z \geq 0) = 1\} \\ S_- &\equiv \{z \in \mathbb{R}^p : P(z'\Gamma_1 z \leq 0) = 1\} \end{aligned} \quad (\text{A.11})$$

it therefore follows from result (A.10) holding with probability one that we must have

$$P(\mathcal{Z} \in S_+ \cup S_-) = 1. \quad (\text{A.12})$$

Next, set $\Gamma^* \equiv \Lambda^* \Sigma^*$ with Σ^* any point in the support of $\text{Var}\{S_1 | W_1\}$ and $\Lambda^* = \text{diag}\{\lambda_1^*, \dots, \lambda_p^*\}$ for any point $(\lambda_1^*, \dots, \lambda_p^*)$ at which the density of $(\lambda_{11}, \dots, \lambda_{1p})$ is strictly positive. We next aim to show that Γ^* is either positive semi-definite or negative semi-definite. We proceed by contradiction: Suppose that Γ^* is neither positive semi-definite nor negative semi-definite. Then, by Lemma A.2.3, Assumption A.2.1(iii), the density of $(\lambda_{11}, \dots, \lambda_{1p})$ being continuous and strictly positive at $(\lambda_1^*, \dots, \lambda_p^*)$, and Assumption A.2.1(v), there are $z_0 \in \text{supp}\{\mathcal{Z}\}$ and $\Lambda_0, \tilde{\Lambda}_0 \in \text{supp}\{\Lambda_1\}$ satisfying

$$z_0'(\Lambda_0 \Sigma^*) z_0 > 0 \quad z_0'(\tilde{\Lambda}_0 \Sigma^*) z_0 < 0.$$

We may therefore select sufficiently small neighborhoods $N(z_0), N(\Lambda_0), N(\tilde{\Lambda}_0), N(\Sigma^*)$ of $z_0, \Lambda_0, \tilde{\Lambda}_0$, and Σ^* such that the following inequalities are satisfied

$$\begin{aligned} & z'(\Lambda \Sigma) z > 0 \text{ for all } (z, \Lambda, \Sigma) \in N(z_0) \times N(\Lambda_0) \times N(\Sigma^*) \\ & z'(\Lambda \Sigma) z < 0 \text{ for all } (z, \Lambda, \Sigma) \in N(z_0) \times N(\tilde{\Lambda}_0) \times N(\Sigma^*). \end{aligned}$$

Then note that since $\Lambda_0 \Sigma^* \in \text{supp}\{\Lambda_1 \text{Var}\{S_1|W_1\}\}$ by Assumption A.2.1(ii), it follows

$$\begin{aligned} P(z'(\Lambda_1 \text{Var}\{S_1|W_1\})z > 0 \text{ for all } z \in N(z_0)) \\ \geq P((\Lambda_1, \text{Var}\{S_1|W_1\}) \in N(\Lambda_0) \times N(\Sigma^*)) > 0. \end{aligned} \quad (\text{A.13})$$

By identical arguments, but relying on $N(\tilde{\Lambda}_0) \times N(\Sigma^*)$ instead, similarly yield that

$$P(z'(\Lambda_1 \text{Var}\{S_1|W_1\})z < 0 \text{ for all } z \in N(z_0)) > 0. \quad (\text{A.14})$$

However, because $P(\mathcal{Z} \in N(z_0)) > 0$ due to $z_0 \in \text{supp}\{\mathcal{Z}\}$, results (A.13) and (A.14) together contradict (A.12) and therefore Γ^* must be either positive semi-definite or negative semi-definite as claimed. Thus, we have so far shown that

$$P(\Gamma_1 \geq 0 \text{ or } \Gamma_1 \leq 0) = 1. \quad (\text{A.15})$$

To conclude, note that by result (A.12) and Lemma A.2.4 either S_+ or S_- (or both) must contain p linearly independent vectors. If S_+ contains p linearly independent vectors $\{s_j\}_{j=1}^p$, then we obtain by the definition of S_+ in (A.11) that

$$\begin{aligned} P(\Gamma_1 = 0) &\leq P(\Gamma_1 \leq 0) \leq P(z'\Gamma_1 z \leq 0 \text{ for all } z \in S_+) \\ &\leq P(s'_j \Gamma_1 s_j \leq 0 \text{ for } 1 \leq j \leq p) = P(s'_j \Gamma_1 s_j = 0 \text{ for } 1 \leq j \leq p). \end{aligned} \quad (\text{A.16})$$

Next note that $z'\Gamma_1 z = z'(\Gamma_1 + \Gamma'_1)z/2$ for any $z \in \mathbb{R}^p$ and use that $(\Gamma_1 + \Gamma'_1)$ is diagonalizable and $s'_j(\Gamma_1 + \Gamma'_1)s_j = 0$ for all $1 \leq j \leq p$ implies that all the eigenvalues of $(\Gamma_1 + \Gamma'_1)$ equal zero due to $\{s_j\}_{j=1}^p$ being linearly independent to conclude that

$$\begin{aligned} P(s'_j \Gamma_1 s_j = 0 \text{ for } 1 \leq j \leq p) &= P(s'_j(\Gamma_1 + \Gamma'_1)s_j = 0 \text{ for } 1 \leq j \leq p) \\ &= P(\Gamma_1 = -\Gamma'_1) \leq P((\lambda_{11}, \dots, \lambda_{1p}) = 0) \leq P(\Gamma_1 = 0), \end{aligned} \quad (\text{A.17})$$

where the first inequality follows from $\text{Var}\{S_{1j}|W_i\} > 0$ for all $1 \leq j \leq p$ by Assumption A.2.1(iii) and the second inequality by $\Gamma_1 = \text{diag}\{(\lambda_{11}, \dots, \lambda_{1p})\} \text{Var}\{S_1|W_1\}$. Hence, combining results (A.16) and (A.17) we obtain that if S_+ contains p linearly independent vectors, then $P(\Gamma_1 = 0) = P(\Gamma_1 \leq 0)$, which together with result (A.15) yields

$$1 = P(\Gamma_1 \geq 0 \text{ or } \Gamma_1 \leq 0) = P(\Gamma_1 \geq 0) + P(\Gamma_1 \leq 0) - P(\Gamma_1 = 0) = P(\Gamma_1 \geq 0).$$

Similarly, if S_- instead contains p linearly independent vectors, then it is possible to show that (A.15) implies that $P(\Gamma_1 \leq 0) = 1$ and therefore the lemma follows. ■

Lemma A.2.3. *Let $\Gamma \equiv D\Sigma$ with D a $p \times p$ diagonal matrix and Σ a $p \times p$ positive semi-definite matrix with $\text{rank}\{\Sigma\} \geq p - 1$. If Γ is neither positive semi-definite nor negative semi-definite, then for any $\delta > 0$ there is a $z \in \mathbb{R}^p$ and diagonal $p \times p$ matrices*

H_1, H_2 satisfying $\|z\| \vee \|D - H_1\| \vee \|D - H_2\| < \delta$, $z'(H_1\Sigma)z > 0$, and $z'(H_2\Sigma)z < 0$.

Proof. Let $\bar{\Gamma} = (\Gamma + \Gamma')/2$ and note that since $a'\Gamma a = a'\Gamma' a$ for any $a \in \mathbb{R}^p$, we have

$$a'\Gamma a = a'\bar{\Gamma} a \text{ for all } a \in \mathbb{R}^p. \quad (\text{A.18})$$

Next observe that $\bar{\Gamma}$ is a symmetric matrix and let $\{\pi_i\}_{i=1}^p$ denote its eigenvalues and $\{v_i\}_{i=1}^p$ its corresponding eigenvectors. If Γ is neither positive semi-definite nor negative semi-definite, then result (A.18) implies that $\bar{\Gamma}$ must have a strictly positive eigenvalue and a strictly negative eigenvalue. Assuming without loss of generality that $\pi_1 > 0$ and $\pi_2 < 0$, let $\alpha > 0$ and define the vectors c_1 and c_2 to be given by

$$c_1 \equiv \alpha v_1 + \alpha \frac{\sqrt{\pi_1}}{\sqrt{|\pi_2|}} v_2 \quad c_2 \equiv \alpha v_1 - \alpha \frac{\sqrt{\pi_1}}{\sqrt{|\pi_2|}} v_2.$$

Next note that $\|c_1\|^2 = \|c_2\|^2 = \alpha^2(1 + \pi_1/|\pi_2|)$ and that therefore we may set $\|c_1\| = \|c_2\| < \delta$ by selecting α to be sufficiently small. Moreover, by direct calculation we have

$$c_1'\bar{\Gamma}c_1 = c_2'\bar{\Gamma}c_2 = \alpha^2\pi_1 + \alpha^2 \frac{\pi_1}{|\pi_2|}\pi_2 = 0, \quad (\text{A.19})$$

where the final equality follows from $\pi_2 < 0$. Next observe that since $\text{rank}\{\Sigma^{1/2}\} = \text{rank}\{\Sigma\} \geq p - 1$ and $c_1, c_2 \in \mathbb{R}^p$ are linearly independent, it follows that $\Sigma^{1/2}c_j \neq 0$ for some $j \in \{1, 2\}$. Assuming without loss of generality that $\Sigma^{1/2}c_1 \neq 0$, we then set $z = c_1$ and note that results (A.18), (A.19), and $\Sigma^{1/2}z \neq 0$ imply that

$$z'\Gamma z = 0 \text{ and } z'\Sigma z > 0. \quad (\text{A.20})$$

Setting $x \equiv \Sigma z$ and letting x_j and z_j denote the j^{th} coordinates of x and z , then note

$$0 < z'\Sigma z = \sum_{j=1}^p z_j x_j,$$

which implies that $z_j x_j > 0$ for some $1 \leq j \leq p$. Assuming without loss of generality that $z_1 x_1 > 0$, then set $e_1 = (1, 0, \dots, 0)'$ and for $\xi > 0$ define the matrices

$$H_1 \equiv D + \xi e_1 e_1' \quad H_2 \equiv D - \xi e_1 e_1',$$

noting that $\|D - H_1\| = \|D - H_2\| < \delta$ provided ξ is chosen sufficiently small. In addition, $\Gamma \equiv D\Sigma$, $x \equiv \Sigma z$, result (A.20), and $x_1 z_1 > 0$ together yield that

$$z'(H_1\Sigma)z = z'((D + \xi e_1 e_1')\Sigma)z = z'\Gamma z + \xi z_1 x_1 > 0.$$

Identical arguments imply $z'(H_2\Sigma)z < 0$, and hence the lemma follows. ■

Lemma A.2.4. *If Assumption A.2.1(v) holds and A_1, A_2 satisfy $P(\mathcal{Z} \in A_1 \cup A_2) = 1$, then it follows that $\text{span}\{A_i\} = \mathbb{R}^p$ for some $i \in \{1, 2\}$.*

Proof. By way of contradiction, suppose that $\text{span}\{A_i\} \neq \mathbb{R}^p$ for $i \in \{1, 2\}$. Then note that $A_i \subseteq V_i$ for some vector subspace $V_i \subset \mathbb{R}^p$ and that for each i we may find a $v_i \in \mathbb{R}^p$ that is orthogonal to V_i and satisfies $\|v_i\|^2 = 1$. Next define v^* to equal

$$v^* \equiv \begin{cases} v_1 & \text{if } |\langle v_1, v_2 \rangle| = 1 \\ v_1 + v_2 & \text{if } |\langle v_1, v_2 \rangle| \neq 1 \end{cases}$$

and note that $\langle v^*, v_i \rangle \neq 0$ for $i \in \{1, 2\}$ due to $\|v_i\|^2 = 1$. Since v_i is orthogonal to V_i it follows that $\xi v^* \notin V_i$ and hence $\xi v^* \notin V_1 \cup V_2$ for any $\xi > 0$. In particular, we obtain that $\xi v^* \notin A_1 \cup A_2$ due to $A_1 \cup A_2 \subseteq V_1 \cup V_2$. However, for ξ sufficiently small, ξv^* belongs to the support of \mathcal{Z} by Assumption A.2.1(v). Thus, selecting $N(\xi v^*)$ to be a neighborhood of ξv^* sufficiently small to ensure that $N(\xi v^*) \subseteq (A_1 \cup A_2)^c$ we obtain

$$P(\mathcal{Z} \in (A_1 \cup A_2)^c) \geq P(\mathcal{Z} \in N(\xi v^*)) > 0, \quad (\text{A.21})$$

where the final inequality follows from ξv^* being in the support of \mathcal{Z} . Since (A.21) contradicts the hypothesis that $P(\mathcal{Z} \in A_1 \cup A_2) = 1$, the lemma follows. ■

A.3 Proofs for Section 3.1.2

The results in Section 3.1.2 rely on the following regularity conditions.

Assumption A.3.1. *For $\mathcal{G}_n = \{S_i, W_i, \Lambda_i, \beta_i, \varepsilon_i, \eta_i\}_{i=1}^n$ and $\mathcal{W}_n = \{W_i\}_{i=1}^n$ we have: (i) $\text{Var}\{\mathcal{Z}|\mathcal{G}_n\} = \text{Var}\{\mathcal{Z}|\mathcal{W}_n\}$ and $\text{Var}\{\mathcal{Z}_j|\mathcal{W}_n\} > 0$ for all $1 \leq j \leq p$ with probability one; (ii) $\{\Lambda_i, S_i\}_{i=1}^n$ are i.i.d. conditionally on \mathcal{W}_n ; (iii) $\text{supp}\{(\Lambda_i, S_i, \mathcal{W}_n)\} = \text{supp}\{\Lambda_i\} \times \text{supp}\{S_i\} \times \text{supp}\{\mathcal{W}_n\}$; (iv) $\text{supp}\{S_i/\|S_i\|_1\} = \{s \in \mathbb{R}^p : s \geq 0 \text{ and } \|s\|_1 = 1\}$.*

Assumptions A.3.1(i)-(iii) impose conditions that simplify our analysis. In turn, Assumption A.3.1(iv) demands that the support of the shares be sufficiently rich – in its statement, $\|\cdot\|_1$ denotes the ℓ_1 norm $\|s\|_1 = \sum_{i=1}^n |s^{(i)}|$ for any vector $s = (s^{(1)}, \dots, s^{(p)})' \in \mathbb{R}^p$. We note that Assumption A.3.1(iv) allows the shares to sum to less than one, as is sometimes the case in empirical applications.

Proof of Proposition 3.2. First note that setting $\mathcal{G}_n = \{S_i, W_i, \Lambda_i, \beta_i, \varepsilon_i, \eta_i\}_{i=1}^n$ implies that $(\Lambda_i, \beta_i, \varepsilon_i, \eta_i) \perp\!\!\!\perp (\mathcal{Z}, S_i)$ conditionally on (W_i, \mathcal{G}_n) . Hence, Assumption 3.2 yields

$$\begin{aligned} E \left[X_i \dot{Z}_{i,n} | \Lambda_i, W_i, \mathcal{G}_n \right] &= E \left[(\mathcal{Z}' \Lambda_i S_i) \dot{Z}_{i,n} | \mathcal{G}_n \right] \\ &= S_i' \Lambda_i E \left[\mathcal{Z}(\mathcal{Z}' - E[\mathcal{Z}' | \mathcal{G}_n]) | \mathcal{G}_n \right] S_i = S_i' (\Lambda_i \text{Var}\{\mathcal{Z} | \mathcal{G}_n\}) S_i, \quad (\text{A.22}) \end{aligned}$$

where the first equality follows from Lemma A.2.1 and $(\Lambda_i, W_i) \in \mathcal{G}_n$, the second equality holds by $Z_i = S'_i \mathcal{Z}$ and Assumption 3.2(ii), and the third equality holds by direct calculation. The claim of the proposition therefore follows from result (A.22), the definition of $\omega_{i,n}$ in (13), and Lemma A.3.1. ■

Proof of Corollary 3.2. Suppose condition (19) holds for some $j \neq k$ and define

$$\sigma_l^2(\mathcal{G}_n) \equiv \text{Var}\{\mathcal{Z}_l | \mathcal{G}_n\} \quad \sigma_{jk}^2(\mathcal{G}_n) = \text{Cov}\{\mathcal{Z}_j, \mathcal{Z}_k | \mathcal{G}_n\} \quad \rho_{jk}(\mathcal{G}_n) \equiv \text{Corr}\{\mathcal{Z}_j, \mathcal{Z}_k | \mathcal{G}_n\}.$$

