Working Paper Series 657
(ISSN 1211-3298)

On Policy Evaluation with Aggregate Time-Series Shocks

Dmitry Arkhangelsky
Vasily Korovkin

CERGE-EI
Prague, May 2020

ISBN 978-80-7343-464-9 (Univerzita Karlova, Centrum pro ekonomický výzkum a doktorské studium)
ISBN 978-80-7344-533-1 (Národohospodářský ústav AV ČR, v. v. i.)

On Policy Evaluation with Aggregate Time-Series Shocks *

Dmitry Arkhangelsky[†] Vasily Korovkin[‡]

May 7, 2020

Abstract

We propose a general strategy for estimating treatment effects, in contexts where the only source of exogenous variation is a sequence of aggregate time-series shocks. We start by arguing that commonly used estimation procedures tend to ignore the crucial time-series aspects of the data. Next, we develop a graphical tool and a novel test to illustrate the issues of the design using data from influential studies in development economics [Nunn and Qian, 2014] and macroeconomics [Nakamura and Steinsson, 2014]. Motivated by these studies, we construct a new estimator, which is based on the time-series model for the aggregate shock. We analyze the statistical properties of our estimator in the practically relevant case, where both cross-sectional and time-series dimensions are of similar size. Finally, to provide causal interpretation for our estimator, we analyze a new causal model that allows taking into account both rich unobserved heterogeneity in potential outcomes and unobserved aggregate shocks.

Keywords: Continuous Difference in Differences, Panel Data, Causal Effects, Treatment Effects, Unobserved Heterogeneity.

JEL Classification: C18, C21, C23, C26.

*This paper benefited greatly from our discussions with Manuel Arellano, Stéphane Bonhomme, and Guido Imbens. We also want to thank Martin Almuzara, Ivan Canay, Alexey Makarin, Eduardo Morales, and Christian Ochsner for helpful comments and suggestions. Asya Evdokimova and Gleb Kurovskiy provided excellent research assistance. Dmitry Arkhangelsky gratefully acknowledges financial support from "María de Maeztu Units of Excellence" Programme MDM-2016-0684. Vasily Korovkin gratefully acknowledges financial support from the Czech Science Foundation grant (19-25383S).

[†]CEMFI, darkhangel@cemfi.es.

[‡]CERGE-EI (a joint workplace of Charles University and the Economics Institute of the Czech Academy of Sciences), vasily.korovkin@cerge-ei.cz.

1 Introduction

Changes in aggregate variables are commonly used to evaluate economic policies. The most popular design of this type is an "event study", where a one-time aggregate shock, e.g., a new law, affects a subpopulation of units, which one observes over time. To quantify the effects of these shocks, practitioners use either difference in differences or, more recently, synthetic control methodology (e.g., Card and Krueger [1993], Abadie and Gardeazabal [2003], Abadie et al. [2010]). Often, when both outcome and treatment variable are unit-specific, this approach is used as a first stage, and the aggregate change effectively plays the role of an instrument. In the absence of a single aggregate shock, researchers often employ more general time-series variation to establish causal links between unit-specific policy and outcome variables. In a typical application, outcomes are observed at some geographical level over time (e.g., Dube and Vargas [2013], Nakamura and Steinsson [2014], and Nunn and Qian [2014]). To address potential endogeneity problems, researchers use aggregate time-series shocks as instruments. A standard econometric tool employed to analyze such data is a two-stage least-squares (TSLS) regression with unit and time fixed effects (see Arellano [2003] for a textbook treatment).

In this paper, we propose an algorithm that can be used to answer causal questions using exogenous aggregate shocks. In other words, our strategy can be applied whenever the sole source of exogenous variation comes from the time-series dimension. Our novel approach has three key steps: first, researchers should report a simple graphical representation of the data, along with a formal test. Next, researchers should construct a time-series model for the aggregate shock. Finally, using the information from the first two steps, researchers should estimate a particular panel model. We show that this strategy, once implemented correctly, delivers meaningful policy-relevant parameters under flexible assumptions about the underlying causal model.

In particular, let Y_{it} be the outcome variable, W_{it} the endogenous regressor, and Z_t the aggregate instrument. To establish a causal link between Y_{it} and W_{it}, the following regression is estimated by TSLS:

$$Y_{it} = \alpha_i + \theta_t + \tau W_{it} + \epsilon_{it}, \tag{1.1}$$

using $D_i Z_t$ as an instrument. Here, D_i is a measure of "exposure" of unit i to the aggregate shock, and τ is the parameter of interest. For example, in Nunn and Qian [2014] Z_t is the amount of wheat produced in the United States in the previous year, W_{it} is the amount of food aid that country i received, and Y_{it} is a measure of local conflict. In this example, D_i is not directly available, but instead is constructed for each country using the whole vector $(W_{i1}, \ldots, W_{iT})$.

The TSLS algorithm uses D_i to aggregate outcomes over units, and it is straightforward to show that $\hat{\tau}_{TSLS}$ is a ratio of two OLS coefficients in the following time-series regressions:

$$
\begin{aligned}
Y_t &= \alpha + \delta Z_t + \epsilon_t \\
W_t &= \beta + \pi Z_t + u_t,
\end{aligned}
\tag{1.2}
$$

where W_t and Y_t are weighted averages of W_{it} and Y_{it}, with the weights proportional to $D_i - \mathbb{E}[D_i]$. These weights average to zero across units to balance away time fixed effects.

Representation (1.2) is the starting point of our analysis. Properties of the aggregate errors (ϵ_t, u_t) depend crucially on the type of moment restrictions that one is willing to assume. These restrictions are rarely explicitly stated in applied papers, so we start with a cross-sectional model that justifies commonly used estimation and inference procedures. In this model, (1.2) effectively reduces to the following error-free time-series model:

$$
\begin{aligned}
Y_t &\approx \alpha + \delta Z_t \\
W_t &\approx \beta + \pi Z_t.
\end{aligned}
\tag{1.3}
$$

Importantly, this is true even in cases where the number of periods is large, e.g., T is of the same order as n. This representation suggests that we can simply plot Y_t and W_t versus Z_t as functions of time to check whether (1.3) holds. In fact, this is equivalent to the standard event-study plot if $Z_t = \{t \geq T_{treat}\}$ is a one-time aggregate shock. In addition to this graphical evidence, we construct a formal test that can be used to find out whether (1.3) holds. Our test is straightforward to implement and is consistent both in the regime where T is small relative to n and when T has the same magnitude as n. We apply these tools to two empirical applications and find compelling evidence against (1.3). While this does not necessarily invalidate the analysis, it makes it more subtle; in particular, time-series properties of Z_t start to play an important role. These results constitute the first contribution of this paper.

If data rejects (1.3), then it is natural to ask whether something can still be done using this design, or whether one has to resort to a different identification strategy altogether. At this point, we go back to (1.2) and argue that if ε_t and u_t are not negligible, then we need to understand what precisely these errors represent. For this purpose, we consider a more general econometric model that combines both time-series and cross-sectional aspects of the data. In this model, we allow D_i to be unobserved and construct it using the data. We also explicitly model the process for Z_t and show how its properties should be used in estimation. Our algorithm can be implemented as the following TSLS regression:

$$Y_{it} = \alpha_i + \eta_i^\top \psi(t) + \theta_t + \tau W_{it} + \epsilon_{it}, \tag{1.4}$$

with $\tilde{D}_i Z_t$ as an instrument. Here, $\psi(t)$ are user-specified functions of time, such as trends or cycles, and $\tilde{D}_i$ is the constructed measure of exposure. Regression (1.4) is done on a subsample of the data that is not used for constructing $\tilde{D}_i$. We analyze the properties of this estimator in the high-dimensional regime where T and n can be of the same order. This is important because in the empirical applications n is of the same magnitude as T (e.g., $n = 51$ and $T = 40$). We also prove that the distribution of the estimator can be uniformly approximated by a normal distribution. These statistical results constitute the second contribution of this paper.

Certain aspects of our econometric model are restrictive. There is no heterogeneity in treatment effects, and errors have a particular structure. As a result, one might worry that the estimator does not have a general causal interpretation under misspecification and is only applicable in a narrow case. To address this issue, we construct a causal model and provide conditions under which our procedure identifies a convex combination of treatment effects. We also use this model to clarify the sources of endogeneity that researchers need to address to conduct valid causal inference. Our model emphasizes different roles that unit weights and aggregate shocks Z_t play for identification. Unit weights are used to integrate the individual outcomes in such a way that the aggregates are still affected by Z_t, but are not exposed to other potential confounders. After the aggregation, the problem reduces to a time-series IV regression with Z_t as an instrument. At this point, time-series properties of Z_t start playing a crucial role. This model and related identification results constitute the third contribution of this paper.

Finally, we compare our algorithm for unit weights with those currently used in empirical practice. We show that commonly used procedures might perform poorly in applications with a moderate number of periods, especially if the instruments are weak. Even with an entirely irrelevant instrument Z_t, the TSLS estimator converges to a concentrated limit. In practice, this means that one would see small confidence intervals and strong first-stage results even with irrelevant instruments. The combination of a small number of periods and an almost perfect first stage is a manifestation of overfitting, and thus using sample-splitting methods of the type that we are proposing is natural in this context. We confirm this intuition in a series of Monte Carlo simulations that are based on the data from Nunn and Qian [2014]. In particular, we show that the estimator that uses the full sample is biased and has asymmetric distribution. In contrast, our estimator performs well both in terms of bias and shape.

Our model is related to shift-share designs (or "Bartik" instruments, after Bartik [1991]), but this relationship is more formal than conceptual. Standard application with the shift-share design has both an outcome and an endogenous treatment that are measured on a location level and instruments that are measured on an industry level. Using industry-specific weights, researchers transform endogenous variables to industry levels and then use IV for estimation (see, e.g., Adão et al. [2018], Borusyak et al. [2018], and Goldsmith-Pinkham et al. [2018] for discussion and details). In our model, the outcome and treatment for every unit, as well as the aggregate shocks, are observed over time. As a result, researchers do not need to transform the outcomes to construct a standard IV estimator. The motivation for aggregating the data is different — researchers are worried about potential unobserved aggregate shocks that are correlated with the instrument.

The paper proceeds as follows: in Section 2, we consider a cross-sectional model that is motivated by the applied work, clarify exactly what (1.3) means, construct a formal test, and use it for two empirical applications. In Section 3, we present a general econometric model, construct a new estimator, and analyze its properties. In Section 4, we present a theoretical causal model and derive identification results. In Section 5, we discuss how our approach is related to current empirical practice, compare our causal model to the recent literature on shift-share designs, and show the results from the Monte Carlo simulations. Section 6 concludes.

2 Cross-sectional Approach

In this section, we consider a cross-sectional model that is motivated by estimation and inference algorithms currently used in empirical work. We argue that this model implies a restrictive time-series representation of the data, and we show that it can be directly tested. We look at two empirical applications and find strong evidence refuting the cross-sectional model. To us, these results indicate that the time-series dimension should play a central role in identification, estimation, and inference.

2.1 Econometric Model

A common algorithm for estimating causal effects with aggregate shocks is a TSLS regression:

$$Y_{it} = \tilde{\alpha}_i + \tilde{\theta}_t + \tau W_{it} + \varepsilon_{it}, \tag{2.1}$$

with $D_i Z_t$ as an instrument. Here W_{it} is the policy variable of interest, Y_{it} is the outcome, Z_t is the aggregate shock (instrument) and D_i measures the exposure of unit i to Z_t. In practice, D_i is usually a function of $\{(W_{it}, Z_t)\}_{t \leq T}$, but in this section we abstract away from this and simply treat it as given. The object of interest is τ, and $\hat{\tau}_{TSLS}$ is the TSLS estimator.

Regression 2.1 can be split into two parts – the reduced form and the first stage:

$$\begin{aligned} Y_{it} &= \alpha_i + \theta_t + \delta D_i Z_t + \epsilon_{it} \\ W_{it} &= \beta_i + \gamma_t + \pi D_i Z_t + u_{it}, \end{aligned} \tag{2.2}$$

where $\alpha_i := \tilde{\alpha}_i + \tau\beta_i$, $\theta_t := \tilde{\theta}_t + \delta\gamma_t$, and $\delta = \tau\pi$. Standard logic implies that a TSLS estimator is a ratio of two OLS estimators:

$$\hat{\tau}_{TSLS} = \frac{\hat{\delta}_{OLS}}{\hat{\pi}_{OLS}}. \tag{2.3}$$

Alternatively, one can represent the same estimator as a ratio of two OLS estimators from the following time-series regressions:

$$\begin{aligned} Y_t &= \alpha + \delta Z_t + \check{\epsilon}_t \\ W_t &= \beta + \pi Z_t + \check{u}_t, \end{aligned} \tag{2.4}$$

where the aggregate variables are defined below:

$$
\begin{aligned}
&Y_t := \frac{1}{n}\sum_{i\le n} Y_{it}\left(\frac{D_i-\overline{D}}{\hat{\mathbb{V}}[D_i]}\right), \quad && W_t := \frac{1}{n}\sum_{i\le n} W_{it}\left(\frac{D_i-\overline{D}}{\hat{\mathbb{V}}[D_i]}\right),\\
&\check{\epsilon}_t := \frac{1}{n}\sum_{i\le n} \epsilon_{it}\left(\frac{D_i-\overline{D}}{\hat{\mathbb{V}}[D_i]}\right), && \check{u}_t := \frac{1}{n}\sum_{i\le n} u_{it}\left(\frac{D_i-\overline{D}}{\hat{\mathbb{V}}[D_i]}\right),\\
&\alpha := \frac{1}{n}\sum_{i\le n} \alpha_i\left(\frac{D_i-\overline{D}}{\hat{\mathbb{V}}[D_i]}\right), && \beta := \frac{1}{n}\sum_{i\le n} \beta_i\left(\frac{D_i-\overline{D}}{\hat{\mathbb{V}}[D_i]}\right).
\end{aligned}
\tag{2.5}
$$

We aggregate Y_{it}, W_{it} with the weights that sum up to zero, and thus time fixed effects are not present in (2.4).

To understand the properties of $\hat{\delta}_{OLS}, \hat{\pi}_{OLS}$, and by extension $\hat{\tau}_{TSLS}$, we need to make assumptions about the underlying statistical model. Let $\mathcal{D}_i := \{(Y_{i1},\dots,Y_{iT}),(W_{i1},\dots,W_{iT}),D_i\}$ – all data available for unit i. We make the following assumption:

Assumption 2.1. (Cross-sectional Model) *Outcomes are generated by the following model:*

$$
\begin{aligned}
&Y_{it} = \alpha_i + \theta_t + \delta D_i Z_t + \epsilon_{it},\\
&W_{it} = \beta_i + \gamma_t + \pi D_i Z_t + u_{it},\\
&\mathbb{E}\left[\begin{pmatrix}1\\ D_i\end{pmatrix}(u_{it},\epsilon_{it})|\{Z_l\}_{l\le T}\right] = 0,
\end{aligned}
\tag{2.6}
$$

and, conditionally on $\{Z_l\}_{l\le T}$, individual-level variables $\{\mathcal{D}_i\}_{i=1}^n$ are i.i.d.