Also note that Assumptions A.3.1(i)(iii) and Proposition 3.2 together imply that a necessary condition for $\omega_{i,n}$ to be positive for all i with probability one is that

$$P(\Lambda_i \geq 0) = 1 \text{ or } P(\Lambda_i \leq 0) = 1.$$

In what follows, we assume that $P(\Lambda_i \geq 0) = 1$ and note that the case in which $P(\Lambda_i \leq 0) = 1$ follows by identical arguments. Next note that by Proposition 3.2, a necessary condition for $\omega_{i,n}$ to be positive for all i with probability one is that

$$(\lambda_j s_j, \lambda_k s_k) \begin{pmatrix} \sigma_j^2(\mathcal{G}_n) & \sigma_{jk}(\mathcal{G}_n) \\ \sigma_{jk}(\mathcal{G}_n) & \sigma_k^2(\mathcal{G}_n) \end{pmatrix} \begin{pmatrix} s_j \\ s_k \end{pmatrix} \quad (\text{A.23})$$

be either positive for all $(s_j, s_k) \in \mathbb{R}_+$ with probability one over \mathcal{G}_n , or negative for all $(s_j, s_k) \in \mathbb{R}_+$ with probability one over \mathcal{G}_n . However, Assumption A.3.1(i), $P(\lambda_{ij} = 0) = 0$, and $P(\lambda_{ij} \geq 0) = 1$ imply that setting $(s_j, s_k) = (1, 0)$ yields a strictly positive number in (A.23) with probability one. Hence, we can conclude that a necessary condition for $\omega_{i,n}$ to be positive for all i with probability one is that we have

$$\begin{aligned} 1 &= P \left(\inf_{s_j, s_k \geq 0} \{s_j^2 \lambda_j \sigma_j^2(\mathcal{G}_n) + s_k^2 \lambda_k \sigma_k^2(\mathcal{G}_n) + s_j s_k \sigma_{jk}(\mathcal{G}_n) (\lambda_j + \lambda_k)\} \geq 0 \right) \\ &= P \left(\inf_{s_j, s_k \geq 0} \{s_j^2 \lambda_j + s_k^2 \lambda_k + s_j s_k \rho_{jk}(\mathcal{G}_n) (\lambda_j + \lambda_k)\} \geq 0 \right), \\ &= P \left(\inf_{s_k \geq 0} \left\{ s_k^2 \left(\lambda_k - \frac{\rho_{jk}^2(\mathcal{G}_n)}{4\lambda_j} (\lambda_j + \lambda_k)^2 \right) \right\} \geq 0 \right) \end{aligned} \quad (\text{A.24})$$

where in the second equality we employed Assumption A.3.1(i) and in the third equality we profiled out s_j by setting $s_j = -s_k \rho(\mathcal{G}_n) (\lambda_j + \lambda_k) / 2\lambda_j$. However, note that

$$P \left(\lambda_k - \frac{\rho_{jk}^2(\mathcal{G}_n)}{4\lambda_j} (\lambda_j + \lambda_k)^2 < 0 \right) = P \left(1 < \frac{\rho_{jk}^2(\mathcal{G}_n)}{4} \left(\frac{\lambda_j}{\lambda_k} + \frac{\lambda_k}{\lambda_j} + 2 \right) \right) > 0 \quad (\text{A.25})$$

where the inequality holds due to condition (3.2), Assumptions A.3.1(i)(iii), and the support of λ_j/λ_k being unbounded. Since result (A.25) implies the necessary condition in (A.24) cannot hold, the claim of the corollary follows. ■

Lemma A.3.1. *Let Assumption A.3.1 hold and $\mathcal{G}_n = \{S_i, W_i, \Lambda_i, \beta_i, \varepsilon_i, \eta_i\}_{i=1}^n$. Then,*

$$P \left(\min_{1 \leq i \leq n} S'_i(\Lambda_i \text{Var}\{\mathcal{Z}|\mathcal{G}_n\})S_i \geq 0 \text{ or } \max_{1 \leq i \leq n} S'_i(\Lambda_i \text{Var}\{\mathcal{Z}|\mathcal{G}_n\})S_i \leq 0 \right) = 1 \quad (\text{A.26})$$

for all n sufficiently large if and only if the distribution of $\Lambda_i \text{Var}\{\mathcal{Z}|\mathcal{G}_n\}$ satisfies

$$P \left(\sup_{s \in \mathbb{R}_+^p} s'(\Lambda_i \text{Var}\{\mathcal{Z}|\mathcal{G}_n\})s \leq 0 \right) = 1 \text{ or } P \left(\inf_{s \in \mathbb{R}_+^p} s'(\Lambda_i \text{Var}\{\mathcal{Z}|\mathcal{G}_n\})s \geq 0 \right) = 1. \quad (\text{A.27})$$

Proof. First note that since $S_i \in \mathbb{R}_+^p$ by Assumption A.3.1(iv), it follows that condition (A.27) implies (A.26) holds. For the reverse direction, set $\mathcal{W}_n = \{W_i\}_{i=1}^n$ and $\Gamma_{i,n} = \Lambda_i \text{Var}\{\mathcal{Z}|\mathcal{W}_n\}$ for notational simplicity, and employ Assumptions A.3.1(i) to obtain

$$\begin{aligned} & P \left(\min_{1 \leq i \leq n} S'_i(\Lambda_i \text{Var}\{\mathcal{Z}|\mathcal{G}_n\})S_i \geq 0 \text{ or } \max_{1 \leq i \leq n} S'_i(\Lambda_i \text{Var}\{\mathcal{Z}|\mathcal{G}_n\})S_i \leq 0 \right) \\ &= P \left(\min_{1 \leq i \leq n} S'_i \Gamma_{i,n} S_i \geq 0 \right) + P \left(\max_{1 \leq i \leq n} S'_i \Gamma_{i,n} S_i \leq 0 \right) - P \left(S'_i \Gamma_{i,n} S_i = 0 \text{ for all } i \right) \\ &= E \left[P^n \left(S'_i \Gamma_{i,n} S_i \geq 0 | \mathcal{W}_n \right) + P^n \left(S'_i \Gamma_{i,n} S_i \leq 0 | \mathcal{W}_n \right) - P^n \left(S'_i \Gamma_{i,n} S_i = 0 | \mathcal{W}_n \right) \right], \end{aligned} \quad (\text{A.28})$$

where the second equality follows from Assumption A.3.1(ii). For $n \geq 2$, result (A.28) and the law of iterated expectations together imply that in order for the equality in (A.26) to hold we must have with probability one over $(\Lambda_i, \mathcal{W}_n)$ that

$$\max \left\{ P \left(S'_i \Gamma_{i,n} S_i \geq 0 | \Lambda_i, \mathcal{W}_n \right), P \left(S'_i \Gamma_{i,n} S_i \leq 0 | \Lambda_i, \mathcal{W}_n \right) \right\} = 1. \quad (\text{A.29})$$

Next note that Assumption A.3.1(iii) implies that the distribution of S_i is absolutely continuous with respect to the distribution of S_i conditionally on $(\Lambda_i, \mathcal{W}_n)$. Therefore, letting \mathbb{M}_p denote the set of $p \times p$ real matrices and defining the sets

$$\begin{aligned} O_+ &\equiv \{G \in \mathbb{M}_p : P(S'GS \geq 0) = 1\} \\ O_- &\equiv \{G \in \mathbb{M}_p : P(S'GS \leq 0) = 1\} \end{aligned}$$

we obtain from result (A.29) holding with probability one over $(\Lambda_i, \mathcal{W}_n)$ that we have

$$P(\Gamma_{i,n} \in O_+ \cup O_-) = 1. \quad (\text{A.30})$$

Moreover, since the sign of $a'Ga$ equals the sign of $a'Ga/\|a\|_1^2$ for any matrix $G \in \mathbb{M}^p$ and vector $0 \neq a \in \mathbb{R}^p$, Assumption A.3.1(iv) allows us to conclude that

$$O_+ = \left\{ G \in \mathbb{M}_p : \inf_{a \in \mathbb{R}_+^p} a'Ga \geq 0 \right\} \quad O_- = \left\{ G \in \mathbb{M}_p : \sup_{a \in \mathbb{R}_+^p} a'Ga \leq 0 \right\}. \quad (\text{A.31})$$

In particular, combining (A.30) with (A.31) yields that the distribution of $\Gamma_{i,n}$ satisfies

$$P\left(\inf_{a \in \mathbb{R}_+^p} a' \Gamma_{i,n} a \geq 0 \text{ or } \sup_{a \in \mathbb{R}_+^p} a' \Gamma_{i,n} a \leq 0\right) = 1. \quad (\text{A.32})$$

To conclude, note that by Lemma A.3.2 we have $P(\Lambda_i \geq 0) = 1$ or $P(\Lambda_i \leq 0) = 1$. However, if $P(\Lambda_i \geq 0) = 1$, then $\Gamma_{i,n} = \Lambda_i \text{Var}\{\mathcal{Z}|\mathcal{W}_n\}$ and Assumption A.3.1(i) imply

$$\begin{aligned} P\left(\sup_{a \in \mathbb{R}_+^p} a' \Gamma_{i,n} a \leq 0\right) &= P\left(\sup_{a \in \mathbb{R}_+^p} a' \Gamma_{i,n} a \leq 0; \Lambda_i \geq 0\right) \\ &\leq P(\Lambda_i = 0) \leq P\left(\sup_{a \in \mathbb{R}_+^p} |a' \Gamma_{i,n} a| = 0\right) \leq P\left(\sup_{a \in \mathbb{R}_+^p} a' \Gamma_{i,n} a \leq 0\right), \end{aligned} \quad (\text{A.33})$$

where the final two inequalities hold by set inclusion. By result (A.32) we thus obtain

$$\begin{aligned} 1 &= P\left(\inf_{a \in \mathbb{R}_+^p} a' \Gamma_{i,n} a \geq 0 \text{ or } \sup_{a \in \mathbb{R}_+^p} a' \Gamma_{i,n} a \leq 0\right) \\ &= P\left(\inf_{a \in \mathbb{R}_+^p} a' \Gamma_{i,n} a \geq 0\right) + P\left(\sup_{a \in \mathbb{R}_+^p} a' \Gamma_{i,n} a \leq 0\right) - P\left(\sup_{a \in \mathbb{R}_+^p} |a' \Gamma_{i,n} a| = 0\right) \\ &= P\left(\inf_{a \in \mathbb{R}_+^p} a' \Gamma_{i,n} a \geq 0\right), \end{aligned}$$

where the final equality follows from (A.33). Similarly, it is possible to show that if $P(\Lambda_i \leq 0) = 1$, then $P(\sup_{a \in \mathbb{R}_+^p} a' \Gamma_{i,n} a \leq 0) = 1$, and therefore the lemma follows. ■

Lemma A.3.2. *Let Assumption A.3.1 hold, set $\mathcal{G}_n = \{S_i, W_i, \Lambda_i, \beta_i, \varepsilon_i, \eta_i\}_{i=1}^n$, and*

$$P\left(\min_{1 \leq i \leq n} S'_i(\Lambda_i \text{Var}\{\mathcal{Z}|\mathcal{G}_n\} S_i) \geq 0 \text{ or } \max_{1 \leq i \leq n} S'_i(\Lambda_i \text{Var}\{\mathcal{Z}|\mathcal{G}_n\} S_i) \leq 0\right) = 1 \quad (\text{A.34})$$

for n sufficiently large. Then, it follows that either $P(\Lambda_i \geq 0) = 1$ or $P(\Lambda_i \leq 0) = 1$.

Proof. We will proceed by contradiction and instead suppose that in fact we have that

$$\min\left\{P\left(\max_{1 \leq j \leq p} \lambda_{ij} > 0\right), P\left(\min_{1 \leq j \leq p} \lambda_{ij} < 0\right)\right\} > 0. \quad (\text{A.35})$$

For notational simplicity, let $\mathcal{W}_n = \{W_i\}_{i=1}^n$, set $\Gamma_{i,n} \equiv \Lambda_i \text{Var}\{\mathcal{Z}|\mathcal{W}_n\}$, and define

$$\begin{aligned} A_+(\Gamma_{i,n}) &\equiv \{s \in \mathbb{R}^p : s \geq 0 \text{ and } s' \Gamma_{i,n} s > 0\} \\ A_-(\Gamma_{i,n}) &\equiv \{s \in \mathbb{R}^p : s \geq 0 \text{ and } s' \Gamma_{i,n} s < 0\}. \end{aligned}$$

Next note that if $\lambda_{ij} > 0$ for some $1 \leq j \leq p$ (resp. $\lambda_{ij} < 0$ for some $1 \leq j \leq p$) then $A_+(\Gamma_{i,n})$ (rep. $A_-(\Gamma_{i,n})$) has non-empty interior whenever $\text{Var}\{\mathcal{Z}_j|\mathcal{W}_n\} > 0$ for all $1 \leq$

$j \leq p$. Therefore, since $s \in A_+(\Gamma_{i,n})$ (resp. $A_-(\Gamma_{i,n})$) if and only if $s/\|s\|_1 \in A_+(\Gamma_{i,n})$ (resp. $s/\|s\|_1 \in A_-(\Gamma_{i,n})$), we obtain by Assumptions A.3.1(i)(iii) that

$$\begin{aligned} P\left(S_i \in A_+(\Gamma_{i,n}) \mid \max_{1 \leq j \leq p} \lambda_{ij} > 0, \mathcal{W}_n\right) &> 0 \\ P\left(S_i \in A_-(\Gamma_{i,n}) \mid \min_{1 \leq j \leq p} \lambda_{ij} < 0, \mathcal{W}_n\right) &> 0. \end{aligned} \quad (\text{A.36})$$

Similarly, also note that display (A.35) holding and Assumption A.3.1(iii) together imply

$$\min \left\{ P\left(\max_{1 \leq j \leq p} \lambda_{ij} > 0 \mid \mathcal{W}_n\right), P\left(\min_{1 \leq j \leq p} \lambda_{ij} < 0 \mid \mathcal{W}_n\right) \right\} > 0. \quad (\text{A.37})$$

To conclude, observe that for $n \geq 2$, display (A.3.2) and Assumptions A.3.1(i)(ii) yield

$$\begin{aligned} 0 &\geq P(S'_1 \Lambda_1 \text{Var}\{\mathcal{Z} \mid \mathcal{W}_n\} S_1 > 0 \text{ and } S'_2 \Lambda_2 \text{Var}\{\mathcal{Z} \mid \mathcal{W}_n\} S_2 < 0) \\ &= E \left[P(S'_1 \Lambda_1 \text{Var}\{\mathcal{Z} \mid \mathcal{W}_n\} S_1 > 0 \mid \mathcal{W}_n) P(S'_1 \Lambda_1 \text{Var}\{\mathcal{Z} \mid \mathcal{W}_n\} S_1 < 0 \mid \mathcal{W}_n) \right] \\ &\geq E \left[P\left(S_1 \in A_+(\Gamma_{1,n}); \max_{1 \leq j \leq p} \lambda_{1j} > 0 \mid \mathcal{W}_n\right) P\left(S_1 \in A_-(\Gamma_{1,n}); \min_{1 \leq j \leq p} \lambda_{1j} < 0 \mid \mathcal{W}_n\right) \right] \\ &> 0, \end{aligned}$$

where the final inequality follows from results (A.36) and (A.37). Hence, we have arrived at a contradiction implying (A.35) cannot hold. ■

A.4 Proofs for Section 3.2

In this appendix we collect the technical results behind Proposition 4.1. Section A.4.1 introduces the framework, Section A.4.2 provides an outline of the main argument, and Sections A.4.3 and A.4.4 establish the two key building blocks for the main result.

A.4.1 Preliminaries and Definitions

We consider a probability space $(\Xi \times \Psi, \mathcal{F} \times \mathcal{C}, P_u \times P_f)$ where Ξ and Ψ are Polish spaces with their respective Borel σ -algebras \mathcal{F} and \mathcal{C} ; see, p. 270 in Dudley (1989). An example of such spaces is $\Xi = (\mathbb{R}^\infty \times \dots \times \mathbb{R}^\infty)$; i.e. a finite number of products of \mathbb{R}^∞ . Let $\Omega \equiv \Xi \times \Psi$, $\mathcal{X} \equiv \mathcal{F} \times \mathcal{C}$ and $P \equiv P_u \times P_f$ so that the probability space can be represented more compactly as (Ω, \mathcal{X}, P) . It is useful for later developments to impose further structure on (Ξ, \mathcal{F}, P_u) . We assume that $\{(\Xi_i, \mathcal{F}_i, P_{u,i})\}_{i=1}^\infty$ is a sequence of probability spaces and define $\Xi \equiv \Xi_1 \times \Xi_2 \times \dots$, $\mathcal{F} \equiv \mathcal{F}_1 \times \mathcal{F}_2 \times \dots$, and $P_u \equiv P_{u,1} \times P_{u,2} \times \dots$ such that (Ξ, \mathcal{F}, P_u) is an infinite dimensional product space. Let $\mathcal{F}_t^i \subset \mathcal{F}_i$ denote an

array of filtrations, where $\mathcal{F}_{-\infty}^i = \{\emptyset, \Xi_i\}$, $\mathcal{F}_{\infty}^i = \mathcal{F}_i$, and $\mathcal{F}_t^i \subset \mathcal{F}_{t+1}^i$ for $i \leq n$.¹⁵ Similarly, let $\mathcal{C}_t \subset \mathcal{C}$ denote a triangular array of filtrations, where $\mathcal{C}_{-\infty} = \{\emptyset, \Psi\}$, $\mathcal{C}_{\infty} = \mathcal{C}$, and $\mathcal{C}_t \subset \mathcal{C}_{t+1}$. In addition, define the filtrations $\mathcal{H}_t^i \equiv \mathcal{F}_t^i \times \mathcal{C}_t$. Then, $\mathcal{F} \times \mathcal{C}_t \subset \mathcal{X}$ for all t and \mathcal{H}_t^i can be embedded in \mathcal{X} for each i and all $t \leq T$ by padding up additional coordinates; see, e.g., page 140 in [Halmos \(1976\)](#). In addition, note that $\mathcal{H}_t^i \subset \mathcal{H}_{t+1}^i$ for all i , and t .

In order to simplify our analysis, we will assume that $\mathcal{F}_t^i = \sigma(\dots, \eta_{i,t-1}, \eta_{it})$, where $(\eta_{it})_{i=1, t=-\infty}^{\infty, \infty}$ is an array of *some* random variables on (Ξ, \mathcal{F}, P_u) where for each i , $(\eta_{it})_{t=-\infty}^{\infty}$ is an array of random variables defined on $(\Xi_i, \mathcal{F}_i, P_{u,i})$. The product space structure of (Ξ, \mathcal{F}, P_u) immediately implies independence of $(\eta_{it})_{t=-\infty}^{\infty}$ and $(\eta_{jt})_{t=-\infty}^{\infty}$ for any $i \neq j$. Likewise, we assume that $\mathcal{C}_t = \sigma(\dots, v_{t-1}, v_t)$, where $(v_t)_{t=-\infty}^{\infty}$ is a sequence of some random variables on (Ψ, \mathcal{C}, P_f) .