We call this model cross-sectional, because it imposes cross-sectional restrictions on the errors conditionally on the whole sequence $\{Z_t\}_{t\le T}$. It is easy to see that in this model parameters δ and π are identified with just two periods as long as there is any variation in Z_t. This model is a direct generalization of the standard difference in differences (two-way fixed effects) model that is routinely used in applied work. There is no heterogeneity in τ and π, which is a clear restriction, because in practice we expect treatment effect to vary across units. We make this assumption for expositional purposes, and as long as heterogeneity is cross-sectional, it does not affect the results that we present below. Of course, the interpretation of $\hat{\tau}_{TSLS}$ changes.

To derive statistical results, we need to impose restrictions on the error terms. The structure of the errors is motivated by our causal model in Section 4. In addition to this assumption we impose a technical condition on the behavior of the tails of the relevant errors, which we state formally in Appendix A.

Assumption 2.2. (Factor errors) *The errors have a factor representation:*

$$\begin{aligned} \epsilon_{it} &= \mu_{\epsilon,t}\nu_i^{(1)} + \nu_{\epsilon,it}^{(2)}, \\ u_{it} &= \mu_{u,t}\nu_i^{(1)} + \nu_{u,it}^{(2)}, \end{aligned} \tag{2.7}$$

where $\mu_{\epsilon,t}, \mu_{u,t} \in \mathbb{R}^{2k}$. *Also* $(\nu_i^{(1)}, D_i)$ *is independent of* $(\nu_{\epsilon,it}^{(2)}, \nu_{u,it}^{(2)})_{t\leq T}$

In the model (2.6), there are $2T + 4$ parameters (time fixed effects, (π, δ), and projections of α_i, β_i on D_i) and $4T$ moment conditions that we can use to estimate them. As a result, there are $2(T-2)$ moment restrictions that can be used to test the model. Representation (2.4) can be used to visualize these restrictions: as we show below under current assumptions $\check{\epsilon}_t, \check{u}_t$ essentially scale as $\frac{1}{\sqrt{n}}$ and thus should be negligible as long as n is sufficiently large. As a result, a simple time-series plot of Y_t and W_t versus their OLS projection on Z_t and a constant can be used as direct graphical evidence to support the model (2.6). In fact, it is a straightforward generalization of standard "parallel trends" plots that are routinely reported in the applied work. We summarize this logic in the following proposition:

Proposition 2.1. (Graphical evidence) *Suppose Assumptions 2.1,2.2, A.1, A.2, A.3 hold. Then the following is true with the probability at least* $1 - \frac{1}{n}$

$$\begin{aligned} \max_t |\check{u}_t| &\lesssim \frac{1}{\mathbb{V}[D_i]}\left(\max_t \|\mu_{u,t}\|_2\sqrt{\frac{\log(nk)}{n}} + \sqrt{\frac{\log(nT)}{n}}\right) \\ \max_t |\check{\epsilon}_t| &\lesssim \frac{1}{\mathbb{V}[D_i]}\left(\max_t \|\mu_{\epsilon,t}\|_2\sqrt{\frac{\log(nk)}{n}} + \sqrt{\frac{\log(nT)}{n}}\right), \end{aligned} \tag{2.8}$$

This proposition shows that $\check{u}_t, \check{\epsilon}_t$ converge to zero uniformly as n goes to infinity. Crucially, this still holds even if T is large (e.g., of the order of n) as long as $\|\mu_{u,t}\|_2, \|\mu_{\epsilon,t}\|_2$ are bounded (over t). From the practical point of view, this says that as long as n is large and factors are not extremely strong, then the aggregate errors in (2.4) are negligible.

2.2 Testing

Plotting Y_t, and W_t versus Z_t is a simple, direct way of assessing (2.6). It is not entirely satisfactory, because in practice we need to have a sense of the size of the error that we can tolerate before rejecting the model (2.6). Proposition 2.1 quantifies this error but does not tell us how we can compute it in practice. To address this issue, we construct a formal test for moment restrictions (2.6). Specifically, we test $\mathbb{E}[\nu_i^{(1)} D_i] = 0$ versus $\|\mathbb{E}[\nu_i^{(1)} D_i]\|_2 > 0$. Our results allow for an empirically relevant case where T is of the same order as n. In this section, we describe the testing procedure informally, leaving the details for Appendix B.1.

We are testing that $\mathbb{E}[\nu_i^{(1)} D_i] = 0$ and $\nu_i^{(1)}$ is present in the model only if $\mu_{\epsilon,t}, \mu_{u,t}$ are not equal to zero. As a result, we need to restrict their size. Let $\mu_\epsilon^\top := (\mu_{\epsilon,1}, \dots, \mu_{\epsilon,T})$, and similarly define μ_u. Let $\mathbf{Z}$ be $T \times 2$ matrix, such that $(\mathbf{Z})_t = (1, Z_t)$ and define $\mathbf{P}_{Z^\perp} := \mathcal{I}_T - \mathbf{Z}\left(\mathbf{Z}^\top \mathbf{Z}\right)^{-1} \mathbf{Z}^\top$. We make the following assumption regarding the strength of the factors:

Assumption 2.3. (Visible factors) *Factors satisfy the following restriction:*

$$\min\{\|\mathbf{P}_{Z^\perp}\mu_\epsilon\|, \|\mathbf{P}_{Z^\perp}\mu_u\|\} \geq \underline{c}_\mu \sqrt{T}. \tag{2.9}$$

This restriction is satisfied if $\mu_{\epsilon,t}, \mu_{u,t}$ are functions of time that are not fully predictable by Z_t. Alternatively, it is satisfied with high probability if $\mu_{\epsilon,t}, \mu_{u,t}$ are realizations of a stationary time-series.

Our test focuses on the sample versions of (2.6). We estimate all the parameters in the model (2.2) by OLS and construct the residuals:

$$\begin{aligned} \hat{\epsilon}_{it} &:= Y_{it} - \hat{\alpha}_{i,OLS} - \hat{\theta}_{t,OLS} - \hat{\delta}_{OLS} D_i Z_t \\ \hat{u}_{it} &:= W_{it} - \hat{\beta}_{i,OLS} - \hat{\gamma}_{t,OLS} - \hat{\pi}_{OLS} D_i Z_t. \end{aligned} \tag{2.10}$$

Let $\hat{\epsilon}_i := (\hat{\epsilon}_{i1}, \dots, \hat{\epsilon}_{iT})$, $\hat{u}_i := (\hat{u}_{i1}, \dots, \hat{u}_{iT})$, and define the following random vectors:

$$\begin{aligned} \hat{\xi}_\epsilon &:= \frac{1}{n} \sum_{i \leq n} (D_i - \overline{D}) \hat{\epsilon}_i \\ \hat{\xi}_u &:= \frac{1}{n} \sum_{i \leq n} (D_i - \overline{D}) \hat{u}_i. \end{aligned} \tag{2.11}$$

Under the null hypothesis that moment restrictions in (2.6) hold, these aggregated residuals should be close to zero. Our test is based on the scaled norm of this object:

$$\hat{L}_n(\alpha) := \sqrt{\frac{n}{(T-2)}}\sqrt{\alpha\|\hat{\xi}_\epsilon\|_2^2 + (1-\alpha)\|\hat{\xi}_u\|_2^2}, \tag{2.12}$$

where $\alpha \in [0,1]$ is a user-specified parameter. In practice, Y_{it} and W_{it} can have very different scales, and thus it is natural to use either $\alpha \in \{0,1\}$, or $\alpha = \frac{\hat{\mathbb{V}}[W_{it}]}{\hat{\mathbb{V}}[W_{it}]+\hat{\mathbb{V}}[Y_{it}]}$. To compute the distribution of this statistic under the null, we use a wild bootstrap scheme, which we describe in detail in Appendix B.2. Let $\hat{L}_n^{(b)}(\alpha)$ be the bootstrap version of the test statistic and define its bootstrap distribution:

$$F_{\hat{L}^{(b)},\alpha}(x) := \mathbb{E}^{(b)}[\{\hat{L}_n^{(b)}(\alpha) \le x\}]. \tag{2.13}$$