We will introduce mixing measures as in [McLeish \(1975\)](#); see also [Andrews \(1988\)](#) for triangular array versions of these measures. Recall $\mathcal{C}_t = \sigma(\dots, v_{t-1}, v_t)$ and define $\mathcal{C}_t^{t+m} \equiv \sigma(v_t, v_{t+1}, \dots, v_{t+m})$ and $\mathcal{C}_t^{\infty} \equiv \sigma(v_t, v_{t+1}, \dots)$. Similarly, we let

$$\mathcal{F}_t^{i,t+m} \equiv \sigma(\eta_{it}, \dots, \eta_{i,t+m}) \text{ for all } i \leq n.$$

Our asymptotic normality results employ α -mixing coefficients, and we therefore define

$$\alpha_f(m) \equiv \sup_t \sup_{A \in \mathcal{C}_t, B \in \mathcal{C}_{t+m}^{\infty}} |P(A \cap B) - P(A)P(B)|.$$

We will now impose the following restrictions on the measures P_u and P_f . By Theorems 4.34 and A.46 in [Breiman \(1968\)](#), a regular conditional distribution on \mathcal{X} given $\mathcal{C} \subset \mathcal{X}$ exists, and by Theorem 10.2.2 in [Dudley \(1989\)](#), the regular conditional distribution is unique for P -almost all $\omega \in \Omega$. As in [Eagleson \(1975\)](#), let $\omega' \in \Omega$ and consider the regular conditional (on \mathcal{C}) probability denoted by $Q_{\omega'}(B, \mathcal{C}) = Q_{\omega'}(B)$. It follows that for fixed $B \in \mathcal{X}$, $Q_{\omega'}(B, \mathcal{C})$ is a version of $P(B|\mathcal{C})$ and for fixed $\omega' \in \Omega$, $Q_{\omega'}(\cdot)$ is a probability measure on \mathcal{X} . Importantly, this means $Q_{\omega'}(\cdot)$ is countably additive which ensures that the law of iterated expectations holds; see p. 270 in [Dudley \(1989\)](#). We note that the results in [Dudley \(1989\)](#) are established for Polish spaces.

Consider the measure space $(\Omega, \mathcal{X}, Q_{\omega'})$ with expectation $E_{\omega'}$, which formalizes the idea of treating the aggregate variables \varkappa_t (to be defined later) as constants through the choice of ω' . By a lemma in p. 558 of [Eagleson \(1975\)](#), the following holds:

Lemma A.4.1 ([Eagleson 1975](#)). *Let \mathcal{Y} be a sub-sigma field of \mathcal{X} such that $\mathcal{C} \subseteq \mathcal{Y}$.¹⁶ Then, for P almost all $\omega' \in \Omega$ and a random variable q with $E[|q|] < \infty$ it follows that $E_{\omega'}[q|\mathcal{Y}](\omega) = E[q|\mathcal{Y}](\omega) Q_{\omega'}\text{-a.s.}$*

¹⁵We can allow for triangular arrays, but suppress it for simplicity of notations.

¹⁶It is easy to see that the proof of Lemma 1 in [Eagleson \(1975\)](#) goes through when $\mathcal{F} = \mathcal{G}$.

For arbitrary t and arbitrary $m \geq 0$, we next define the filtration $\mathcal{Y}_t^{i,t+m}$ by

$$\mathcal{Y}_t^{i,t+m} \equiv \mathcal{F}_t^{i,t+m} \times \mathcal{C} \quad (\text{A.38})$$

To establish marginal convergence as $T \rightarrow \infty$ for each $i \leq n$ fixed we follow the strategy of the proof of Theorem 2 in [Eagleson \(1975\)](#). This requires modifying the regularity conditions to the measure $Q_{\omega'}$. Define the conditional L_q norm $\|y\|_{q|\mathcal{C}} = (\int |y|^q dP(y|\mathcal{C}))^{1/q}$. By Lemma [A.4.1](#) it follows that $\|y\|_{q|\mathcal{C}} = (\int |y|^q dQ_{\omega'})^{1/q} Q_{\omega'}$ a.s. for P -almost all ω' . Similarly, define the conditional mixing coefficients

$$\alpha_{\xi|\mathcal{C}}(m) \equiv \sup_t \sup_{A \in \mathcal{Y}_{-\infty}^{i,t}, B \in \mathcal{Y}_{t+m}^{i,\infty}} |P(A \cap B|\mathcal{C}) - P(A|\mathcal{C})P(B|\mathcal{C})|, \quad a.s. \quad (\text{A.39})$$

where $\alpha_{\xi|\mathcal{C}}(m)$ are \mathcal{C} -measurable random variables.

Now, recall that our analysis is predicated on the moment restriction in [\(21\)](#). In order to facilitate our asymptotic analysis, we impose the following:

Condition 1. (i) There are some random vectors f_t and u_{it} defined on (Ψ, \mathcal{C}, P_f) and $(\Xi_i, \mathcal{F}_i, P_{u,i})$ such that $S_i \varepsilon_{it} = \psi(f_t, u_{it})$; (ii) The \mathcal{Z}_t is a component of f_t ; and (iii) The joint distribution of $(u_{it})_{t=1}^{\infty}$ is identical across i .

It is important to note that the product structure $P_u = P_{u,1} \times P_{u,2} \times \dots$ implies that the variables $(u_{it})_{t=1}^{\infty}$ are independent over i . Finally, we define the variables

$$\xi_{it} \equiv (\mathcal{Z}_t)' \nu_{it}, \quad \varkappa_t \equiv (\mathcal{Z}_t)' \zeta_t,$$

where ζ_t and ν_{it} are given by $\zeta_t \equiv E[S_{it} \varepsilon_{it} | \mathcal{C}] = E[S_{it} \varepsilon_{it} | f_t]$ and $\nu_{it} \equiv S_{it} \varepsilon_{it} - \zeta_t$.

A.4.2 Proof of Proposition 3.3 (Outline)

Recall that Proposition 3.3 derives an asymptotic distribution for the numerator (i.e. the score) of $(\hat{\beta}_n - \beta)$; see [\(22\)](#). Given the introduced notation, we can decompose

$$\frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T (\mathcal{Z}_t' S_{it}) \varepsilon_{it} = \frac{1}{T} \sum_{t=1}^T \varkappa_t + \frac{1}{nT} \sum_{t=1}^T \sum_{i=1}^n \xi_{it}.$$

We start the analysis by studying the conditional distribution given \mathcal{C} of the term

$$\frac{1}{\sqrt{nT}} \sum_{t=1}^T \sum_{i=1}^n \xi_{it}. \quad (\text{A.40})$$

By Condition 1, ξ_{it} are independent over i conditional on \mathcal{C} , and we therefore expect that [\(A.40\)](#) is asymptotically normal conditional on \mathcal{C} . Under additional regularity

conditions, the limit distribution has a variance that does not depend on \mathcal{C} . Indeed, this is formally established in Theorem A.4.2 below. In particular, defining

$$\phi_{nT}(\varsigma_1 | \mathcal{C}) \equiv E \left[\exp \left(\iota_{\varsigma_1} \frac{1}{\sqrt{nT}} \sum_{t=1}^T \sum_{i=1}^n \xi_{it} \right) \middle| \mathcal{C} \right], \quad (\text{A.41})$$

we have that $\phi_{nT}(\varsigma_1 | \mathcal{C}) \rightarrow \phi(\varsigma_1)$ almost surely in \mathcal{C} , where $\phi(\varsigma_1)$ denotes the characteristic function of the limiting normal distribution.

Now, let's consider the joint characteristic function of the vector in display (25):

$$E \left[\exp \left(\iota_{\varsigma_2} \frac{1}{\sqrt{T}} \sum_{t=1}^T \varkappa_t + \iota_{\varsigma_1} \frac{1}{\sqrt{nT}} \sum_{t=1}^T \sum_{i=1}^n \xi_{it} \right) \right] = E \left[\exp \left(\iota_{\varsigma_2} \frac{1}{\sqrt{T}} \sum_{t=1}^T \varkappa_t \right) \phi_{nT}(\varsigma_1 | \mathcal{C}) \right]$$

where we note that \varkappa_t is measurable with respect to \mathcal{C} . Because $\sum_{t=1}^T E[\varkappa_t] = 0$ due to the moment restriction in (21), a time series CLT should apply to the term

$$\frac{1}{\sqrt{T}} \sum_{t=1}^T \varkappa_t. \quad (\text{A.42})$$

Indeed, the asymptotic normality of (A.42) is formally established in Theorem A.4.1 in Appendix A.4.3 below. Letting $\varphi(\varsigma_2)$ denote the characteristic function of the limiting normal distribution of the term in (A.42), we then obtain that

$$\begin{aligned} & E \left[\exp \left(\iota_{\varsigma_2} \frac{1}{\sqrt{T}} \sum_{t=1}^T \varkappa_t + \iota_{\varsigma_1} \frac{1}{\sqrt{nT}} \sum_{t=1}^T \sum_{i=1}^n \xi_{it} \right) \right] - \varphi(\varsigma_2) \phi(\varsigma_1) \\ &= E \left[\exp \left(\iota_{\varsigma_2} \frac{1}{\sqrt{T}} \sum_{t=1}^T \varkappa_t \right) \phi_{nT}(\varsigma_1 | \mathcal{C}) \right] - E \left[\exp \left(\iota_{\varsigma_2} \frac{1}{\sqrt{T}} \sum_{t=1}^T \varkappa_t \right) \phi(\varsigma_1) \right] \end{aligned} \quad (\text{A.43})$$

$$+ E \left[\exp \left(\iota_{\varsigma_2} \frac{1}{\sqrt{T}} \sum_{t=1}^T \varkappa_t \right) \phi(\varsigma_1) \right] - \varphi(\varsigma_2) \phi(\varsigma_1). \quad (\text{A.44})$$

Because characteristic functions are bounded by one and $\phi_{nT}(\varsigma_1 | \mathcal{C}) \rightarrow \phi(\varsigma_1)$ almost surely in \mathcal{C} we obtain from the dominated convergence theorem that

$$\begin{aligned} |(\text{A.43})| &\leq E \left[\exp \left(\iota_{\varsigma_2} \frac{1}{\sqrt{T}} \sum_{t=1}^T \varkappa_t \right) |\phi_{nT}(\varsigma_1 | \mathcal{C}) - \phi(\varsigma_1)| \right] \\ &\leq E [|\phi_{nT}(\varsigma_1 | \mathcal{C}) - \phi(\varsigma_1)|] \rightarrow 0. \end{aligned}$$

Similarly, since (A.42) is asymptotically normally distributed and $\varphi(\varsigma_2)$ denotes the

characteristic function of its limiting distribution, we also have that

$$\begin{aligned} |(\text{A.44})| &\leq \left| E \left[\exp \left(\iota_{\varsigma_2} \frac{1}{\sqrt{T}} \sum_{t=1}^T \varkappa_t \right) \right] - \varphi(\varsigma_2) \right| |\phi(\varsigma_1)| \\ &\leq \left| E \left[\exp \left(\iota_{\varsigma_2} \frac{1}{\sqrt{T}} \sum_{t=1}^T \varkappa_t \right) \right] - \varphi(\varsigma_2) \right| \rightarrow 0. \end{aligned}$$

To conclude, we see that (A.40) and (A.42) are jointly asymptotically normally distributed and independent under the conditions laid out in Theorems A.4.1 and A.4.2.

Remark A.4.1. While we have focused on the numerator of $(\hat{\beta}_n - \beta)$, the denominator can be studied under similar conditions. In particular it is possible to show that

$$\frac{1}{nT} \sum_{t=1}^T \sum_{i=1}^n (\mathcal{Z}'_t S_{it}) X_{it} = \frac{1}{nT} \sum_{t=1}^T \sum_{i=1}^n E[(\mathcal{Z}'_t S_{it}) X_{it}] + o_P(1),$$

though, for conciseness, we do not provide the details of the relevant argument. ■

A.4.3 Unconditional Convergence

The main purpose of this section is to establish a time series central limit theorem for

$$\frac{1}{\sqrt{T}} \sum_{t=1}^T \varkappa_t,$$

which is formally established in Theorem A.4.1 below. For this purpose, it is convenient to assume that \varkappa_t is L_2 near-epoch-dependent (NED). The concept of NED sequences was introduced by Billingsley (1968). McLeish (1975) or Andrews (1988) show that a NED process is also a mixingale. Therefore assumptions that impose NED type conditions lead to strong laws by showing that these processes also satisfy the requirements for strong laws of related mixingales; see Theorem 3 in McLeish (1975) for the first result of this nature.

We impose the following condition that establishes the NED property for \varkappa_t , and imposes sufficient conditions for \varkappa_t to satisfy the conditions for the SLLN in de Jong (1996) and the CLT in de Jong (1997). In the statement below, we say that a sequence δ_m is of size $-\lambda$ if $\delta_m = O(m^{-\lambda-\omega})$ for some $\omega > 0$; see p. 335 in de Jong (1997).

Condition 2. For $r > 2$, assume that $\sup_t \|\varkappa_t\|_r < \infty$, and that there exists a bounded array of non-random constants c_t such that

$$\left\| \varkappa_t - E[\varkappa_t | \mathcal{C}_{t-m}^{t+m}] \right\|_2 \leq c_t \beta_f(m) \text{ for all } t$$

where $\beta_f(m)$ is of size $-1/2$. Further assume that for $r > 2$, $\alpha_f(m)$ is of size $r/(r-2)$.

To establish the CLT we adopt the conditions given in [de Jong \(1997\)](#). The result is based on a blocking scheme that needs to be defined. Let b_T and l_T be positive, non-decreasing integer valued sequences that are the lengths of included and discarded blocks. Assume $b_T \geq l_T + 1$, $l_T \rightarrow \infty$, $l_T \geq 1$, $b_T \leq T$, $b_T/T \rightarrow 0$ and $l_T/b_T \rightarrow 0$. Let $r_T \equiv \lceil T/b_T \rceil$ and define $\tilde{\varkappa}_{T,t} \equiv \varkappa_t/\sigma_{T,\varkappa}$, where where $\sigma_{T,\varkappa}^2$ is given by

$$\sigma_{T,\varkappa}^2 \equiv E \left[\left(\sum_{t=1}^T \varkappa_t \right)^2 \right].$$

The following condition is sufficient for obtaining the desired CLT.

Condition 3. Assume that $\{\varkappa_t, \mathcal{C}_t\}$ satisfies [Condition 2](#). There exists a positive constant array $e_{T,t}$ such that $c_t/e_{T,t}$ is uniformly bounded in $t \leq T$ and $T \geq 1$ and $\{\tilde{\varkappa}_{T,t}/e_{T,t}\}$ is L_r bounded for $r > 2$ uniformly in $t \leq T$ and $T \geq 1$. Let $M_{T,j} \equiv \max_{(j-1)b_T+1 \leq t \leq jb_T} e_{T,t}$ for $1 \leq j \leq r_T$ and $M_{T,r_T+1} \equiv \max_{r_T b_T+1 \leq t \leq T} e_{T,t}$. Then, $\max_{1 \leq j \leq r_T+1} M_{T,j} = o(b_T^{-1/2})$ and $\sum_{j=1}^{r_T} M_{T,j}^2 = O(b_T^{-1})$.

Theorem A.4.1. Assume that [Condition 3](#) holds and suppose that there is some non-random constant σ_\varkappa^2 such that $\lim_{T \rightarrow \infty} \sigma_{T,\varkappa}^2/T = \sigma_\varkappa^2$. It then follows that

$$\frac{1}{\sqrt{T}} \sum_{t=1}^T \varkappa_t \xrightarrow{d} N(0, \sigma_\varkappa^2).$$

Proof. Under [Condition 3](#), we can use [Theorem 2](#) in [de Jong \(1997\)](#) to conclude that $\sum_{t=1}^T \tilde{\varkappa}_{T,t} \xrightarrow{d} N(0, 1)$. Under the additional assumption that $T^{-1}\sigma_{T,\varkappa}^2 \rightarrow \sigma_\varkappa^2$ for some non-random constant σ_\varkappa^2 , it follows from the continuous mapping theorem that

$$\frac{1}{\sqrt{T}} \sum_{t=1}^T \varkappa_t \xrightarrow{d} N(0, \sigma_\varkappa^2), \tag{A.45}$$

which establishes [Theorem A.4.1](#). ■

A.4.4 Conditional Analysis

The primary purpose of this section is to establish the asymptotic normality of the term

$$\frac{1}{\sqrt{nT}} \sum_{t=1}^T \sum_{i=1}^n \xi_{it}$$

conditional on \mathcal{C} , which is formally shown in [Theorem A.4.2](#). The proof uses ideas similar to the development in [Hahn et al. \(2019\)](#), which in turn relies on arguments in [Eagleson \(1975\)](#), to handle the conditioning step in the proof of the CLT; see also

Lemma 2.9.5 in [van der Vaart and Wellner \(1996\)](#). However, the dependence structure of the panel is more complicated here than in [Hahn et al. \(2019\)](#) and requires a different approach to prove the CLT. [Hahn et al. \(2019\)](#) consider a scenario where conditional on \mathcal{C} , a cross-sectional average over i.i.d. draws is analyzed. Here, we need to extend these results to a panel setting with joint asymptotics as $N, T \rightarrow \infty$ and where we allow for general dependence and possible non-stationarity in the time series direction. We extend the notation from [Hahn et al. \(2019\)](#) to account for these extensions. The result of [Eagleson \(1975\)](#) formalizes the intuition that conditional on \mathcal{C} the processes \mathcal{Z}_t and f_t can be treated as a fixed constant in deriving limiting results.

We assume that $\alpha_{\xi|\mathcal{C}}(m) \rightarrow 0$ almost surely as $m \rightarrow \infty$. This is done in the next condition, which introduces conditional NED, by requiring $\alpha_{\xi|\mathcal{C}}$ be of size $r/(r-2)$.

Condition 4. For $r > 2$, assume that $\sup \|\xi_{it}\|_{r|\mathcal{C}} < \infty$ a.s., and that there exists an array of \mathcal{C} -measurable random variables $d_{it}(\omega)$ and a sequence of \mathcal{C} -measurable random variables $\beta_{\xi|\mathcal{C}}(m)$ of size $-1/2$ a.s. such that the following inequality holds:

$$\left\| \xi_{it} - E \left[\xi_{it} | \mathcal{Y}_{t-m}^{i,t+m} \right] \right\|_{2|\mathcal{C}} \leq d_{it}(\omega) \beta_{\xi|\mathcal{C}}(m).$$

Further assume that for $r > 2$, $\alpha_{\xi|\mathcal{C}}(m)$ is of size $r/(r-2)$ a.s.

We now turn to establishing our main result. The argument is based on the one-to-one mapping h_T of the double index (i, t) into a single index $s = h_T(i, t) \equiv (i-1)T + t$ for $i \leq n$ and $L \equiv nT$. We will let $(\bar{i}_T(s), \bar{t}_T(s))$ denote the (i, t) that corresponds to s , i.e., $(\bar{i}_T(s), \bar{t}_T(s)) \equiv h_T^{-1}(s)$.¹⁷ With some abuse of notation, we then set

$$(nT)^{-1/2} \sum_{i=1}^n \sum_{t=1}^T \xi_{it} = L^{-1/2} \sum_{s=1}^L \xi_{L,s}.$$

The sum on the right hand side sums over components starting with $i = 1, t = 1, \dots, T$ followed by $i = 2, t = 1, \dots, T$, and so on. Similarly, we construct the array of filtrations based on (A.38), using coordinate identification rules of [Halmos \(1976\)](#) (see p. 151 and p. 155) to organize coordinates, by defining

$$\begin{aligned} \mathcal{K}_{L,s} &\equiv \left(\bigtimes_{j=1}^{\bar{i}_T(s)-1} \mathcal{F}_{-\infty}^{j,\infty} \right) \times \mathcal{F}_{-\infty}^{\bar{i}_T(s), \bar{t}_T(s)} \times \left(\bigtimes_{j=\bar{t}_T(s)+1}^{\infty} \mathcal{F}_{-\infty}^j \right) \times \mathcal{C} \quad \text{if } \bar{t}_T(s) < T, \\ \mathcal{K}_{L,s} &\equiv \left(\bigtimes_{j=1}^{\bar{i}_T(s)} \mathcal{F}_{-\infty}^{j,\infty} \right) \times \left(\bigtimes_{j=\bar{t}_T(s)+1}^{\infty} \mathcal{F}_{-\infty}^j \right) \times \mathcal{C} \quad \text{if } \bar{t}_T(s) = T. \end{aligned}$$

The construction guarantees that $\mathcal{K}_{L,s} \subset \mathcal{X}$ by padding missing coordinates with the trivial field $\mathcal{F}_{-\infty}^j = \{\emptyset, \Xi_j\}$.

¹⁷Note $\bar{i}_T(s)$ is the smallest integer larger than or equal to s/T , and $\bar{t}_T(s) = s - (\bar{i}_T(s) - 1) \cdot T$.

We introduce the binary operator \ominus that maps two σ -fields generated by a sequence of random variables into a σ -field generated by the non-overlapping portion of members of the sequence. In particular, for $0 < s < s'$ and $\mathcal{F}_{-\infty}^{j,s'}$, $\mathcal{F}_{-\infty}^{j,s}$ we define $\mathcal{F}_{-\infty}^{j,s'} \ominus \mathcal{F}_{-\infty}^{j,s} \equiv \mathcal{F}_{-\infty}^{j,s'}$, where $\mathcal{F}_t^{i,t+m} = \sigma(\eta_{it}, \dots, \eta_{i,t+m})$. We further define \ominus to have the properties $\left(\prod_{j=1}^{\infty} \mathcal{A}_{nT}^j\right) \ominus \left(\prod_{j=1}^{\infty} \mathcal{B}_{nT}^j\right) \equiv \prod_{j=1}^{\infty} \left(\mathcal{A}_{nT}^j \ominus \mathcal{B}_{nT}^j\right)$ for $\mathcal{A}_{nT}^j, \mathcal{B}_{nT}^j \subset \mathcal{F}_{-\infty}^{j,\infty}$, where we understand $\mathcal{F}_{-\infty}^{j,\infty} \ominus \mathcal{F}_{-\infty}^{j,\infty} \equiv \mathcal{F}_{-\infty}^j$, $\mathcal{F}_{-\infty}^j \ominus \mathcal{F}_{-\infty}^j \equiv \mathcal{F}_{-\infty}^j$ and $\mathcal{C} \ominus \mathcal{C} \equiv \mathcal{C}$. Using these properties we define the sigma fields $\mathcal{K}_{L,s}^{s'}$ by $\mathcal{K}_{L,s}^{s'} \equiv \mathcal{K}_{L,s'} \ominus \mathcal{K}_{L,s}$.

Because $\mathcal{K}_{L,s}$ contains the coordinate \mathcal{C} for all L and s the construction implies that when conditioning on $\mathcal{K}_{L,s}$, all \varkappa_t and ζ_t are held fixed. We have the following Lemma relating mixing coefficients for $\{\xi_{L,s}, \mathcal{K}_{L,s}\}$ to mixing coefficients for $\{\xi_{it}, \mathcal{Y}_{-\infty}^{i,t}\}$.¹⁸

Lemma A.4.2. *Assume that $\{\xi_{it}, \mathcal{Y}_{-\infty}^{i,t}\}$ satisfies Condition 4 for each i . Assume that Condition 1 holds. Then, for $r > 2$, $\{\xi_{L,s}, \mathcal{K}_{L,s}\}$ satisfies $\sup_{L,s \leq L} \|\xi_{L,s}\|_{r|\mathcal{C}} < \infty$ a.s., and there exists an array of \mathcal{C} -measurable, almost surely bounded, random variables $d_{L,s}(\omega)$ and a sequence of \mathcal{C} -measurable random variables $\beta_{\xi|\mathcal{C}}(m)$ such that*

$$\left\| \xi_{L,s} - E \left[\xi_{L,s} | \mathcal{K}_{L,s-m}^{s+m} \right] \right\|_{2|\mathcal{C}} \leq d_{L,s}(\omega) \beta_{\xi|\mathcal{C}}(m) \text{ for each } s \leq L \quad (\text{A.46})$$

where $\beta_{\xi|\mathcal{C}}(m)$ is of size $-1/2$ a.s.. Furthermore, for $r > 2$, and

$$\alpha_{\xi|\mathcal{C}}(m) = \sup_t \sup_{A \in \mathcal{K}_{L,-\infty}^t, B \in \mathcal{K}_{L,t+m}^{\infty}} |P(A \cap B|\mathcal{C}) - P(A|\mathcal{C})P(B|\mathcal{C})| \quad (\text{A.47})$$

it follows that $\alpha_{\xi|\mathcal{C}}(m)$ is of size $r/(r-2)$ a.s.

Proof. The first claim follows immediately from $\xi_{L,s} = \xi_{it}$, where the $(i, t) = h_T^{-1}(s)$, and by applying Condition 4 to ξ_{it} . To establish (A.46), we first consider the case where $m > \max(T - \bar{t}_T(s), \bar{t}_T(s))$. The way the filtration $\mathcal{K}_{L,s-m}^{s+m}$ is constructed, it follows that $\xi_{L,s}$ is measurable with respect to $\mathcal{K}_{L,s-m}^{s+m}$. This then leads to

$$\left\| \xi_{L,s} - E \left[\xi_{L,s} | \mathcal{K}_{L,s-m}^{s+m} \right] \right\|_{2|\mathcal{C}} = 0.$$

Now, consider the case where $0 \leq m \leq \max(T - \bar{t}_T(s), \bar{t}_T(s))$. We distinguish two cases (a) $T - \bar{t}_T(s) \geq \bar{t}_T(s)$, and (b) $T - \bar{t}_T(s) < \bar{t}_T(s)$. For Case (a), if $0 \leq m < \bar{t}_T(s)$, then¹⁹

$$E \left[\xi_{L,s} | \mathcal{K}_{L,s-m}^{s+m} \right] = E \left[\xi_{it} | \mathcal{F}_{t-m}^{i,t+m} \vee \mathcal{C} \right],$$

and if $\bar{t}_T(s) \leq m < T - \bar{t}_T(s)$, then by independence

$$E \left[\xi_{L,s} | \mathcal{K}_{L,s-m}^{s+m} \right] = E \left[\xi_{it} | \mathcal{F}_{-\infty}^{i,t+m} \vee \mathcal{C} \right].$$

¹⁸Condition 4 includes uniformity of the mixing and approximation coefficients, which is used here.

¹⁹For simplicity of notations, we adopt the convention that $(i, t) = (\bar{t}_T(s), \bar{t}_T(s))$.

In the latter case, we have

$$\begin{aligned} \left\| \xi_{L,s} - E \left[\xi_{L,s} | \mathcal{K}_{L,s-m}^{s+m} \right] \right\|_{2|\mathcal{C}} &= \left\| \xi_{it} - E \left[\xi_{it} | \mathcal{F}_{nT,-\infty}^{i,t+m} \vee \mathcal{C} \right] \right\|_{2|\mathcal{C}} \\ &\leq \left\| \xi_{it} - E \left[\xi_{it} | \mathcal{F}_{t-m}^{i,t+m} \vee \mathcal{C} \right] \right\|_{2|\mathcal{C}}, \end{aligned}$$

because the residual in the projection on $\mathcal{F}_{t-m}^{i,t+m} \vee \mathcal{C} \subset \mathcal{F}_{-\infty}^{i,t+m} \vee \mathcal{C}$ has smaller variance, whereas in the former case we have

$$\left\| \xi_{L,s} - E \left[\xi_{L,s} | \mathcal{K}_{L,s-m}^{s+m} \right] \right\|_{2|\mathcal{C}} = \left\| \xi_{it} - E \left[\xi_{it} | \mathcal{F}_{t-m}^{i,t+m} \vee \mathcal{C} \right] \right\|_{2|\mathcal{C}}.$$

Therefore, in both cases, we have

$$\left\| \xi_{L,s} - E \left[\xi_{L,s} | \mathcal{K}_{n,s-m}^{s+m} \right] \right\|_{2|\mathcal{C}} \leq d_{nT,it}(\omega) \beta_{\xi|\mathcal{C}}(m)$$

by Condition 4. Relabeling the term $d_{nT,it}(\omega) = d_{L,s}(\omega)$ then yields the claimed inequality in (A.46). For Case (b), if $0 \leq m \leq T - \bar{t}_T(s)$, then

$$E \left[\xi_{L,s} | \mathcal{K}_{L,s-m}^{s+m} \right] = E \left[\xi_{it} | \mathcal{F}_{t-m}^{i,t+m} \vee \mathcal{C} \right],$$

and if $T - \bar{t}_T(s) < m \leq \bar{t}_T(s)$, then

$$E \left[\xi_{L,s} | \mathcal{K}_{L,s-m}^{s+m} \right] = E \left[\xi_{it} | \mathcal{F}_{t-m}^{i,\infty} \vee \mathcal{C} \right].$$

In both cases, we have by similar reasoning

$$\left\| \xi_{L,s} - E \left[\xi_{L,s} | \mathcal{K}_{L,s-m}^{s+m} \right] \right\|_{2|\mathcal{C}} \leq d_{nT,it}(\omega) \beta_{\xi|\mathcal{C}}(m).$$

Finally, to establish (A.47) note that if $m > T - \bar{t}_T(s)$, it follows for $A \in \mathcal{K}_{L,-\infty}^s$ and $B \in \mathcal{K}_{L,s+m}^\infty$ that

$$|P(A \cap B | \mathcal{C}) - P(A | \mathcal{C}) P(B | \mathcal{C})| = 0 \text{ a.s.} \quad (\text{A.48})$$

because in that case A and B are independent by Condition 1. When $m \leq T - \bar{t}_T(s)$, which implies that $\bar{t}_T(s) = \bar{t}_T(s+m)$, or in other words that conditioning happens within the same cross-sectional unit, we consider the following. The product space structure of $\mathcal{K}_{L,-\infty}^s$ implies that $A = A_1 \times A_2 \times A_3 \times A_4$, where²⁰ $A_1 \in \times_{j=1}^{\bar{t}_T(s)-1} \mathcal{F}_{-\infty}^j$, $A_2 \in \mathcal{F}_{-\infty}^{\bar{t}_T(s), \bar{t}_T(s)}$, $A_3 \in \times_{j=\bar{t}_T(s)+1}^{\infty} \mathcal{F}_{nT,-\infty}^j$, and $A_4 \in \mathcal{C}$. In the same way, $B = B_1 \times B_2 \times B_3 \times B_4$, where $B_1 \in \times_{j=1}^{\bar{t}_T(s+m)-1} \mathcal{F}_{-\infty}^j$, $B_2 \in \mathcal{F}_{\bar{t}_T(s+m), \infty}^{\bar{t}_T(s+m)}$, $B_3 \in \times_{j=\bar{t}_T(s+m)+1}^{\infty} \mathcal{F}_{-\infty}^j$, $B_4 \in \mathcal{C}$. Note that

$$A_2 \times A_4 \in \mathcal{Y}_{-\infty}^{\bar{t}_T(s), \bar{t}_T(s)} = \mathcal{F}_{-\infty}^{\bar{t}_T(s), \bar{t}_T(s)} \times \mathcal{C} \quad (\text{A.49})$$

²⁰See (Halmos, 1976, p.154 and theorem 38.A).

and

$$B_2 \times B_4 \in \mathcal{Y}_{\bar{t}_T(s+m), \infty}^{\bar{t}_T(s+m), \infty} = \mathcal{F}_{\bar{t}_T(s+m), \infty}^{\bar{t}_T(s+m), \infty} \times \mathcal{C}. \quad (\text{A.50})$$

Since $A \cap B = \times_{j=1}^4 (A_j \cap B_j)$ by p. 139 in [Halmos \(1976\)](#), and

$$P(A \cap B | \mathcal{C}) = \left(\prod_{j=1}^3 \tilde{P}_j(A_j \cap B_j) \right) P_f(A_4 \cap B_4 | \mathcal{C})$$

where \tilde{P}_j are themselves products of measures $P_{u,j}$ corresponding to the respective coordinates in A_j and B_j and where A_j, B_j for $j \leq 3$ are independent of \mathcal{C} . We note that A_1 and B_3 consist of coordinates that are either Ξ_j or \emptyset . By Theorem 33.A in [Halmos \(1976\)](#), $B = \emptyset$ if and only if at least one coordinate is equal to \emptyset . Thus, by the properties of conditional expectations, $P(A \cap B | \mathcal{C}) = 0$ if one of the coordinates of B or A is equal to \emptyset . Thus, the mixing coefficient is zero in this case. We therefore assume without loss of generality that for A_1 and B_3 all coordinates are equal to Ξ_j . This implies that $A_1 \cap B_1 = B_1$ and $A_3 \cap B_3 = A_3$, and we have $\tilde{P}_1(A_1 \cap B_1) = \tilde{P}_1(B_1)$ and $\tilde{P}_1(A_3 \cap B_3) = \tilde{P}_3(A_3)$. Note that $P_f(A_4 \cap B_4 | \mathcal{C}) = E[1_{A_4} 1_{B_4} | \mathcal{C}] = 1_{A_4} 1_{B_4}$ since A_4 and B_4 are measurable with respect to \mathcal{C} . Finally, note that $\tilde{P}_2(A_2 \cap B_2) = P_{u, \bar{t}_T(s)}(A_2 \cap B_2)$. (Recall that $\bar{t}_T(s) = \bar{t}_T(s+m)$.) These arguments lead to

$$P(A \cap B | \mathcal{C}) = \tilde{P}_1(B_1) \tilde{P}_3(A_3) P_{u, \bar{t}_T(s)}(A_2 \cap B_2) 1_{A_4} 1_{B_4}. \quad (\text{A.51})$$

Similar arguments also show that

$$P(A | \mathcal{C}) = \tilde{P}_1(A_1) P_{u, \bar{t}_T(s)}(A_2) 1_{A_4} \quad (\text{A.52})$$

and

$$P(B | \mathcal{C}) = \tilde{P}_3(B_3) P_{u, \bar{t}_T(s)}(B_2) 1_{B_4} \quad (\text{A.53})$$

such that

$$\begin{aligned} & |P(A \cap B | \mathcal{C}) - P(A | \mathcal{C}) P(B | \mathcal{C})| \quad (\text{A.54}) \\ &= \left| P_{u, \bar{t}_T(s)}(A_2 \cap B_2) - P_{u, \bar{t}_T(s)}(A_2) P_{u, \bar{t}_T(s)}(B_2) \right| \tilde{P}_1(B_1) \tilde{P}_3(A_3) 1_{A_4} 1_{B_4} \\ &\leq \left| P_{u, \bar{t}_T(s)}(A_2 \cap B_2) - P_{u, \bar{t}_T(s)}(A_2) P_{u, \bar{t}_T(s)}(B_2) \right| 1_{A_4} 1_{B_4} \\ &\leq \sup_{A \in \mathcal{Y}_{-\infty}^{\bar{t}_T(s), \bar{t}_T(s)}, B \in \mathcal{Y}_{\bar{t}_T(s+m), \infty}^{\bar{t}_T(s), \infty}} |P(A \cap B | \mathcal{C}) - P(A | \mathcal{C}) P(B | \mathcal{C})| \\ &\leq \alpha_{\xi | \mathcal{C}}(m), \end{aligned}$$

where the equality combines results [\(A.51\)](#), [\(A.52\)](#) and [\(A.53\)](#), the first inequality uses

$\tilde{P}_1(B_1) \tilde{P}_3(A_3) \leq 1$, the second inequality uses

$$\begin{aligned} P_{u, \bar{v}_T(s)}(A_2 \cap B_2) 1_{A_4} 1_{B_4} &= P_{u, \bar{v}_T(s)}(A_2 \cap B_2) P_f(A_4 \cap B_4 | \mathcal{C}) \\ &= P_{u, \bar{v}_T(s)}((A_2 \times A_4) \cap (B_2 \times B_4) | \mathcal{C}) \end{aligned}$$

as well as (A.49) and (A.50). The last inequality in (A.54) uses (A.39). Then, combining (A.48) and (A.54) show that (A.47) holds. ■

Lemma A.4.2 is the basis for establishing a joint CLT as both $n, T \rightarrow \infty$. For this, the earlier blocking definitions need to be adjusted. Let $L \equiv nT$, and $n, T \rightarrow \infty$. Define b_L and l_L to be positive, non-decreasing integer valued sequences that are the length of included and discarded blocks. Assume $b_L \geq l_L + 1$, $l_L \rightarrow \infty$, $l_L \geq 1$, $b_L \leq n$, $b_L/L \rightarrow 0$ and $l_L/b_L \rightarrow 0$. Let $r_L \equiv \lfloor L/b_L \rfloor$ and define $\tilde{\xi}_{L,s} \equiv \xi_{L,s}/\sigma_{L,\xi}(\omega)$,²¹ where

$$\sigma_{L,\xi}^2 = E \left[\left(\sum_{s=1}^L \xi_{L,s} \right)^2 \middle| \mathcal{C} \right]. \quad (\text{A.55})$$

We impose the following conditions that are modifications of the conditions for the case of marginal convergence when $T \rightarrow \infty$.