The following theorem compares the properties of $F_{\hat{L}^{(b)},\alpha}(x)$ with the distribution $F_{\hat{L},\alpha}(x)$ of $\hat{L}_n(\alpha)$:

Theorem 2.2. (Test for cross-sectional model) *Suppose Assumptions 2.1, 2.2,2.3, A.1, A.2, A.3 hold, and rank*$(\mathbf{Z}) = 2$. *In addition, suppose that either both T and k are fixed, or* $\frac{\log(n)\sqrt{\log(n)+k}}{\sqrt{T}} = o(1)$ *and* $\frac{k^{\frac{7}{4}}\log^3(n)}{\sqrt{n}} = o(1)$. *Then for any $\alpha \in [0,1]$, the following is true:*

$$\sup_x \left| F_{\hat{L}^{(b)},\alpha}(x) - F_{\hat{L},\alpha}(x)\right| = o_p(1). \tag{2.14}$$

This theorem describes two different regimes. In the first one, T is fixed and the fact that bootstrap works is expected. At the same time, in our empirical applications n is at most three times larger than n, so the second regime is more relevant. In the second regime, T can be as large as n (or even larger). One could be surprised that bootstrap works in this regime, because it is known that standard central limit theorem approximation fails (e.g., Belloni et al. [2018]). Here things are different because the factors dominate the remaining part of the error.

We can construct a p-value:

$$\hat{p} := 1 - F_{\hat{L}^{(b)},\alpha}(\hat{L}_n) \tag{2.15}$$

and use it to evaluate the strength of evidence against the cross-sectional model. If $\hat{p}$ is close to zero (say below 10%), then Assumption 2.1 is questionable, and thus alternative models should be considered for estimation and inference. Formally, we have the following result:

Proposition 2.3. (Consistency) *Suppose all the assumptions of Theorem 2.2 are satisfied, except* $\|\mathbb{E}[\nu_i^{(1)} D_i]\| > 0$*. Then* $\hat{p} = o_p(1)$*.*

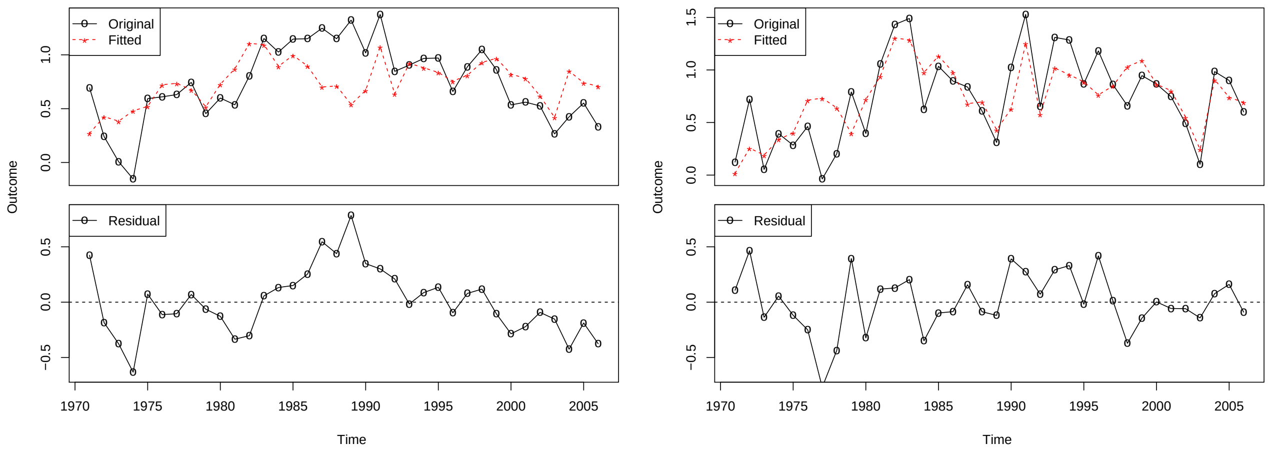
We do not formally investigate local power properties of our test, but heuristically the size of alternatives that are hard to detect scales as $\frac{1}{\sqrt{n}}$. Under current assumptions, the standard errors of $(\hat{\delta}_{OLS}, \hat{\pi}_{OLS})$ scale as $\frac{1}{\sqrt{n}}$, so it is natural to consider local alternatives of this order. As long as OLS results are available, our test is extremely simple to construct, it does not require any additional estimation, and it can be used to quantify the evidence against the cross-sectional model.

2.3 Empirical Illustrations

In this section, we give two examples from the empirical literature, which use aggregate variables as instruments. We illustrate our point with graphical evidence, and then apply our test from the previous subsection.

The first example is based on Nunn and Qian [2014] — a study of causal effect of foreign aid on the onset and length of civil conflicts. The authors use U.S. wheat production as a time-series instrument Z_t for the U.S. foreign aid W_{it}. The foreign-aid example allows us to plot the graphs that correspond to (1.3) — an aggregate representation of the model. We discover that the behavior of the aggregate (Y_t, W_t) series is not the same as suggested by (1.3), or, more formally, it is not in line with Proposition 2.1. Motivated by this fact, we examine the behavior of residuals' vectors $(\hat{\xi}_\epsilon, \hat{\xi}_u)$. Namely, our bootstrap procedure finds strong evidence against the null hypothesis, and hence against the aggregate model (1.3). Overall, the combination of graphical evidence and formal testing suggests that time-series aspects play an important role in this example and should not be ignored for purposes of identification, estimation, and inference.

The second example comes from the Nakamura and Steinsson [2014] study of fiscal multipliers in the U.S. economy. The aggregate instrument Z_t, in this case, is fluctuation in federal military buildups. Regional variation in the military share of GDP allows running TSLS the way we discussed in the Introduction. We point out that both our graphical approach and our formal test are highly in favor of rejecting the aggregate model (1.3). In other words, time-series aspects are of the first order in this application as well.



Panel A: Observed Data Panel B: Simulated Data

Figure 1: Aggregated Time Series Y_t, $\hat{Y}_t$ Along with $\hat{\xi}_\epsilon$, Nunn and Qian [2014]

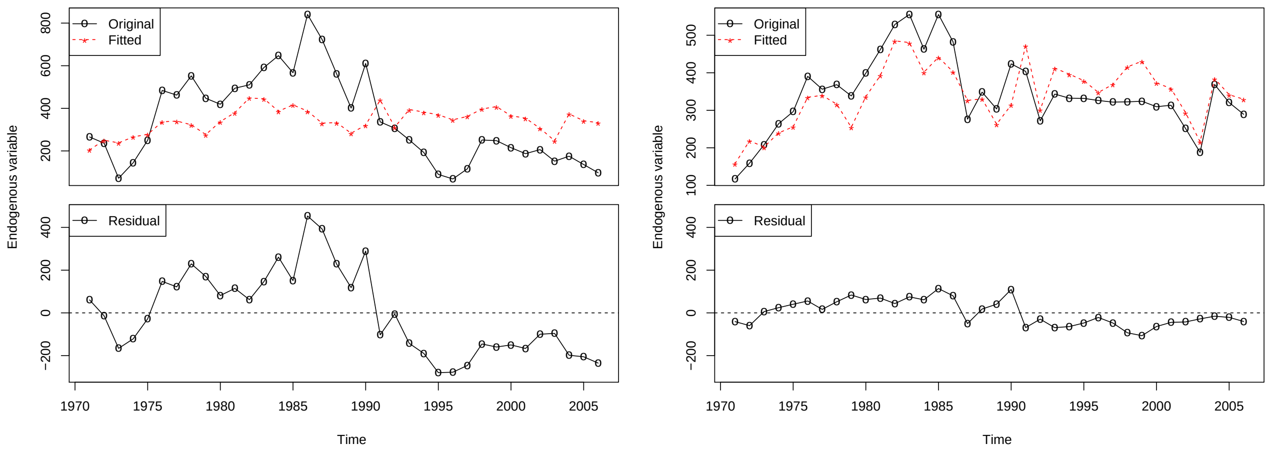
Notes: Panel A shows the aggregated series from the observed data. Y_t is defined in equation (2.4), and $\hat{Y}_t$ is a predicted value from a regression of Y_t on Z_t. The lower graph in Panel A portrays ξ_ϵ as defined in (2.10). Panel B shows data, simulated from $Y_t = \alpha + \delta Z_t + \epsilon_t$, where ϵ_t satisfies Proposition 2.1.

2.3.1 Foreign Aid and Civil Conflict

In this subsection, we turn to studies in development economics that link endowment shocks to the onset and length of civil conflicts. Frequently, these studies use instrumental variables to achieve causal interpretation of the results [Miguel et al., 2004]. A popular approach is to use aggregate fluctuations such as changes in commodity prices [Dube and Vargas, 2013] or changes in outside production [Nunn and Qian, 2014] as a source of exogenous variation.

Specifically, Nunn and Qian [2014] examine how U.S. food aid affects conflict in non-OECD recipient countries. The primary motivation is a long-lasting debate among aid workers on whether food aid provides relief for populations in poverty or leads to negative spillovers, and whether those spillovers can be large enough to defeat the purpose of aid. Potentially the main such spillover is an increased probability and length of civil conflict due to aid looting and fighting over it.

Nunn and Qian [2014] develop a strategy to estimate the effect of U.S. wheat aid to non-OECD countries using time variation in U.S. wheat production, which is arguably exogenous to the conditions in the recipient country. A fundamental feature of the setting is that while U.S. aid varies in cross-section and in time, wheat production fluctuates only over time. Hence, the authors need to interact aid intensity with cross-sectional variation in the country's likelihood



Panel A: Observed Data

Panel B: Simulated Data

Figure 2: Aggregated Time Series W_t, $\hat{W}_t$ Along with $\hat{\xi}_u$, Nunn and Qian [2014]

Notes: Panel A shows the aggregated series from the observed data. W_t is defined in equation (2.4), and $\hat{W}_t$ is a predicted value from a regression of W_t on Z_t. The lower graph of Panel A portrays ξ_u as defined in (2.10). Panel B shows data simulated from $W_t = \beta + \pi Z_t + u_t$, where u_t satisfies Proposition 2.1.

of being a recipient to construct a meaningful instrument for aid.

Formally, a generic cross-sectional observation i is a country, while t is a year. Y_{it} are various conflict-related outcomes: incidence, duration, and the onset of both interstate and civil conflicts. For this section, we concentrate on the definition of conflict as an indicator that equals one if there is conflict in country i at year t, similar to the baseline specification of Nunn and Qian [2014].

Next, we denote W_{it} as the quantity of wheat aid shipped from the United States to recipient country i in a year t. The main instrument in the original paper, which we denote as Z_t is the amount of U.S. wheat production in the previous year, and it varies only over time. What the authors call an "instrument" instead is a variable that varies both by time and in a cross-section: $Z_{t-1} \times \bar{D}_i$, where $\bar{D}_i = (1/36)\sum_{t=1971}^{2006}\{W_{it} > 0\}$. That is, $\bar{D}_i$ is the fraction of years in which the country received the U.S. food aid in the whole study period.

As the main exercise of this subsection, we follow (1.3) and Proposition 2.1 and plot the outcome Y_t together with its predicted value $\hat{Y}_t$ f, and residuals $\hat{\xi}_\epsilon$. Then, we repeat the exercise for the first stage and W_t instead of Y_t. We show simulated data as a benchmark, where ϵ_{it} and u_{it} are drawn so they satisfy Assumption 2.1, while Z_t is taken from the data. For parameters in simulations, we use the estimates from the original paper (see Section 5.3 for details).

Figure 1 shows the results for Y_t. Panel A plots the observed data, while Panel B shows the simulated data. The deviations between Y_t and $\hat{Y}_t$ in the observed data are noticeably larger than in the simulated data. By definition, it translates into the behavior of residuals, as depicted in the lower graph of Panels A and B. Specifically, the fluctuation in simulated errors is smaller than in the observed data.

The same is true if we turn to Figure 2 and show the behavior of the endogenous variable in the observed data and the simulated data. Again, for the simulated data, the deviations of W_t and $\hat{W}_t$ are smaller, and so are the residuals. Moreover, the deviations of simulated errors for the first stage are also lower than in the simulated reduced form. The reason for it is that Y_{it} is binary, and the two-way fixed-effect model does not give a good fit, even with simulated data.

A natural next step is to apply our bootstrap procedure and examine the p-values for the null that the aggregate model is the correct one. We run our bootstrap procedure separately for the reduced form and the first stage. The p-values are reported in the first two columns of Table 1, Panel A. The p-values in both cases are quite low, which is highly suggestive of rejecting the null hypothesis that Assumption 2.1 holds. The p-value for the first stage is specifically low, which is in line with the patterns from Panels A of Figures 1 and 2. The fit for the first stage is much worse in Figure 2 than in Figure 1.

Using bootstrap only for the reduced form or first stage corresponds to picking extreme values of the weight α in equation 2.10. One can ask what happens if we pick an intermediate value of α. We report the resulting p-value in the last column of Panel A, Table 1. The p-value of 0.017 is lower than both of the p-values for the reduced form and for the first stage. Overall, our bootstrap exercise suggests that one should switch to a more general model. We specify and discuss this model in Section 3.

2.3.2 Government Spending Multipliers

In this subsection, we replicate an influential study by Nakamura and Steinsson [2014] on the fiscal multiplier for the U.S. economy. A crucial feature of the authors' setting is that they are interested in an *open economy relative multiplier*. That is, they compare different U.S. states, and study their reaction to fluctuations in aggregate military spending in a panel setting. This strategy allows them to control for common shocks such as monetary policy. It also allows them to account for potential endogeneity of local procurement spending.

Table 1: Empirical Illustrations: Bootstrap

	Reduced form Y_t	First-Stage: W_t	Weighted
	$\alpha = 1$	$\alpha = 0$	$\alpha = 0.5$
Panel A: Nunn and Qian [2014]			
p-value	0.185	0.025	0.017
n	98	98	98
T	36	36	36
Panel B: Nakamura and Steinsson [2014]			
p-value	0.073	0.187	0.066
n	51	51	51
T	39	39	39

Notes: p-values are calculated based on $B = 100,000$ bootstrap repetitions. Panel A: Y_{it} is conflict incidence in country i at year t, W_{it} is U.S. wheat aid to country i at year t. Panel B: Y_{it} is two-year state GDP growth, W_{it} is two-year growth in per capita military procurement spending in state i in year t, normalized by output. The last column corresponds to $\hat{p} = 1 - F_{\hat{L}^{(b)},\alpha}(\hat{L}_n)$, where $\hat{L}_n(\alpha) := \sqrt{n/(T-2)}\sqrt{\alpha\|\hat{\xi}_\epsilon\|_2^2 + (1-\alpha)\|\hat{\xi}_u\|_2^2}$ with $\alpha = 0.5$. We use standardized outcomes and treatments Y_{it}, W_{it}, but it only matters for the p-value in the case of $\alpha = 0.5$, and is identical to the non-standardized variables case in the first two columns.