Condition 5. Assume that $\{\xi_{it}, \mathcal{Y}_{-\infty}^{i,t}\}$ satisfies Condition 4 for each $i \leq n$ and all $T, n \geq 1$. For $\{\xi_{L,s}, \mathcal{K}_{L,s}\}$ impose the following additional restrictions. There exists a positive array of random variables $g_{L,s}(\omega)$ such that $d_{L,s}(\omega)/g_{L,s}(\omega)$ is uniformly bounded P -a.s. for all $s \leq L$ and all $L \geq 1$ and

$$\left\| \tilde{\xi}_{L,s}/g_{L,s}(\omega) \right\|_{r|\mathcal{C}} < \infty \quad (\text{A.56})$$

for $r > 2$ for all $s \leq L$ and all $L \geq 1$. Let $M_{L,j} \equiv \max_{(j-1)b_L+1 \leq s \leq jb_L} g_{L,s}(\omega)$ for $1 \leq j \leq r_L$, and $M_{L,r_L+1} \equiv \max_{r_L b_L+1 \leq s \leq L} g_{L,s}(\omega)$. Then,

$$b_L^{1/2} \left(\max_{1 \leq j \leq r_L+1} M_{L,j} \right) \rightarrow 0 \text{ } P\text{-a.s.}$$

and

$$\left| b_L \left(\sum_{j=1}^{r_L} M_{L,j}^2 \right) \right| \leq C_M < \infty \text{ } P\text{-a.s.}$$

for some constant C_M and all L large enough.

We now state the following conditional CLT for joint convergence.

²¹Note that $\sigma_{L,\xi}$ in (A.55) is a random variable. In order to emphasize it, we write $\tilde{\xi}_{L,s} \equiv \xi_{L,s}/\sigma_{L,\xi}(\omega)$.

Theorem A.4.2. *Assume that Conditions 1 and 5 hold. Then, as $n, T \rightarrow \infty$,*

$$\lim_{T, n \rightarrow \infty} E \left[e^{\iota \tau L^{-1/2} \sum_{s=1}^L \tilde{\xi}_{L,s}} \middle| \mathcal{C} \right] = e^{-\frac{1}{2} \tau^2} \text{ a.s.}$$

If in addition, $L^{-1} \sigma_{L,\xi}^2(\omega) \rightarrow \sigma_\xi^2$ P -a.s. where σ_ξ^2 is a constant, then it follows that

$$\lim_{T, n \rightarrow \infty} E \left[e^{\iota \tau L^{-1/2} \sum_{s=1}^L \xi_{L,s}} \middle| \mathcal{C} \right] = e^{-\frac{1}{2} \tau^2 \sigma_\xi^2} \text{ a.s.}$$

Proof. The proof follows the proof of Theorem 2 in [Eagleson \(1975\)](#). Lemma [A.4.2](#) shows that $\{\xi_{L,s}, \mathcal{K}_{L,s}\}$ is a NED process with coefficients $\beta_{\xi|\mathcal{C}}(m)$ of size $-1/2$ and is based on a mixing sequence with coefficients $\alpha_{\xi|\mathcal{C}}(m)$ of size $-1/2$, where $\beta_{\xi|\mathcal{C}}(m, \omega)$ and $\alpha_{\xi|\mathcal{C}}(m, \omega)$ are \mathcal{C} -measurable random variables and where we add the ω argument to emphasize this fact. By Lemma [A.4.1](#), the conditional expectations on (Ω, \mathcal{X}, P) and $(\Omega, \mathcal{X}, Q_{\omega'})$ coincide with P -probability one. Therefore, for all $\omega' \in \Omega$, except for a set of P -measure zero, $\{\xi_{it}, \mathcal{Y}_{-\infty}^{i,t}\}$ satisfies the same NED process conditions with the same mixing coefficients Q_ω -a.s. The quantities $\beta_{\xi|\mathcal{C}}(m, \omega')$ and $\alpha_{\xi|\mathcal{C}}(m, \omega')$ are constants for the probability space $(\Omega, \mathcal{X}, Q_{\omega'})$. Therefore, the remaining assumptions in Condition 5 guarantee that for all $\omega' \in \Omega$, except for a set of P -measure zero, $\{\xi_{n,s}, \mathcal{K}_{L,s}\}$ satisfies the conditions of Theorem 2 in [de Jong \(1997\)](#) for the probability space $(\Omega, \mathcal{X}, Q_{\omega'})$, except on a set of $Q_{\omega'}$ -probability of zero. Then, for P -almost all ω'

$$\lim_{T, n \rightarrow \infty} E \left[e^{\iota \tau L^{-1/2} \sum_{s=1}^L \tilde{\xi}_{L,s}} \middle| \mathcal{C} \right] = \lim_{T, n \rightarrow \infty} E_{\omega'} \left[e^{\iota \tau L^{-1/2} \sum_{s=1}^L \tilde{\xi}_{L,s}} \right] = e^{-\frac{1}{2} \tau^2} \text{ a.s.} \quad (\text{A.57})$$

where the last equality follows from applying Theorem 2 in [de Jong \(1997\)](#) on $(\Omega, \mathcal{X}, Q_{\omega'})$ to a set of $Q_{\omega'}$ probability one, which establishes the first result.

For the second statement, define the $(\Omega, \mathcal{X}, Q_{\omega'})$ -constant random variable

$$\sigma_{L,\xi}^2(\omega') = E_{\omega'} \left[\left(\sum_{s=1}^L \xi_{L,s} \right)^2 \right].$$

By Lemma [A.4.1](#) $\sigma_{L,\xi}^2(\omega')$ is identical to $\sigma_{L,\xi}^2$ defined in [\(A.55\)](#) except on a set of $Q_{\omega'}$ probability zero. Then, it follows that $L^{-1} \sigma_{L,\xi}^2(\omega') \rightarrow \sigma_\xi^2$ with $Q_{\omega'}$ -probability one on $(\Omega, \mathcal{X}, Q_{\omega'})$ because of the maintained assumption of the theorem that $L^{-1} \sigma_{L,\xi}^2(\omega) \rightarrow \sigma_\xi^2$ P -a.s. The result follows from applying the continuous mapping theorem to [\(A.57\)](#). ■

A.5 Proofs for Section 4.1

In this appendix we provide a formal justification of the inference procedure proposed in Section 4.1 by relying on the high dimensional central limit theorem of [Chernozhuikov et al. \(2022\)](#). To this end, we impose the following assumptions.

Assumption A.5.1. There are $\{\psi_i\}_{i=1}^{b_n}$ that are independent conditionally on \mathcal{G}_n and satisfy $\psi_i \in \mathbb{R}^q$, $E[\psi_i|\mathcal{G}_n] = 0$ for all $1 \leq i \leq b_n$, and for any $\varrho > 0$

$$P \left(\left\| \sum_{i=1}^n f(V_i, \hat{\theta}_n) - \sum_{i=1}^{b_n} \psi_i \right\|_{\infty} > \frac{\varrho \sqrt{b_n}}{\sqrt{\log(q)}} \middle| \mathcal{G}_n \right) \leq r_{1n}(\varrho, \mathcal{G}_n).$$

Assumption A.5.2. For some $B_n(\mathcal{G}_n)$ satisfying $1 \leq B(\mathcal{G}_n) < \infty$ with probability one

$$\max_{1 \leq i \leq b_n} \max_{1 \leq k \leq q} E \left[\exp \left(\frac{|\psi_{ki}|}{B_n(\mathcal{G}_n)} \right) \middle| \mathcal{G}_n \right] \leq 2.$$

Assumption A.5.3. There are $\underline{\sigma} > 0$ and $\bar{\sigma} < \infty$ not depending on \mathcal{G}_n satisfying

$$P \left(\underline{\sigma}^2 \leq \min_{1 \leq k \leq q} \frac{1}{b_n} \sum_{i=1}^{b_n} E[\psi_{ki}^2 | \mathcal{G}_n], \max_{1 \leq k \leq q} \frac{1}{b_n} \sum_{i=1}^{b_n} E[\psi_{ki}^4 | \mathcal{G}_n] \leq \bar{\sigma}^2 B_n(\mathcal{G}_n) \middle| \mathcal{G}_n \right) = 1.$$

Assumption A.5.4. For $\mathcal{D}_n \equiv (\{Y_i, X_i, S_i\}_{i=1}^n, \mathcal{Z}, \mathcal{G}_n)$ we have for any $\varrho > 0$ that

$$P \left(\left\| \sum_{i=1}^{b_n} \omega_i \left((\hat{\psi}_i - \psi_i) - \frac{1}{b_n} \sum_{j=1}^{b_n} (\hat{\psi}_j - \psi_j) \right) \right\|_{\infty} > \frac{\varrho \sqrt{b_n}}{\sqrt{\log(q)}} \middle| \mathcal{D}_n \right) \leq r_{2n}(\varrho, \mathcal{D}_n).$$

Assumption A.5.5. The weights $\{\omega_i\}_{i=1}^{b_n}$ are independent of $\{\mathcal{D}_n\}$ and either: (i) Follow a multinomial distribution with parameter b_n and probabilities $(1/b_n, \dots, 1/b_n)$, or (ii) Are i.i.d. with $\omega_i = \omega_{i,1} + \omega_{i,2}$ with $\omega_{i,1} \sim N(0, \sigma_{\omega}^2)$ for some $\sigma_{\omega}^2 \geq 0$ and $|\omega_{i,2}| \leq 3$.

Assumption A.5.1 demands that the vector of moments be asymptotically equivalent to a q -dimensional sample mean of random variables $\{\psi_i\}_{i=1}^{b_n}$. We note that Assumption A.5.1 effectively requires that the null hypothesis be true by requiring that the variables $\{\psi_i\}_{i=1}^{b_n}$ have mean zero. In turn, Assumptions A.5.2 and A.5.3 imposes moment restrictions on the variables $\{\psi_i\}_{i=1}^{b_n}$ that ensure that the high dimensional central limit theorem of Chernozhukov et al. (2022) is applicable. Finally, Assumption A.5.4 demands a linearization requirement on our bootstrap statistic, while Assumption A.5.5 states requirements on the weights $\{\omega_i\}_{i=1}^{b_n}$ that we may employ. We note, in particular, that Assumption A.5.5 allows for the empirical bootstrap (through Assumption A.5.5(i)) and the use of Standard Gaussian, Rademacher, or Mammen (1993) weights (through Assumption A.5.5(ii)).

Our next result encompasses Proposition 4.1 as a special case. The first and second parts of the result provide conditions under which the level of the test is $1 - \alpha$ unconditionally on \mathcal{G}_n and conditionally on \mathcal{G}_n respectively. We view the unconditional result as appropriate for the asymptotic framework in Adao et al. (2019) (in which elements of \mathcal{G}_n are resampled), and the conditional result as more suitable for the asymptotic framework in Goldsmith-Pinkham et al. (2020) (in which \mathcal{G}_n is not resampled).

Lemma A.5.1. *Let Assumptions A.5.1, A.5.2, A.5.3, A.5.4, and A.5.5 hold. (i) If $r_{1n}(\varrho, \mathcal{G}_n) \vee r_{2n}(\varrho, \mathcal{D}_n) = o_P(1)$ for any $\varrho > 0$ and $B_n^2(\mathcal{G}_n) \log^5(qb_n)/b_n = o_P(1)$, then*

$$\lim_{n \rightarrow \infty} P(T_n \leq \hat{c}_n) = 1 - \alpha.$$

(ii) If for any constants $\epsilon, \varrho > 0$ we have $r_{1n}(\varrho, \mathcal{G}_n) = o_{as}(1)$ and $P(r_{2n}(\varrho, \mathcal{D}_n) > \epsilon | \mathcal{G}_n) = o_{as}(1)$, and in addition $B_n^2(\mathcal{G}_n) \log^5(qb_n)/b_n = o_{as}(1)$, then it follows that

$$\lim_{n \rightarrow \infty} P(T_n \leq \hat{c}_n | \mathcal{G}_n) = 1 - \alpha \quad a.s.$$

Proof. We begin by defining the conditional covariance matrix $\Sigma(\mathcal{G}_n)$ to be given by

$$\Sigma(\mathcal{G}_n) \equiv \frac{1}{b_n} \sum_{i=1}^{b_n} E[\psi_i \psi_i' | \mathcal{G}_n]$$

and letting $\mathbb{T}_n \equiv \|\mathbb{G}_n\|_\infty$ for $\mathbb{G}_n \in \mathbb{R}^q$ a Gaussian vector satisfying $\mathbb{G}_n \sim N(0, \Sigma(\mathcal{G}_n))$. Further denote the linearized versions of T_n and T_n^* by letting L_n and L_n^* equal

$$L_n \equiv \left\| \frac{1}{\sqrt{b_n}} \sum_{i=1}^{b_n} \psi_i \right\|_\infty \quad L_n^* \equiv \left\| \frac{1}{\sqrt{b_n}} \sum_{i=1}^{b_n} \omega_i \left(\psi_i - \frac{1}{b_n} \sum_{j=1}^{b_n} \psi_j \right) \right\|_\infty \quad (\text{A.58})$$

and for notational convenience set $\delta_n(\mathcal{G}_n) \equiv (B_n^2(\mathcal{G}_n) \log^5(qb_n)/b_n)^{1/4}$. By Theorem 2.1 in Chernozhuokov et al. (2022) there then exists a C_1 not depending on \mathcal{G}_n such that

$$\sup_{t \in \mathbb{R}} |P(L_n \leq t | \mathcal{G}_n) - P(\mathbb{T}_n \leq t | \mathcal{G}_n)| \leq C_1 \delta_n(\mathcal{G}_n). \quad (\text{A.59})$$

Next note that, for any constant $\varrho > 0$, result (A.58) allows us to conclude that

$$\begin{aligned} & P\left(b_n^{-1/2} T_n \leq t \mid \mathcal{G}_n\right) \\ & \geq P(L_n \leq t - \varrho \mid \mathcal{G}_n) - P\left(|b_n^{-1/2} T_n - L_n| > \varrho \mid \mathcal{G}_n\right) \\ & \geq P(\mathbb{T}_n \leq t - \varrho \mid \mathcal{G}_n) - C_1 \delta_n(\mathcal{G}_n) - P\left(|b_n^{-1/2} T_n - L_n| > \varrho \mid \mathcal{G}_n\right) \\ & \geq P(\mathbb{T}_n \leq t \mid \mathcal{G}_n) - C_2 \varrho \sqrt{\log(q)} - C_1 \delta_n(\mathcal{G}_n) - P\left(|b_n^{-1/2} T_n - L_n| > \varrho \mid \mathcal{G}_n\right) \\ & \geq P(L_n \leq t \mid \mathcal{G}_n) - C_2 \eta - 2C_1 \delta_n(\mathcal{G}_n) - r_{1n}(\eta, \mathcal{G}_n), \end{aligned} \quad (\text{A.60})$$

where the third inequality holds for some constant C_2 not depending on \mathcal{G}_n by Lemma J.3 in Chernozhuokov et al. (2022), and the final inequality holds for any $\eta > 0$ by result (A.59), Assumption A.5.1, and setting $\varrho = \eta / \sqrt{\log(q)}$.

For $\mathcal{D}_n \equiv (\{Y_i, X_i, S_i\}_{i=1}^n, \mathcal{Z}, \mathcal{G}_n)$ and any constant $C_3 > 0$ next define the event

$$\mathcal{E}(\mathcal{D}_n) \equiv 1 \left\{ \sup_{t \in \mathbb{R}} |P(L_n \leq t | \mathcal{G}_n) - P(L_n^* \leq t | \mathcal{D}_n)| \leq C_3 \delta_n(\mathcal{G}_n) \right\}.$$

By Assumption A.5.5 and Lemmas 4.5 and 4.6 in Chernozhuokov et al. (2022) it then follows that we may select a constant C_3 not depending on \mathcal{G}_n under which we have

$$P(\mathcal{E}(\mathcal{D}_n) = 1 | \mathcal{G}_n) \geq 1 - \frac{2}{b_n} - 3v_n(\mathcal{G}_n), \quad v_n^2(\mathcal{G}_n) = \frac{B_n^2(\mathcal{G}_n) \log^3(qb_n)}{b_n}. \quad (\text{A.61})$$

Applying (A.61) and the same arguments as in (A.60) we obtain that if $\mathcal{E}(\mathcal{D}_n) = 1$, then

$$\begin{aligned} & P\left(b_n^{-1/2} T_n^* \leq t \mid \mathcal{D}_n\right) \\ & \leq P(L_n^* \leq t + \varrho \mid \mathcal{D}_n) + P\left(|b_n^{-1/2} T_n^* - L_n^*| > \varrho \mid \mathcal{D}_n\right) \\ & \leq P(L_n^* \leq t \mid \mathcal{D}_n) + C_2 \varrho \sqrt{\log(q)} + 2(C_1 + C_3) \delta_n(\mathcal{G}_n) + P\left(|b_n^{-1/2} T_n^* - L_n^*| > \varrho \mid \mathcal{D}_n\right) \\ & \leq P(L_n^* \leq t \mid \mathcal{D}_n) + C_2 \eta + 2(C_1 + C_3) \delta_n(\mathcal{G}_n) + r_{2n}(\eta, \mathcal{D}_n) \end{aligned} \quad (\text{A.62})$$

for any constant $\eta > 0$. Next, plug in $t = \hat{c}_n b_n^{-1/2}$ into result (A.62) and note that result (A.60) then implies that whenever $\mathcal{E}(\mathcal{D}_n) = 1$ we must have

$$\begin{aligned} 1 - \alpha & \leq P\left(L_n^* \leq \hat{c}_n b_n^{-1/2} \mid \mathcal{D}_n\right) + C_2 \eta + 2(C_1 + C_3) \delta_n(\mathcal{G}_n) + r_{2n}(\eta, \mathcal{D}_n) \\ & \leq P\left(L_n \leq \hat{c}_n b_n^{-1/2} \mid \mathcal{G}_n\right) + C_2 \eta + (2C_1 + 3C_3) \delta_n(\mathcal{G}_n) + r_{2n}(\eta, \mathcal{D}_n) \\ & \leq P(T_n \leq \hat{c} \mid \mathcal{G}_n) + 2C_2 \eta + 4(C_1 + C_3) \delta_n(\mathcal{G}_n) + r_{2n}(\eta, \mathcal{D}_n) + r_{1n}(\eta, \mathcal{G}_n) \\ & \equiv P(T_n \leq \hat{c}_n \mid \mathcal{G}_n) + s_n(\eta, \mathcal{D}_n), \end{aligned} \quad (\text{A.63})$$

where the final equality is definitional.

To conclude, let \mathcal{A}_n be a sigma field satisfying $\mathcal{G}_n \subseteq \mathcal{A}_n$ and note that for any $\varrho > 0$ we obtain from result (A.61) and the law of iterated expectations that

$$\begin{aligned} P(\mathcal{E}(\mathcal{D}_n) = 1 | \mathcal{A}_n) & \geq E[1\{v_n(\mathcal{G}_n) \leq \varrho\} P(\mathcal{E}(\mathcal{D}_n) = 1 | \mathcal{G}_n) | \mathcal{A}_n] \\ & \geq P(v_n(\mathcal{G}_n) \leq \varrho | \mathcal{A}_n) \left(1 - \frac{2}{b_n} - 3\varrho\right). \end{aligned} \quad (\text{A.64})$$

Similarly, for any $\varrho > 0$ we may select $\eta > 0$ sufficiently small so as to ensure that

$$P(s_n(\eta, \mathcal{D}_n) < \varrho | \mathcal{A}_n) \geq P(\delta_n(\mathcal{G}_n) \vee r_{1n}(\eta, \mathcal{G}_n) \vee r_{2n}(\eta, \mathcal{D}_n) > \epsilon \varrho | \mathcal{A}_n) \quad (\text{A.65})$$

for some $\epsilon > 0$. Finally, observe that (A.63) and the law of iterated expectations yield

$$\begin{aligned} P(T_n \leq \hat{c}_n | \mathcal{A}_n) &\geq E[1\{T_n \leq \hat{c}_n\}1\{P(T_n \leq \hat{c}_n | \mathcal{G}_n) \geq 1 - \alpha - \varrho\} | \mathcal{A}_n] \\ &\geq (1 - \alpha - \varrho)E[1\{P(T_n \leq \hat{c}_n | \mathcal{G}_n) \geq 1 - \alpha - \varrho\} | \mathcal{A}_n] \\ &\geq (1 - \alpha - \varrho)P(s(\eta, \mathcal{D}_n) < \varrho, \mathcal{E}(\mathcal{D}_n) = 1 | \mathcal{A}_n). \end{aligned} \quad (\text{A.66})$$

Part (i) of the lemma therefore follows from ϱ being arbitrary, results (A.64), (A.65), (A.66), and setting \mathcal{A}_n to be the trivial sigma field. Part (ii) of the lemma similarly follows from ϱ being arbitrary, results (A.64), (A.65), (A.66), and setting $\mathcal{A}_n = \mathcal{G}_n$. ■

References

- ADAO, R., KOLESÁR, M. and MORALES, E. (2019). Shift-share designs: Theory and inference. *The Quarterly Journal of Economics*, **134** 1949–2010.
- ANDREWS, D. (1988). Laws of large numbers for dependent non-identically distributed random variables. *Econometric Theory*, **4** 458–467.
- AUTOR, D. H., DORN, D. and HANSON, G. H. (2013). The china syndrome: Local labor market effects of import competition in the united states. *American economic review*, **103** 2121–68.
- BAI, Y., SANTOS, A. and SHAIKH, A. M. (2022). A two-step method for testing many moment inequalities. *Journal of Business & Economic Statistics*, **40** 1070–1080.
- BARTIK, T. J. (1991). *Who Benefits from State and Local Economic Development Policies?* W.E. Upjohn Institute for Employment Research, Kalamazoo, MI.
- BILLINGSLEY, P. (1968). *Convergence of Probability Measures*. Wiley Series in Probability and Mathematical Statistics, Wiley.
- BLANCHARD, O. and KATZ, L. F. (1992). Yregional evolutions, brooking papers on economic activity. *Economic Studies Program, The Brookings Institution*, **23** 76.
- BLANDHOL, C., BONNEY, J., MOGSTAD, M. and TORGOVITSKY, A. (2022). When is tsls actually late? Tech. rep., National Bureau of Economic Research.
- BORUSYAK, K., HULL, P. and JARAVEL, X. (2022). Quasi-experimental shift-share research designs. *The Review of Economic Studies*, **89** 181–213.
- BREIMAN, L. (1968). *Probability*. Classics in Applied Mathematics, Society for Industrial and Applied Mathematics.
- CARD, D. (2001). Immigrant inflows, native outflows, and the local labor market impacts of higher immigration. *Journal of Labor Economics*, **19** 22–64.
- CHAO, J. C., HAUSMAN, J. A., NEWEY, W. K., SWANSON, N. R. and WOUTERSEN, T. (2014). Testing overidentifying restrictions with many instruments and heteroskedasticity. *Journal of Econometrics*, **178** 15–21.
- CHEN, X. and SANTOS, A. (2018). Overidentification in regular models. *Econometrica*, **86** 1771–1817.

- CHERNOZHUKOV, V., CHETVERIKOV, D., KATO, K. and KOIKE, Y. (2022). Improved central limit theorem and bootstrap approximations in high dimensions. *The Annals of Statistics*, **50** 2562–2586.
- DE JONG, R. M. (1996). A strong law of large numbers for triangular mixingale arrays. *Statistics & Probability Letters*, **27** 1–9.
- DE JONG, R. M. (1997). Central limit theorems for dependent heterogeneous random variables. *Econometric Theory*, **13** 353–367.
- DIAMOND, R. (2016). The determinants and welfare implications of us workers’ diverging location choices by skill: 1980–2000. *American Economic Review*, **106** 479–524.
- DUDLEY, R. (1989). *Real Analysis and Probability*. Wadsworth & Brooks Cole.
- EAGLESON, G. K. (1975). Martingale convergence to mixtures of infinitely divisible laws. *The Annals of Probability*, **3** 557–562.
- GOLDSMITH-PINKHAM, P., SORKIN, I. and SWIFT, H. (2020). Bartik instruments: What, when, why, and how. *American Economic Review*, **110** 2586–2624.
- HAHN, J., KUERSTEINER, G. and MAZZOCCO, M. (2019). Estimation with Aggregate Shocks. *The Review of Economic Studies*, **87** 1365–1398.
- HAHN, J., KUERSTEINER, G. and MAZZOCCO, M. (2022). Joint time-series and cross-section limit theory under mixingale assumptions. *Econometric Theory*, **38** 942–958.
- HALMOS, P. (1976). *Measure Theory*. Graduate Texts in Mathematics, Springer New York.
- IMBENS, G. W. and ANGRIST, J. D. (1994). Identification and estimation of local average treatment effects. *Econometrica*, **62** 467–475.
- KOLESAR, M. (2013). Estimation in an instrumental variables model with treatment effect heterogeneity. Tech. rep., Princeton University.
- MAMMEN, E. (1993). Bootstrap and wild bootstrap for high dimensional linear models. *The Annals of Statistics*, **21** 255–285.
- MCLEISH, D. L. (1975). A Maximal Inequality and Dependent Strong Laws. *The Annals of Probability*, **3** 829 – 839.
- TOPALOVA, P. (2010). Factor immobility and regional impacts of trade liberalization: Evidence on poverty from india. *American Economic Journal: Applied Economics*, **2** 1–41.
- VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes: with Applications to Statistics*. Springer, New York.

Supplemental Appendix

Jinyong Hahn
hahn@econ.ucla.edu
UCLA

Guido Kuersteiner
gkuerte@umd.edu
University of Maryland

Andres Santos
andres@econ.ucla.edu
UCLA

Wavid Willigrod
dwwilligrod@gmail.com
UCLA

April 25, 2024

This supplemental appendix includes: (i) Calculations that justify the asymptotic validity of the proposed overidentification tests; and (ii) A set of Monte Carlo experiments evaluating the finite sample performance of such tests.

S.1 Influence Functions

In this section, we discuss how to verify the conditions of Lemma A.5.1, and hence Proposition 4.1, in the context of the overidentification tests of Sections 4.2 and 4.3.

S.1.1 Conditioning on Shocks

We first examine the overidentification test introduced in Section 4.2, which is designed for applications in which \mathcal{G}_n denotes a set of aggregate shocks that include \mathcal{Z} . Recall that in Section 4.2 we set $\psi_{ij} \equiv U_{ij}/\sigma_j$ and $\hat{\psi}_{ij} \equiv \hat{U}_{ij}/\hat{\sigma}_j$ with U_{ij} and \hat{U}_{ij} given by

$$\begin{aligned} U_{ij} &\equiv (S_{ij}\varepsilon_i - (E[S_{ij}(X_i, W_i')|\mathcal{G}_n])(E[A_i(X_i, W_i')|\mathcal{G}_n])^{-1}A_i\varepsilon_i \\ \hat{U}_{ij} &\equiv (S_{ij}\hat{\varepsilon}_i - (\frac{1}{n}\sum_{i=1}^n S_{ij}(X_i, W_i'))(\frac{1}{n}\sum_{i=1}^n A_i(X_i, W_i)')^{-1}A_i\hat{\varepsilon}_i, \end{aligned} \quad (\text{S.1})$$

where $A_i = (Z_i, W_i)'$ and σ_j^2 and $\hat{\sigma}_j^2$ denote the population and sample variances

$$\sigma_j^2 \equiv \text{var}\{U_{ij}|\mathcal{G}_n\} \quad \hat{\sigma}_j^2 \equiv \frac{1}{n}\sum_{i=1}^n (\hat{U}_{ij} - \frac{1}{n}\sum_{k=1}^n \hat{U}_{kj})^2. \quad (\text{S.2})$$

It will also prove helpful to define the variables $\{R_{ij}\}_{i=1}^n$ for $1 \leq j \leq p$ according to

$$R_{ij} \equiv S_{ij}\varepsilon_i - \left(\frac{1}{n} \sum_{k=1}^n S_{kj}(X_k, W'_k)\right) \left(\frac{1}{n} \sum_{k=1}^n A_k(X_k, W'_k)\right)^{-1} A_i \varepsilon_i. \quad (\text{S.3})$$

In particular, by definition of T_n and standard manipulations it then follows that

$$T_n = \max_{1 \leq j \leq p} \left| \sum_{i=1}^n \frac{R_{ij}}{\hat{\sigma}_j} \right|. \quad (\text{S.4})$$

In order to apply Lemma A.5.1(ii), to justify the asymptotic validity of the over-identification test of Section 4.2, we require suitable moment conditions and that

$$\max_{1 \leq j \leq p} \left| \frac{\sqrt{\log(p)}}{\sqrt{n}} \sum_{i=1}^n \frac{R_{ij} - U_{ij}}{\sigma_j} \right| = o_P(1) \quad (\text{S.5})$$

$$\max_{1 \leq j \leq p} \frac{\log^2(p)}{n} \sum_{i=1}^n \frac{(\hat{U}_{ij} - U_{ij})^2}{\sigma_j^2} = o_P(1), \quad (\text{S.6})$$

where probability statements are understood to be conditionally on \mathcal{G}_n and requirements (S.5) and (S.6) to hold almost surely in \mathcal{G}_n . By relying on Lemma D.5 in Chernozhukov et al. (2019), it is possible to show that requirement (S.6) in fact implies that

$$\log(p) \times \max_{1 \leq j \leq p} \left| \frac{\sigma_j}{\hat{\sigma}_j} - 1 \right| = o_P(1), \quad (\text{S.7})$$

where, again, probabilities are understood to be conditionally on \mathcal{G}_n and (S.7) to hold almost surely in \mathcal{G}_n . Moreover, by the triangle inequality and condition (S.5) we have

$$\begin{aligned} & \max_{1 \leq j \leq p} \left| \frac{\sqrt{\log(p)}}{\sqrt{n}} \sum_{i=1}^n \frac{R_{ij}}{\hat{\sigma}_j} - \frac{U_{ij}}{\sigma_j} \right| \\ & \leq \left(\max_{1 \leq j \leq p} \left| \frac{\sqrt{\log(p)}}{\sqrt{n}} \sum_{i=1}^n \frac{U_{ij}}{\sigma_j} \right| + o_P(1) \right) \times \max_{1 \leq j \leq p} \left| \frac{\sigma_j}{\hat{\sigma}_j} - 1 \right| = o_P(1), \end{aligned} \quad (\text{S.8})$$

where the final result follows from result (S.7) and a standard maximal inequality; see, e.g., Lemma 2.2.2 in van der Vaart and Wellner (1996). Result (S.8) together with (S.4) and $\psi_{ij} \equiv U_{ij}/\sigma_j$ imply that Assumption A.5.1 holds with $b_n = n$, $q = p$, and $r_{1n}(\varrho, \mathcal{G}_n)$ satisfying $r_{1n}(\varrho, \mathcal{G}_n) = o_{as}(1)$ for any $\varrho > 0$.

In order to verify Assumption A.5.3, recall that $\mathcal{D}_n \equiv (\{Y_i, X_i, S_i\}_{i=1}^n, \mathcal{Z}, \mathcal{G}_n)$ and note that if $\{\omega_i\}_{i=1}^n$ are i.i.d. standard normal random variables independent of \mathcal{D}_n ,

then a standard maximal inequality yields that

$$E\left[\max_{1 \leq j \leq p} \left| \frac{\sqrt{\log(p)}}{\sqrt{n}} \sum_{i=1}^n \omega_i ((\hat{\psi}_{ij} - \psi_{ij}) - \frac{1}{n} \sum_{k=1}^n (\hat{\psi}_{kj} - \psi_{kj})) \right| \right] \lesssim \sqrt{\log(p)} \times \max_{1 \leq j \leq p} \left(\frac{\log(p)}{n} \sum_{i=1}^n (\hat{\psi}_{ij} - \psi_{ij})^2 \right)^{1/2}. \quad (\text{S.9})$$

Moreover, employing that $\hat{\psi}_{ij} \equiv \hat{U}_{ij}/\hat{\sigma}_j$ and $\psi_{ij} \equiv U_{ij}/\sigma_j$ we can bound (S.9) by

$$\begin{aligned} & \max_{1 \leq j \leq p} \frac{\log^2(p)}{n} \sum_{i=1}^n (\hat{\psi}_{ij} - \psi_{ij})^2 \\ & \lesssim \frac{\log^2(p)}{n} \times \left(\max_{1 \leq j \leq p} \sum_{i=1}^n \frac{(\hat{U}_{ij} - U_{ij})^2}{\sigma_j^2} (1 + o_P(1)) + \max_{1 \leq j \leq p} \sum_{i=1}^n \frac{U_{ij}^2}{\sigma_j^2} \left(\frac{\sigma_j}{\hat{\sigma}_j} - 1 \right)^2 \right) = o_P(1), \end{aligned}$$

where the final result holds by (S.6) and a standard maximal inequality. By Markov's inequality and result (S.9) it follows that Assumption A.5.3 holds for some $r_{2n}(\varrho, \mathcal{D}_n)$ satisfying $P(r_{2n}(\varrho, \mathcal{D}_n) > \epsilon | \mathcal{D}_n) = o_{as}(1)$ for any $\varrho > 0$ as required by Lemma A.5.1.