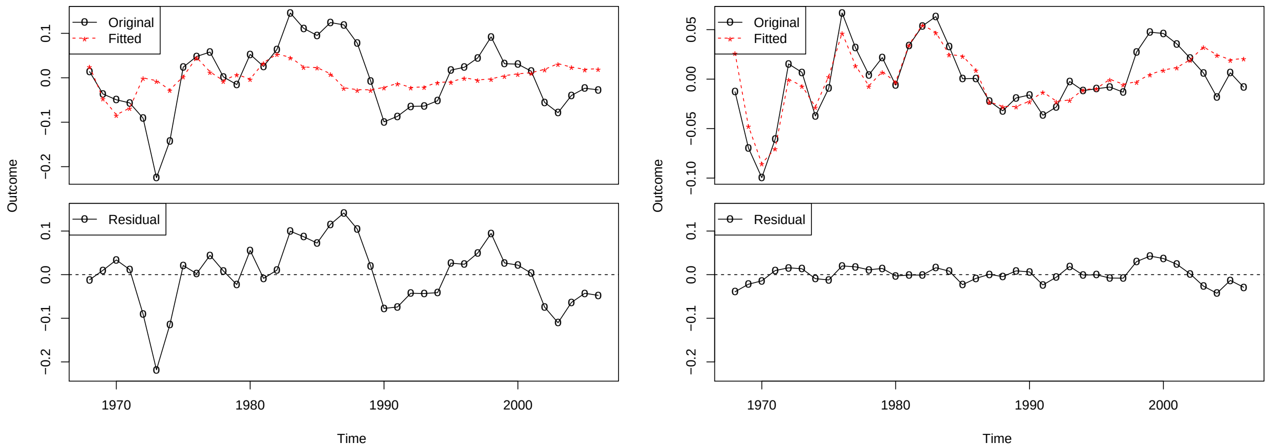
In their baseline specification, Nakamura and Steinsson [2014] interact state fixed effects with the fluctuations in aggregate military spending and use this interaction as an IV for state-level military procurement. To illustrate their approach, we introduce some notations. For a generic observation — a state — i, and a generic time period t, we denote per capita output growth in state i from year $t-2$ to t by Y_{it}.[1] Similarly, we denote two-year growth in per capita military procurement spending in state i in year t, normalized by output, in year $t-2$, by W_{it}. Finally, we use Z_t changes in total national procurement from year $t-2$ to t as an aggregate instrument. Our main empirical exercise in this paper is described by the following model:

$$\begin{aligned} Y_{it} &= \alpha_i + \theta_t + \tau W_{it} + \epsilon_{it} \\ W_{it} &= \beta_i + \gamma_t + \pi_i Z_t + u_{it}, \end{aligned} \tag{2.16}$$

where $(\alpha_i, \beta_i, \theta_t, \gamma_t, \pi_i)$ are fixed effects. This strategy is equivalent to using unit weights of a specific form and constructing aggregate variables. However, as we argue in more detail in Section 5, these weights yield inconsistent estimators for any fixed T.

We draw plots that are similar to Figures 1 and 2. Namely, Figure 3 shows the aggregated outcome and the fitted values for observed and simulated data together with the residuals $\hat{\xi}_\epsilon$.

[1] The authors advocate for using two-year changes instead of one-year changes together with leads and lags by arguing that a fully dynamic specification does not change the results much.



Panel A: Observed Data

Panel B: Simulated Data

Figure 3: Aggregated Time Series Y_t, $\hat{Y}_t$ Along with $\hat{\xi}_\epsilon$, Nakamura and Steinsson [2014]

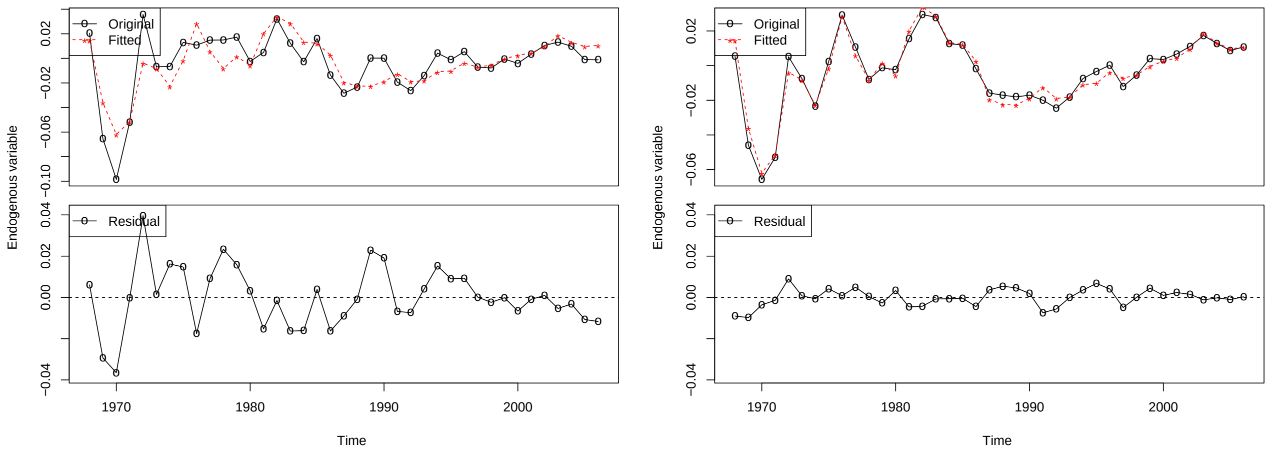
Notes: Panel A shows the aggregated series from the observed data. Y_t is defined in equation (2.4), and $\hat{Y}_t$ is a predicted value from a regression of Y_t on Z_t. The lower graph in Panel A portrays ξ_ϵ as defined in (2.10). Panel B shows data, simulated from $Y_t = \alpha + \delta Z_t + \epsilon_t$, where ϵ_t satisfies Proposition 2.1.

As before, the residuals for the observed data are much larger than for the simulations. The fit is much better in the first stage plotted in Panel A of Figure 4, which is the opposite of the Nunn and Qian [2014] example considered above. However, even for the first stage, one can conjecture that the fit is drastically different than in the simulated data based on Assumption 2.1. We rerun the bootstrap test and report the results in Panel B of Table 1.

Panel B of Table 1 shows that the p-values are low — all of them are lower than 0.2. Indeed, the p-value for the first stage is now larger, which confirms the graphical evidence. The p-value 0.066 from the weighted combination of residuals is again lower than both of the p-values for the extreme cases.

3 Joint Approach

In this section, we consider a general econometric model that combines time-series and cross-sectional aspects. The model is motivated by the fact that in empirical applications, there is strong evidence that time dimension plays an important role. We also address the question that was ignored in Section 2: where do unit exposures D_i come from? As we showed in Section 2.3, the answer to this question varies across applications. We propose a unified approach and



Panel A: Observed Data Panel B: Simulated Data

Figure 4: Aggregated Time Series W_t, $\hat{W}_t$ Along with $\hat{\xi}_u$, Nakamura and Steinsson [2014]

Notes: Panel A shows the aggregated series from the observed data. W_t is defined in equation (2.4), and $\hat{W}_t$ is a predicted value from a regression of W_t on Z_t. The lower graph of Panel A portrays ξ_u as defined in (2.10). Panel B shows data simulated from $W_t = \beta + \pi Z_t + u_t$, where u_t satisfies Proposition 2.1.

describe the assumptions under which it leads to consistent estimation.