S.1.2 Identification Through Shocks

We next discuss the overidentification test of Section 4.3, which is designed for applications in which identification is driven by exogeneity of the shocks \mathcal{Z} . Recall that in the corresponding asymptotic framework, originally developed by Adao et al. (2019), we set $\mathcal{G}_n = \{S_i, W_i, \varepsilon_i\}_{i=1}^n$. Following the notation in Section 4.3, we further set

$$\begin{aligned} \delta_j & \equiv \left(\sum_{i=1}^n W_i W_i' \right)^{-1} \sum_{i=1}^n W_i g_j(\varepsilon_i, W_i, S_i) \\ \kappa_j & \equiv \left(\sum_{i=1}^n E[S_i' \mathcal{E} X_i | \mathcal{G}_n] \right)^{-1} \left(\sum_{i=1}^n E[S_i' \mathcal{E} X_i | \mathcal{G}_n] \frac{\partial}{\partial \varepsilon} g_j(\varepsilon_i, W_i, S_i) \right), \end{aligned} \quad (\text{S.10})$$

where $\mathcal{E} \equiv \mathcal{Z} - E[\mathcal{Z} | \mathcal{G}_n]$ and we note that δ_j and κ_j depend on n (through \mathcal{G}_n), but we suppress the dependence from the notation. As estimators for δ_j and κ_j we employ

$$\begin{aligned} \hat{\delta}_j & \equiv \left(\sum_{i=1}^n W_i W_i' \right)^{-1} \sum_{i=1}^n W_i g_j(\hat{\varepsilon}_i, W_i, S_i) \\ \hat{\kappa}_j & \equiv \left(\sum_{i=1}^n (Z_i - W_i' \hat{\pi}_n) X_i \right)^{-1} \left(\sum_{i=1}^n (Z_i - W_i' \hat{\pi}_n) X_i \frac{\partial}{\partial \varepsilon} g_j(\hat{\varepsilon}_i, W_i, S_i) \right), \end{aligned}$$

where $\hat{\pi}_n$ denotes the coefficient from regressing $\{Z_i\}_{i=1}^n$ on $\{W_i\}_{i=1}^n$. In addition let

$$U_{ij} \equiv \mathcal{E}_i \times \sum_{k=1}^n S_{ki} (g_j(\varepsilon_k, W_k, S_k) - W'_k \delta_j - \varepsilon_k \kappa_j)$$

$$\hat{U}_{ij} \equiv \hat{\mathcal{E}}_i \times \sum_{k=1}^n S_{ki} (g_j(\hat{\varepsilon}_k, W_k, S_k) - W'_k \hat{\delta}_j - \hat{\varepsilon}_k \hat{\kappa}_j)$$

for $\hat{\mathcal{E}}_i$ an estimator of \mathcal{E}_i (see Remark 4.2), and recall that $\psi_{ij} \equiv U_{ij}/\sigma_j$ and $\hat{\psi}_{ij} \equiv \hat{U}_{ij}/\hat{\sigma}_j$, where σ_j and $\hat{\sigma}_j$ respectively denote the population and finite sample variances

$$\sigma_j^2 \equiv \frac{1}{p} \sum_{i=1}^p \text{Var}\{U_{ij}|\mathcal{G}_n\} \quad \hat{\sigma}_j^2 \equiv \frac{1}{p} \sum_{i=1}^p \left(\hat{U}_{ij} - \frac{1}{p} \sum_{k=1}^p \hat{U}_{kj} \right)^2.$$

It will also prove convenient to define the variables $\{R_{ij}\}_{i=1}^n$ for $1 \leq j \leq q$ according to

$$R_{ij} \equiv g_j(\hat{\varepsilon}_i, W_i, S_i)(Z_i - W'_i \hat{\pi}_n).$$

The asymptotic validity of the overidentification test of Section 4.3 may be justified by employing Lemma A.5.1(i). In order to appeal to Lemma A.5.1(i), first note that if $\{\mathcal{E}_i\}_{i=1}^p$ are (uniformly) Sub-Gaussian almost surely in \mathcal{G}_n , then Assumption A.5.2 can be verified by setting $B_n(\mathcal{G}_n) = KC_n$ for K large enough and C_n given by

$$C_n \equiv \max_{1 \leq i \leq p} \max_{1 \leq j \leq q} \left(\frac{\text{Var}\{U_{ij}|\mathcal{G}_n\}}{\frac{1}{p} \sum_{i=1}^p \text{Var}\{U_{ij}|\mathcal{G}_n\}} \right)^{1/2}.$$

In turn, Assumptions A.5.1 and A.5.2 can be verified under the key requirements

$$\max_{1 \leq j \leq q} \frac{\sqrt{\log(q)}}{\sigma_j \sqrt{p}} \left| \sum_{i=1}^n R_{ij} - \sum_{i=1}^p U_{ij} \right| = o_P(1) \quad (\text{S.11})$$

$$\max_{1 \leq j \leq q} \frac{\log^2(q)}{p} \sum_{i=1}^p \frac{(\hat{U}_{ij} - U_{ij})^2}{\sigma_j^2} = o_P(1), \quad (\text{S.12})$$

where the convergence in probability statement should be understood as jointly over all the data (rather than conditionally on \mathcal{G}_n). In particular, under the condition that

$$\frac{C_n^2 \log^2(q)}{p^{(1-c)/2}} = o_P(1) \quad (\text{S.13})$$

for some $0 < c < 1$, it is possible to argue by relying on Lemma D.5 in Chernozhukov et al. (2019) that requirement (S.12) in fact implies that

$$\log(q) \times \max_{1 \leq j \leq q} \left| \frac{\sigma_j}{\hat{\sigma}_j} - 1 \right| = o_P(1). \quad (\text{S.14})$$

Moreover, by applying Lemma D.3 in [Chernozhukov et al. \(2019\)](#) and relying on the rate condition in [\(S.13\)](#) it is also possible to obtain the rate bounds

$$\max_{1 \leq j \leq q} \left| \frac{1}{\sqrt{p}} \sum_{i=1}^p \frac{U_{ij}}{\sigma_j} \right| = O_P(\sqrt{\log(q)}) \quad \max_{1 \leq j \leq q} \left| \frac{1}{p} \sum_{i=1}^p \frac{U_{ij}}{\sigma_j} \right|^2 = O_P(1). \quad (\text{S.15})$$

Combining results [\(S.14\)](#) and [\(S.15\)](#) with the same arguments employed in Section [S.1.1](#), it is then straightforward to show that conditions [\(S.11\)](#) and [\(S.12\)](#) imply Assumptions [A.5.1](#) and [A.5.4](#) hold with $b_n = p$ and $r_{1n}(\varrho, \mathcal{G}_n) \vee r_{2n}(\varrho, \mathcal{G}_n) = o_P(1)$. To conclude verifying the main requirements of Lemma [A.5.1\(i\)](#) we note that the condition $B_n^2(\mathcal{G}_n) \log^5(qp)/p = o_P(1)$ is implied by requirement [\(S.13\)](#) (up to logs).

Condition [\(S.11\)](#) is more challenging to verify than its analogue in Section [S.1.1](#) (i.e. [\(S.5\)](#)) because there are n terms $\{R_{ij}\}$ but p terms $\{U_{ij}\}$. Fortunately, as we next outline, it is possible to establish that [\(S.11\)](#) holds by building on the assumptions and arguments in [Adao et al. \(2019\)](#). To this end, we start with the decomposition

$$\sum_{i=1}^n R_{ij} = \sum_{i=1}^n g_j(\varepsilon_i, W_i, S_i)(Z_i - W_i' \hat{\pi}_n) \quad (\text{S.16})$$

$$+ \sum_{i=1}^n (g_j(\hat{\varepsilon}_i, W_i, S_i) - g_j(\varepsilon_i, W_i, S_i))(Z_i - W_i' \pi_n) \quad (\text{S.17})$$

$$+ \sum_{i=1}^n (g_j(\hat{\varepsilon}_i, W_i, S_i) - g_j(\varepsilon_i, W_i, S_i))W_i'(\hat{\pi}_n - \pi_n). \quad (\text{S.18})$$

It is also helpful to note that since $Z_i = S_i' \mathcal{Z}$, $E[Z_i | \mathcal{G}_n] = W_i' \pi_n$ under the null hypothesis, and $E[Z_i | \mathcal{G}_n] = S_i' E[\mathcal{Z} | \mathcal{G}_n]$ due to $S_i \in \mathcal{G}_n$, it follows that

$$Z_i - W_i' \pi_n = S_i' \mathcal{E}. \quad (\text{S.19})$$

Next, note that [\(S.19\)](#), the definition of δ_j in [\(S.10\)](#), the equality $S_k' \mathcal{E} = \sum_{i=1}^p S_{ki} \mathcal{E}_i$, and some algebra allows us to express term [\(S.16\)](#) as being equal to

$$\begin{aligned} \sum_{i=1}^n g_j(\varepsilon_i, W_i, S_i)(Z_i - W_i' \hat{\pi}_n) &= \sum_{i=1}^n g_j(\varepsilon_i, W_i, S_i) \{ (Z_i - W_i' \pi_n) + W_i' (\pi_n - \hat{\pi}_n) \} \\ &= \sum_{i=1}^n g_j(\varepsilon_i, W_i, S_i) \{ S_i' \mathcal{E} - W_i' \left(\sum_{k=1}^p W_k W_k' \right)^{-1} \sum_{l=1}^p W_l S_l' \mathcal{E} \} \\ &= \sum_{i=1}^p \mathcal{E}_i \times \left(\sum_{k=1}^n S_{ki} (g_j(\varepsilon_k, W_k, S_k) - W_k' \delta_j) \right). \end{aligned}$$

We analyze the term in (S.17) through a linearization argument. To this end, we set

$$M_{1q} \equiv \max_{1 \leq j \leq q} \left\| \frac{\partial}{\partial \varepsilon} g_j \right\|_{\infty} \quad M_{2q} \equiv \max_{1 \leq j \leq q} \left\| \frac{\partial^2}{\partial \varepsilon^2} g_j \right\|_{\infty}$$

and, following [Adao et al. \(2019\)](#), we let $n_k \equiv \sum_{i=1}^n S_{ki}$ and set $r_n = (\sum_{k=1}^p n_k^2)^{-1}$. Then note that by result (S.19) and a standard Taylor expansion we obtain that

$$\begin{aligned} & \sum_{i=1}^n (g_j(\hat{\varepsilon}_i, W_i, S_i) - g_j(\varepsilon_i, W_i, S_i))(Z_i - W_i' \pi_n) \\ &= \sum_{i=1}^n \frac{\partial}{\partial \varepsilon} g_j(\varepsilon_i, W_i, S_i) S_i' \mathcal{E}(X_i(\beta - \hat{\beta}_n) + W_i'(\gamma_s - \hat{\gamma}_n)) \\ & \quad + \sum_{i=1}^n \frac{\partial^2}{\partial \varepsilon^2} g_j(\tilde{\varepsilon}_i, W_i, S_i) (\hat{\varepsilon}_i - \varepsilon_i)^2 S_i' \mathcal{E}, \end{aligned} \quad (\text{S.20})$$

where $\tilde{\varepsilon}_i$ is some intermediate value between $\hat{\varepsilon}_i$ and ε_i . If the covariates W_i are bounded almost surely, then a maximal inequality (applied conditionally on \mathcal{G}_n) yields

$$\begin{aligned} E \left[\max_{1 \leq j \leq q} \left\| \sum_{i=1}^p \mathcal{E}_i \left\{ \sum_{k=1}^n S_{ik} \frac{\partial}{\partial \varepsilon} g_j(\varepsilon_k, W_k, S_k) W_k' \right\} \right\| \middle| \mathcal{G}_n \right] \\ \lesssim \sqrt{\log(q)} M_{1q} \times \left(\sum_{i=1}^p \left(\sum_{k=1}^n S_{ik} \right)^2 \right)^{1/2} = \frac{\sqrt{\log(q)} M_{1q}}{\sqrt{r_n}}, \end{aligned}$$

where the final equality follows by definition of r_n . We can therefore conclude that

$$\max_{1 \leq j \leq q} \left| \sum_{i=1}^n \frac{\partial}{\partial \varepsilon} g_j(\varepsilon_i, W_i, S_i) S_i' \mathcal{E} W_i' (\gamma_s - \hat{\gamma}_n) \right| = O_P \left(\frac{\sqrt{\log(q)} M_{1q}}{\sqrt{r_n}} \|\gamma_s - \hat{\gamma}_n\| \right),$$

where the probability is understood to be over the entire data. Similarly, adapting the arguments in the proof of Proposition 3 in [Adao et al. \(2019\)](#) (see in particular the proof of their display (A.4)) and employing a maximal inequality for degenerate U-statistics (see, e.g., equation (3.5) in [Giné et al. \(2000\)](#)) it is possible to establish that

$$\max_{1 \leq j \leq q} \left| \sum_{i=1}^n \frac{\partial}{\partial \varepsilon} g_j(\varepsilon_i, W_i, S_i) (S_i' \mathcal{E} X_i - E[S_i' \mathcal{E} X_i | \mathcal{G}_n]) (\beta - \hat{\beta}_n) \right| = O_P \left(\frac{\log(q) M_{1q}}{\sqrt{r_n}} |\hat{\beta}_n - \beta| \right).$$

Moreover, the arguments in [Adao et al. \(2019\)](#) can additionally be used to conclude that

$$\hat{\beta}_n - \beta = \frac{\sum_{i=1}^n S_i' \mathcal{E} \varepsilon_i}{\sum_{i=1}^n E[S_i' \mathcal{E} X_i | \mathcal{G}_n]} + O_P \left(\frac{\|\hat{\gamma} - \gamma_s\|}{n \sqrt{r_n}} + \frac{1}{n^2 r_n} \right), \quad (\text{S.21})$$

while sup norm bound on the quadratic term in (S.20) and the definition of M_{2q} imply

$$\max_{1 \leq j \leq q} \left| \sum_{i=1}^n \frac{\partial^2}{\partial \varepsilon^2} g_j(\hat{\varepsilon}_i, W_i, S_i) (\hat{\varepsilon}_i - \varepsilon_i)^2 S_i' \mathcal{E} \right| = O_P(nM_{2q}(|\hat{\beta}_n - \beta|^2 + \|\hat{\gamma} - \gamma_s\|^2)).$$

Thus, since result (S.21) implies that $|\hat{\beta}_n - \beta| = O_P((n\sqrt{r_n})^{-1})$, our analysis so far yields

$$\begin{aligned} \max_{1 \leq j \leq q} \left| \sum_{i=1}^n (g_j(\hat{\varepsilon}_i, W_i, S_i) - g_j(\varepsilon_i, W_i, S_i))(Z_i - W_i' \pi_n) - \kappa_j \sum_{i=1}^n S_i' \mathcal{E} \varepsilon_i \right| \\ = O_P\left(\frac{\sqrt{\log(q)} M_{1q}}{\sqrt{r_n}} \|\hat{\gamma} - \gamma_s\| + nM_{2q} \|\hat{\gamma} - \gamma_s\|^2 + \log(q) \frac{M_{1q} \vee M_{2q}}{nr_n}\right). \end{aligned}$$

Finally, using that $\|\hat{\pi}_n - \pi_n\| \vee |\hat{\beta} - \beta| = O_P((n\sqrt{r_n})^{-1})$ and relying on the mean value theorem allows to bound in probability the term in (S.18) by

$$\max_{1 \leq j \leq q} \left| \sum_{i=1}^n (g_j(\hat{\varepsilon}_i, W_i, S_i) - g_j(\varepsilon_i, W_i, S_i)) W_i' (\hat{\pi}_n - \pi_n) \right| = O_P\left(\frac{M_{1q}}{\sqrt{r_n}} \left(\frac{1}{n\sqrt{r_n}} + \|\hat{\gamma} - \gamma_s\|\right)\right).$$

To simplify our bounds, we suppose that $\|\hat{\gamma} - \gamma_s\|$ has the same rate of convergence as $|\hat{\beta} - \beta|$ so that $\|\hat{\gamma} - \gamma_s\| = O_P((n\sqrt{r_n})^{-1})$. Combining our analysis of the terms in (S.16)-(S.18) together with the definition of R_{ij} and U_{ij} we can then conclude that

$$\max_{1 \leq j \leq q} \left| \sum_{i=1}^n R_{ij} - \sum_{i=1}^p U_{ij} \right| = O_P(\log(q) \frac{M_{1q} \vee M_{2q}}{nr_n}).$$

Thus, finally setting $\underline{\sigma} \equiv \min_{1 \leq j \leq q} \sigma_j$ we obtain that (S.11) is implied by the condition

$$\frac{\log^{3/2}(q)(M_{1q} \vee M_{2q})}{\underline{\sigma} \sqrt{pnr_n}} = o_P(1).$$

S.2 Simulation Evidence

We next conduct a series of Monte Carlo simulations to evaluate the finite sample performance of the overidentification tests proposed in Section 4. With the goal of informing the implementation of our tests in the empirical application of Section 5, we employ simulation designs based on the Autor et al. (2013) dataset. In particular, as in Autor et al. (2013), our designs consist of short panels with $T = 2$ time periods, $n = 722$ commuting zones, and $p = 397$ sectors defined by four digit SIC codes.

In what follows, we incorporate the short panel structure into our notation by letting Y_{it} , X_{it} , W_{it} , and Z_{it} respectively denote the outcome, regressor, controls, and instrument for commuting zone i at time periods t . We also note that in Autor et al. (2013) both the regressor X_{it} and instrument Z_{it} have a Bartik structure and hence we now

index shares and aggregate shocks by subscripts x and z . Concretely, we have

$$X_{it} = S'_{xit} \mathcal{Z}_{xt} \quad Z_{it} = S'_{zit} \mathcal{Z}_{zt} \quad (\text{S.22})$$

where S_{xit} and S_{zit} represent share vectors for commuting zone i at time t and \mathcal{Z}_{xt} and \mathcal{Z}_{zt} denote aggregate shocks at time t . Finally, because [Autor et al. \(2013\)](#) weight all observations by the start of period commuting zone population, in our simulations we employ the same weights throughout the analysis.

S.2.1 Conditioning on Shocks

We begin by examining the finite sample performance of the overidentification test proposed in [Section 4.2](#), which recall was designed for applications that implicitly condition on the aggregate shocks – i.e. that employ asymptotic approximations based on only n growing. As discussed in [Remark 2.1](#), implicitly conditioning on aggregate shocks in short panels yields the overidentifying moment restrictions

$$E[S_{zit}\varepsilon_{it}] = 0 \quad \text{for } 1 \leq t \leq T. \quad (\text{S.23})$$

Since $T = 2$ and $p = 397$ in the context of [Autor et al. \(2013\)](#), result (S.23) represents a total of 794 possible moment restrictions. Moreover, because [Autor et al. \(2013\)](#) cluster observations at the state level, their effective number of observations is 48.