3.1 Econometric Model

We start by specifying a model for the unit-level outcomes and treatments:

Assumption 3.1. (General model) *For each unit i outcomes are generated in the following way:*

$$\begin{aligned} Y_{it} &= \alpha_i + \theta_t + \tau \pi_i Z_t + \epsilon_{it}, \\ W_{it} &= \beta_i + \gamma_t + \pi_i Z_t + u_{it}, \end{aligned} \tag{3.1}$$

where we observe (Y_{it}, W_{it}, Z_t). Errors ϵ_{it}, u_{it} have the following factor structure:

$$\begin{aligned} \epsilon_{it} &= \mu_{\epsilon,t}^\top \nu_i^{(1)} + H_t \nu_{\epsilon,it}^{(2)}, \\ u_{it} &= \mu_{u,t}^\top \nu_i^{(1)} + H_t \nu_{u,it}^{(2)} \\ \mathbb{E}[\nu_i^{(1)}] &= 0, \\ \mathbb{E}[(\nu_{\epsilon,it}^{(2)}, \nu_{u,it}^{(2)})] &= 0, \end{aligned} \tag{3.2}$$

where $\nu_i^{(1)} \in \mathbb{R}^{2k}$*, and* $(\nu_i^{(1)}, \pi_i)$ *is independent of* $\{\nu_{\epsilon,it}^{(2)}, \nu_{u,it}^{(2)}\}_{t\leq T}$*. Sequence* $\{\alpha_i, \beta_i, \pi_i, \nu_i^{(1)}, \nu_i^{(2)}\}_{i=1}^n$ *is i.i.d. and is independent of* $\{H_t, Z_t\}_{t\leq T}$*.*

This structure is different from the one specified by Assumptions 2.1 and 2.2 in several important ways. First, we do not assume that $\pi_i = \delta D_i$ where D_i is observed; instead π_i is completely unrestricted. Next, errors do not satisfy a cross-sectional restriction $\mathbb{E}[\pi_i(\epsilon_{it}, u_{it})] = 0$. The model for errors ϵ_i, u_i is similar to the one specified in Assumption 2.2. The key difference is that now we explicitly state that errors are affected by some unobserved aggregate shock H_t.

Assumption 3.1 specifies a complicated panel model with many components: factor errors, unobserved aggregate shocks, and additional independence restrictions. It should be contrasted with the cross-sectional model specified in Assumptions 2.1, which appears to be much simpler. This simplicity is deceiving: Assumption 2.1 does not describe the joint behavior of unit-level outcomes and aggregate shocks. As long as this behavior is important for identification, estimation, or inference, one has to take a stand on the joint model. This particular specification is motivated by a causal model that we describe in Section 4

Our next assumption restricts the process Z_t:

Assumption 3.2. Z_t *has the following representation:*

$$\begin{aligned} Z_t &= \eta_Z^{(0)} + \eta_Z^{(1)}\psi(t) + \tilde{Z}_t \\ \mathbb{E}[\tilde{Z}_t] &= 0, \end{aligned} \tag{3.3}$$

where $\psi(t) \in \mathbb{R}^p$ *is a known function of time.*

In a stationary case $\eta_Z^{(1)} \equiv 0$, and $\mathbb{E}[Z_t] = \eta_Z^{(0)}$, but in general we can allow the mean to be a function of time, as long as this function — $\psi(t)$ is known. In practice $\psi(t)$ can include trends, cyclical components, and other aggregate series, as long as they can be regarded as fixed for the purpose of the empirical exercise.

3.2 Estimation

Our approach to estimation follows the same TSLS logic that we outlined at the beginning of Section 2.1. We aggregate units across cross-sectional dimension using some weights, run two time-series OLS regressions, and take the ratio of the OLS coefficients. There are two important challenges in this process. First, we need to deal with the fact that π_i is not observed, as opposed to the simple case, when it is proportional to observed D_i. Second, because our weights can be correlated with idiosyncratic errors, the aggregate behavior of outcomes can be quite complicated. Thus we need to exploit the structure of Z_t given by Assumption 3.2 and run appropriate OLS regressions.

Our estimation algorithm has the following steps:

1. Fix $T_0 = \left\lfloor \frac{T}{2} \right\rfloor$ and estimate $\hat{\pi}_i$ by running the following OLS regression for each unit i using first T_0 periods:

$$W_{it} = \beta_i + \pi_i Z_t + \eta_i \psi(t) + u_{it}, \tag{3.4}$$

2. Construct unit weights $\hat{D}_i := \hat{\pi}_i - \overline{\hat{\pi}}$ and aggregate the outcomes:

$$W_t := \frac{1}{n}\sum_{i=1}^{n} W_{it}\hat{D}_i, Y_t := \frac{1}{n}\sum_{i=1}^{n} Y_{it}\hat{D}_i. \tag{3.5}$$

3. Run the following two regressions using data for periods $t > T_0$:

$$\begin{aligned} Y_t &= \eta_y^{(0)} + \eta_y^{(1)}\psi(t) + \delta Z_t + \epsilon_t, \\ W_t &= \eta_w^{(0)} + \eta_w^{(1)}\psi(t) + \pi Z_t + u_t. \end{aligned} \tag{3.6}$$

4. Construct the estimator for the treatment effect:

$$\hat{\tau}_{TS} := \frac{\hat{\delta}_{OLS}}{\hat{\pi}_{OLS}}. \tag{3.7}$$

This estimator can be computed simply by running a $TSLS$ regression with $\hat{D}_i Z_t$ as an instrument for W_{it} using only data for $t > T_0$:

$$Y_{it} = \alpha_i + \theta_t + \eta_i^\top \psi(t) + \tau W_{it} + \epsilon_{it}. \tag{3.8}$$

This algorithm is different from those that are currently used in empirical applications: first, D_i is not observed but rather constructed from the data in a disciplined way; second, $\psi(t)$ is included in the regression to extract mean-zero innovations from Z_t; and finally, only half of the periods are used to run the regression.

To describe the properties of $\hat{\tau}_{TS}$, we need to make additional statistical assumptions. The first assumption specifies the structure of $\tilde{Z}_t$:

Assumption 3.3. (Innovations) *Innovations $\tilde{Z}_t$ are independent and satisfy the following condition:*

$$\mathbb{E}[\tilde{Z}_t^2] = \sigma_Z^2. \tag{3.9}$$

This assumption makes two restrictions: it forbids any dependence between errors, and it assumes that the variance of innovations is constant over time. The latter restriction does not play an important role and can be safely dropped as long as we assume that the variances are bounded from above and below. The first restriction is important but can be substantially weakened (at some technical cost). In particular, we can allow $\tilde{Z}_t$ to be a martingale difference sequence. We also need to make other technical assumptions on tail behavior of relevant errors, which we state in Appendix A.

To state the formal result, we need to make additional definitions. We define the following sequence:

$$\mu_t := \mu_{\epsilon,t}^\top \mathbb{E}[\nu_i^{(1)} \pi_i], \tag{3.10}$$

and we let $\tilde{\mu}_t$ be the residuals in the following OLS regression:

$$\mu_t = \eta_\mu^{(0)} + \eta_\mu^{(1)} \psi(t) + e_t, \tag{3.11}$$

where $t > T_0$. Conceptually $\tilde{\mu}_t$ quantifies the size of the relevant aggregate error that affects the time-series behavior of the estimator. The following theorem makes this precise.