In designing our simulations, we aimed to reflect the clustering structure in [Autor et al. \(2013\)](#) by employing a heteroskedastic version of the group shock model of [Moulton \(1986\)](#). To this end, we let c denote a cluster, which consists of the commuting zone time pairs (i, t) for which i belongs to the state represented by c , and let C denote the collection of all clusters. Employing the [Autor et al. \(2013\)](#) dataset, we then estimate a model in which the errors ε_{it} are assumed to have the structure

$$\varepsilon_{it} = \eta_c + \zeta_{it}$$

where η_c are i.i.d. cluster level shocks and ζ_{it} are i.i.d. shocks and independent of η_c . We further impose a parsimonious heteroskedasticity specification by supposing that

$$E[\eta_c^2 | \mathcal{A}_n] = a_\eta + s_\eta \left(\sum_{(i,t) \in c} S'_{zit} S_{zit} \right)$$

$$E[\zeta_{it}^2 | \mathcal{A}_n] = a_\zeta + s_\zeta (S'_{zit} S_{zit})$$

for some constants a_η, s_η, a_ζ , and s_ζ and $\mathcal{A}_n \equiv \{S_{zit}, S_{xit}, W_{it}\}$. In order to estimate this model, we employ the fitted residuals $\{\hat{\varepsilon}_{it}\}$ from the weighted instrumental variable

estimation in the main specification of Autor et al. (2013) and let

$$\begin{aligned}
(\hat{a}_\eta, \hat{s}_\eta) &\equiv \arg \min_{a, s \in \mathbf{R}} \sum_{c \in C} \sum_{(it) \neq (\tilde{i}, \tilde{t}) \in c} (\hat{e}_{it} \hat{e}_{\tilde{i}\tilde{t}} - a - s(\sum_{(i't') \in c} S'_{zi't'} S_{zi't'}))^2 \\
(\hat{a}_\zeta, \hat{s}_\zeta) &\equiv \arg \min_{a, s \in \mathbf{R}} \sum_{c \in C} \sum_{(i,t) \in c} (\hat{e}_{it}^2 - \hat{\sigma}_{\eta,c}^2 - a - s(S'_{zit} S_{zit}))^2
\end{aligned} \tag{S.24}$$

where

$$\hat{\sigma}_{\eta,c}^2 \equiv \max\{\hat{a}_\eta + \hat{s}_\eta(\sum_{(i,t) \in c} S'_{zit} S_{zit}), 0\}.$$

Given these estimates, we generate our Monte Carlo samples as follows:

STEP 1. We employ the same controls $\{W_{it}\}$, aggregate shocks \mathcal{Z}_{zt} and \mathcal{Z}_{xt} , and regression weights as in the main specification of Autor et al. (2013), which we keep fixed across all the simulations. ■

STEP 2. For each $t \in \{1, 2\}$ we draw n observations $\{S_{xit}^*, S_{zit}^*\}_{i=1}^n$ with replacement from the original full sample set of shares $\{S_{xit}, S_{zit}\}_{i=1}^n$. ■

STEP 3. Given the sample $\{S_{xit}^*, S_{zit}^*\}$, we create a sample of instruments $\{Z_{it}^*\}$ and endogenous variables $\{X_{it}^*\}$ by setting $Z_{it}^* \equiv (S_{zit}^*)' \mathcal{Z}_{zt}$ and letting $X_{it}^* \equiv (S_{xit}^*)' \mathcal{Z}_{xt}$. ■

STEP 4. For each commuting zone time pair (i, t) and cluster c we create the variances

$$\begin{aligned}
(\hat{\sigma}_{\zeta,it}^*)^2 &\equiv \max\{\hat{a}_\zeta + \hat{s}_\zeta(S_{zit}^*)' S_{zit}^*, 0\} \\
(\hat{\sigma}_{\eta,c}^*)^2 &\equiv \max\{\hat{a}_\eta + \hat{s}_\eta \sum_{(i,t) \in c} (S_{zit}^*)' S_{zit}^*, 0\}
\end{aligned}$$

by employing the full sample estimates $\hat{a}_\eta, \hat{s}_\eta, \hat{a}_\zeta$ and \hat{s}_ζ from (S.24). ■

STEP 5. To create a sample of outcomes for our simulations, we draw $|C|$ i.i.d. standard normal variables $\{V_c\}_{c \in C}$, $n \times T$ i.i.d. standard normal variables $\{U_{it}\}$, and set

$$Y_{it}^* \equiv X_{it}^* \hat{\beta} + W_{it}' \hat{\gamma}_s + V_c \hat{\sigma}_{\eta,c}^* + U_{it} \hat{\sigma}_{\zeta,it}^*,$$

where $(\hat{\beta}, \hat{\gamma}_s)$ denote the weighted instrumental variable estimators from the main specification in Autor et al. (2013) and the “ c ” subscript is understood to refer to the state to which commuting zone i belongs. ■

By repeating Steps 1-5 we generate one thousand samples $\{Y_{it}^*, X_{it}^*, Z_{it}^*, W_{it}\}$ on which we evaluate the finite sample properties of our test. We note that the number of moment restrictions ($p \times T = 794$) in (S.23) far exceeds the number of clusters in the simulations (48 states). Since any linear combination of the moment restrictions in (S.23) is also a valid moment restriction, we also examine the performance of our test when adding restrictions across time periods and/or different levels of SIC codes. Table

Table 3: Rejection Probabilities

Moment Restrictions	# Moments	Significance Level			p-value
		1%	5%	10%	
Four Digit SIC & all time periods	794	0.000	0.004	0.013	0.1274
Four Digit SIC & time aggregated	397	0.000	0.008	0.031	0.168
Three Digit SIC & all time periods	272	0.000	0.007	0.029	0.0488
Three Digit SIC & time aggregated	136	0.000	0.007	0.025	0.0748
Two Digit SIC & all time periods	40	0.001	0.030	0.072	0.0054
Two Digit SIC & time aggregated	20	0.004	0.034	0.095	0.0018

Finite sample rejection probabilities for overidentification tests of the validity of the moment restrictions imposed when asymptotic approximation implicitly condition on aggregate shocks.

3 reports the finite sample rejection probabilities for tests based on different choices of moments and significance levels. The final column of Table 3 additionally reports the p-value obtained when the test is implemented in the data of Autor et al. (2013). All critical values were obtained by following the procedure in Section 4.2 with one thousand bootstrap draws.

Overall we find that the test is able to control size across all specifications. However, for larger values of the number of moments, the finite sample rejection probability of the test is significantly below its nominal level. The test performs best when aggregating across time periods and two digit SIC codes. For this specification, which consists of twenty moments, the finite sample rejection probabilities of the test are close to the nominal levels. Because the design only has 48 clusters, we view this specification as still employing a large number of moments relative to the sample size.

S.2.2 Identification Through Shocks

We conclude by examining the finite sample performance of the overidentification test proposed in Section 4.3, which was designed for applications in which the exogeneity of the instrument is due to the exogeneity of the aggregate shocks. Recall that in these applications $\mathcal{G}_n = \{S_{zit}, S_{xit}, W_{it}, \varepsilon_{it}\}$ and the overidentifying restriction is given by

$$E[Z_{it}|\mathcal{G}_n] = W_{it}'\pi_n \tag{S.25}$$

where

$$\pi_n = \left(\sum_{t=1}^T \sum_{i=1}^n W_{it} W_{it}' \right)^{-1} \sum_{t=1}^T \sum_{i=1}^n W_{it} (S_{zit}' E[Z_{it}|\mathcal{G}_n]).$$

In order to ensure that the null hypothesis holds in our simulation design, we rely on a model proposed by Adao et al. (2019) as a sufficient condition for (S.25). Specifically,

we suppose that for some $p \times d$ matrix of shock \mathcal{Z}_{wt} and $d \times 1$ vector Γ we have

$$E[\mathcal{Z}_{zt}|\mathcal{G}_n] = \mathcal{Z}_{wt}\Gamma \quad W_{it} = \mathcal{Z}'_{wt}S_{zit}. \quad (\text{S.26})$$

To estimate this model in the original [Autor et al. \(2013\)](#) dataset, we first compute a ridge regression of the coordinates of $W_{it} \in \mathbb{R}^d$ on S_{zit} by setting

$$\hat{\delta}_{jt} \equiv \left(\sum_{i=1}^n S_{zit}S'_{zit} + \lambda I_p \right)^{-1} \sum_{i=1}^n S_{zit}W_{itj},$$

for each j and t , where W_{itj} denotes the j^{th} coordinate of the vector $W_{it} \in \mathbb{R}^d$, I_p is a $p \times p$ identity matrix, and we set the penalty λ to equal 0.1. Given these estimates, we let $\hat{\mathcal{Z}}_{wt} \equiv [\hat{\delta}_{1t}, \dots, \hat{\delta}_{dt}]$ and estimate Γ through the regression

$$\hat{\Gamma} \equiv \arg \min_{g \in \mathbb{R}^d} \sum_{t=1}^T \|\mathcal{Z}_{zt} - \hat{\mathcal{Z}}_{wt}g\|^2.$$

In what follows, it will be helpful to define $\hat{E}[\mathcal{Z}_{zt}|\mathcal{G}_n] \equiv \hat{\mathcal{Z}}_{wt}\hat{\Gamma}$ and $\hat{\nu}_t \equiv \mathcal{Z}_{zt} - \hat{\mathcal{Z}}_{wt}\hat{\Gamma}$.

We further aim to reflect the clustering structure in [Autor et al. \(2013\)](#). To this end, we follow [Adao et al. \(2019\)](#) and [Borusyak et al. \(2022\)](#) who in re-examining the empirical analysis of [Autor et al. \(2013\)](#) cluster shocks at the three digit SIC code. As in [Section S.2.1](#), we estimate a common shock model in which \mathcal{Z}_{zt} satisfies

$$\mathcal{Z}_{ztj} = \eta_c + \zeta_{tj}$$

where \mathcal{Z}_{ztj} denotes the j^{th} coordinate of \mathcal{Z}_{zt} , η_c are i.i.d. cluster level shocks, and ζ_{tj} are i.i.d. shocks independent of η_c . We estimate the variance of these shocks by setting

$$\begin{aligned} \hat{\sigma}_\eta^2 &\equiv \frac{1}{|C|} \sum_{c \in C} \frac{1}{n_c(n_c - 1)} \sum_{(t,j) \neq (\tilde{t}, \tilde{j}) \in c} \hat{\nu}_{tj} \hat{\nu}_{\tilde{t}\tilde{j}} \\ \hat{\sigma}_\zeta^2 &\equiv \frac{1}{|C|} \sum_{c \in C} \frac{1}{n_c} \sum_{(s,t) \in c} \hat{\nu}_{tj}^2 - \hat{\sigma}_\eta^2, \end{aligned}$$

where n_c denotes the number of observations in cluster c and ν_{tj} denotes the j^{th} coordinate of the vector $\hat{\nu}_t \in \mathbb{R}^p$.

Finally, in order to reflect the strength of the instrument in [Autor et al. \(2013\)](#) in our simulation design, we run the following regression on the aggregate shocks

$$(\hat{\alpha}, \hat{\kappa}) \equiv \arg \min_{a, k \in \mathbb{R}} \sum_{t=1}^T \sum_{j=1}^p (\mathcal{Z}_{xtj} - a - k\mathcal{Z}_{ztj})^2$$

and let $\hat{\sigma}_\xi^2$ denote the sample variance of the residuals $\hat{\xi}_{tj} \equiv \mathcal{Z}_{xtj} - \hat{\alpha} - \hat{\kappa}\mathcal{Z}_{ztj}$.

Given these estimates, we generate our Monte Carlo samples as follows:

STEP 1. We first create controls $\hat{W}_{it} \equiv \hat{Z}'_{wt} S_{zit}$, which we note have the structure required by model (S.26). We combine the controls $\{\hat{W}_{it}\}$ with the shares $\{S_{zit}, S_{xit}\}$ in Autor et al. (2013) and keep all of them fixed across simulation designs. ■

STEP 2. To generate our instrument, we first draw $|C|$ i.i.d. standard normal variables $\{V_c\}_{c \in C}$, $p \times T$ i.i.d. standard normal variables $\{U_{ztj}\}$, and set

$$Z_{ztj}^* = \hat{E}[Z_{ztj} | \mathcal{G}_n] + V_c \hat{\sigma}_\eta + U_{ztj} \hat{\sigma}_\zeta,$$

where the “ c ” subscript refers to the three digit SIC code to which sector j belongs. As our instrument we then employ $Z_{it}^* \equiv S'_{zit} Z_{zt}^*$. Note that, because $\hat{E}[Z_{zt} | \mathcal{G}_n] \equiv \hat{Z}_{wt} \hat{\Gamma}$, the shocks Z_{zt}^* have the structure required by model (S.26). ■

STEP 3. Similarly, in order to generate aggregate shocks for our regressor, we draw $p \times T$ i.i.d. standard normal random variables $\{U_{xtj}\}$ and let

$$Z_{xtj} = \hat{\alpha} + \hat{\kappa} Z_{ztj}^* + V_{xtj} \hat{\sigma}_\xi.$$

As our regressor, we then employ $X_{it}^* \equiv S'_{xit} Z_{xt}$. ■

STEP 4. Finally, we generate a sample of outcomes Y_{it}^* by simply setting Y_{it}^* to equal

$$Y_{it}^* = X_{it}^* \hat{\beta} + \hat{W}'_{it} \hat{\gamma}_s + \hat{e}_{it},$$

where $(\hat{\beta}, \hat{\gamma}_s)$ and $\{\hat{e}_{it}\}$ denote the estimators and residuals obtained from replacing W_{it} with \hat{W}_{it} in the main specification of Autor et al. (2013). ■

By repeating Steps 1-4 we generate one thousand samples $\{Y_{it}^*, X_{it}^*, Z_{it}^*, \hat{W}_{it}\}$ on which we evaluate the performance of the overidentification test proposed in Section 4.3. In order to implement the test, we need to select the moments to employ (i.e. the functions g_j in (30)) and an estimator $\hat{\mathcal{E}}_t^*$ for the demeaned shock

$$\mathcal{E}_t^* \equiv Z_{zt}^* - E[Z_{zt}^* | \mathcal{G}_n].$$

In the main specification of Autor et al. (2013), there are no control variables with the structure required by the estimator for \mathcal{E}_t^* proposed by Borusyak et al. (2022). We therefore instead adapt the estimator of \mathcal{E}_t^* advocated by Adao et al. (2019) by employing

$$\hat{\mathcal{E}}_t^* = \left(\sum_{i=1}^n S_{zit} S'_{zit} + \lambda I_p \right)^{-1} \sum_{i=1}^n S_{zit} (Z_{it}^* - \hat{W}'_{it} \hat{\pi}_n^*) \quad (\text{S.27})$$

for $\hat{\pi}_n^*$ the coefficient obtained from a weighted regression of $\{Z_{it}^*\}$ of $\{\hat{W}_{it}\}$. We introduce ridge regression in (S.27) because the design matrix is ill-conditioned. In this regard, our

Table 4: Rejection Probabilities

Ridge Parameter	Significance Level			p-value
	1%	5%	10%	
$\lambda = 1e - 3$	0.029	0.103	0.177	0.0012
$\lambda = 1e - 4$	0.011	0.061	0.127	0.0074
$\lambda = 1e - 5$	0.008	0.049	0.114	0.0368
$\lambda = 1e - 6$	0.008	0.047	0.100	0.0650

Finite sample rejection probabilities for overidentification tests of the validity of the moment restrictions imposed when identification is driven by the exogeneity of the aggregate shocks.

estimator differs from that in [Adao et al. \(2019\)](#) who use ordinary regression (i.e. $\lambda = 0$), but instead drop sectors from the regression to address the ill conditioning of the design matrix.¹ The p-values of the test can depend on λ , and we employ the simulations to inform the choice of penalty λ for our application.²

Finally, for our moments we select the square of the residual and moments based on the pdf of the Logit distribution, which may be interpreted as different kernel estimators. Specifically, we employ a total of twenty moments by setting

$$g_j(\varepsilon, w, s) = \begin{cases} \varepsilon^2 & \text{if } j = 1 \\ \frac{\exp(\varepsilon - a_j)}{(\exp(\varepsilon - a_j) + 1)^2} & \text{if } 2 \leq j \leq 20 \end{cases}$$

with $a_2, \dots, a_{20} = -2.25, -2, \dots, 2, 2.25$. Table 4 reports the finite sample rejection probabilities of the resulting test for different choices of the ridge parameter λ . The final column of Table 4 additionally reports the p-value obtained when the test is implemented in the data of [Autor et al. \(2013\)](#). The results are based on one thousand simulations with the bootstrap implementation relying on one thousand replications. Overall, we find that the rejection probability is close to the nominal level of the test provided that the ridge parameter is sufficiently small.

References

- ADAO, R., KOLESÁR, M. and MORALES, E. (2019). Shift-share designs: Theory and inference. *The Quarterly Journal of Economics*, **134** 1949–2010.
- AUTOR, D. H., DORN, D. and HANSON, G. H. (2013). The china syndrome: Local labor market effects of import competition in the united states. *American economic review*, **103** 2121–68.
- BORUSYAK, K., HULL, P. and JARAVEL, X. (2022). Quasi-experimental shift-share research designs. *The Review of Economic Studies*, **89** 181–213.

¹See page 3 in <https://github.com/kolesarm/ShiftShareSE/blob/master/doc/ShiftShareSE.pdf>

²Analogously, as noted by [Adao et al. \(2019\)](#), dropping sectors to ensure that the design matrix is well conditioned similarly affects the resulting standard errors.

- CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2019). Inference on causal and structural parameters using many moment inequalities. *The Review of Economic Studies*, **86** 1867–1900.
- GINÉ, E., LATAŁA, R. and ZINN, J. (2000). Exponential and moment inequalities for u-statistics. In *High Dimensional Probability II*. Springer, 13–38.
- MOULTON, B. R. (1986). Random group effects and the precision of regression estimates. *Journal of econometrics*, **32** 385–397.
- VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes: with Applications to Statistics*. Springer, New York.