Theorem 3.1. (Statistical properties) *Suppose Assumptions 3.1, 3.2, 3.3, A.4 hold. Additionally suppose that $\frac{T}{n} = O(1)$ and $\frac{p}{T} = o(1)$. Then the following inequality holds with* *probability of at least* $1 - \frac{1}{T} + \frac{1}{n}$*:*

$$|\hat{\tau}_{TS} - \tau| \lesssim \frac{\log(T)\sqrt{\log(n)}}{\sigma_Z T} + \frac{\log(T)\|\mathbb{E}[\nu_i^{(1)}\pi_i]\|_2}{\sigma_Z \mathbb{V}[\pi_i]\sqrt{T}} \tag{3.12}$$

If, in addition $\|\mathbb{E}[\nu_i^{(1)}\pi_i]\|_2 > c$*, then the following statement is true:*

$$\sup_x \left|\mathbb{E}\left[\frac{\hat{\tau}_{TS} - \tau}{\sigma_{est}} \leq x\right] - \mathbb{E}[\mathcal{N}(0,1) \leq x]\right| \lesssim \frac{\log(T)\sqrt{\log(n)}}{\sigma_Z \mathbb{V}[\pi_i]\sqrt{T}} \tag{3.13}$$

where $\sigma_{est} := \frac{\sqrt{\sum_{t>T_0} \tilde{\mu}_t^2}}{\sigma_Z \mathbb{V}[\pi_i](T-T_0-p-1)}$

This theorem describes the behavior of the estimator in the practically relevant regime where $T \approx n$. The first part of the theorem says that if $\|\mathbb{E}[\nu_i^{(1)}\pi_i]\|_2$ is small (e.g., equal to zero) then the estimator behaves as $\mathcal{O}_p\left(\frac{1}{\sqrt{nT}}\right)$. This is important because this is the rate that one would get from the best estimator in a simple two-way model without any time-series complications. When $\|\mathbb{E}[\nu_i^{(1)}\pi_i]\|_2$ is not small, the time-series dimension dominates and the estimator behaves as $\mathcal{N}(0, \sigma_{est})$, where $\sigma_{est} = O(\frac{1}{\sqrt{T}})$.

Of course, this result is valid under quite restrictive time-series assumptions on $\tilde{Z}_t$. Given that the assumptions are restrictive, we view it as describing the first-best case for our estimator. All additional complications that can arise in practice are likely to make the result weaker. This is particularly important if one wants to use this result for inference.

3.3 Discussion of Inference

We leave the formal approach to inference in this model for future investigation. Theorem 3.1 can be used to construct a variance estimator and conduct normal-based inference. In practice, this is problematic for two reasons: first, the time-series process for $\tilde{Z}_t$ can be more complicated, and second, it is not robust to weak instruments. As a practical recommendation, we suggest that applied researchers use the robust t-test procedure developed in Ibragimov and Müller [2010], or the permutation-based test developed in Canay et al. [2017].

The standard approach to weak inference in the literature assumes that the first-stage and reduced-form estimators have a joint normal distribution with known variances and build a test based on this information (see Andrews et al. [2019] for a recent overview). In practice, these variances are not known, but with i.i.d. data they can be easily consistently estimated. For time series, consistent estimation of such variance is a much harder problem: many different solutions are proposed in the literature. While consistent estimation is possible in certain models, e.g., using HAC-type estimators (e.g., Newey and West [1986]), there is a common understanding that it does not work very well, especially with moderate T (see Müller [2014] for a discussion).

Another complication with time-series data is that the distribution of most statistics is not known under the null hypothesis. Sometimes the distribution contains parameters that cannot be consistently estimated (see Müller and Watson [2008, 2015]). As a result, to be fully robust one has to specify a class of time-series models and build a test that efficiently incorporates this information (see Elliott et al. [2015] for a general treatment). We believe that this is the first-best approach, but fully developing it for our model is beyond the scope of this paper.

4 Causal Model with Exogenous Aggregate Shocks

In this section, we present a parsimonious causal model that captures both cross-sectional and time-series aspects of the problem. Our primary goal is to motivate the econometric model we presented in Section 3. The secondary goal is to prove an identification result for the model with general heterogeneity in the first stage and in the treatment effects. We also use this model to clarify the sources of endogeneity that researchers need to address to conduct valid causal inference.

4.1 Framework

We observe n units (i being a generic one) over T periods (t is a generic period). For each unit, we observe an outcome variable Y_{it}, an endogenous policy variable W_{it}, and an aggregate shock Z_t. Our goal is to estimate a causal relationship between Y_{it} and W_{it}. We abstract away from additional unit-specific time-invariant covariates: they can be incorporated in a straightforward way.

To formalize what we mean by causality, we start with a model of potential outcomes. In addition to w_t (potential value of W_{it}) and z_t (potential value of Z_t), we also introduce h_t — an unobserved aggregate shock which causally affects both outcome variables. We define $w^t := (\dots, w_1, \dots, w_t)$, $z^t := (\dots, z_1, \dots, z_t)$, and $h^t := (\dots, h_1, \dots, h_t)$, and make the following assumption:

Assumption 4.1. (Potential outcomes)
Potential outcomes are generated in the following way:

$$\begin{aligned} Y_{it}(w^t, h^t) &:= \alpha_{it} + \tau_{it} w_t + \theta_{it} h_t \\ W_{it}(h^t, z^t) &= \beta_{it} + \pi_{it} z_t + \gamma_{it} h_t. \end{aligned} \tag{4.1}$$

As a result, the observed outcomes behave in the following way:

$$\begin{aligned} Y_{it} &= \alpha_{it} + \tau_{it} W_{it} + \theta_{it} H_t \\ W_{it} &= \beta_{it} + \pi_{it} Z_t + \gamma_{it} H_t. \end{aligned} \tag{4.2}$$

This assumption imposes multiple restrictions on the structure of potential outcomes: (a) there are no dynamic effects, only current values of shocks and endogenous variable matter; (b) the model is linear in w_t, z_t, and h_t with no interaction terms; and (c) there is an exclusion restriction for z_t — it is only present in the second equation. Linearity in (z_t, h_t) can be relaxed, even though it plays a role in estimation. Separability between z_t and h_t is crucial for our results. Finally, the exclusion restriction emphasizes the role of Z_t as an instrument.

Our first identification assumption describes the relation between the aggregate shocks and the potential outcomes:

Assumption 4.2. (Exogeneity)
Aggregate shocks are independent of potential outcomes:

$$\{Z_t, H_t\}_{t=1}^T \perp\!\!\!\perp \{\tau_{it}, \alpha_{it}, \beta_{it}, \theta_{it}, \pi_{it}, \gamma_{it}\}_{t=1}^T. \tag{4.3}$$

The importance of this assumption depends on the type of application one is considering. In standard event studies, Z_t is an indicator of the postreatment period – $\{t \geq T_{treat}\}$ – and is fixed. In this case, Assumption 4.2 trivially holds if we assume that the support of $\{Z_t, H_t\}_{t=1}^T$ is a singleton, or, to put if differently, if we condition on $\{Z_t, H_t\}_{t=1}^T$. This type of reasoning is predominant in event studies and other applications that use the difference in differences methodology. A notable exception is the model of staggered adoption considered in Athey and Imbens [2018], but there the shocks are unit-specific, not aggregate. In other applications, Z_t

and H_t are shocks that are determined on a macro level and are treated as random variables (in time-series sense). The model in this paper is designed for these applications, not event studies, although some insights apply there as well.

Assumption 4.2 becomes restrictive once it is combined with Assumption 4.1. To understand why this is the case, it is natural to think about an economic model where Z_t is determined in an equilibrium together with unit-level outcomes and is therefore endogenous. At the same time, Z_t is an aggregate variable and thus can be correlated with unit-level outcomes only if those are subject to some aggregate shocks that are correlated with Z_t. In principle, Assumption 4.2 allows for this as long as these shocks are captured by H_t, but at the same time Assumption 4.1 restricts the role H_t can play. It follows that Assumption 4.2 is more natural if (Z_t, H_t) is determined outside of the economic model for unit-level outcomes, or in other words, if the aggregate shocks are exogenous. Overall this assumption is conceptually different from design assumptions that are made in experimental or observational studies, e.g., random assignment or unconfoundedness (see Imbens and Rubin [2015]).

It follows from Assumptions 4.1 and 4.2 that the main problem for causal inference in the current model is unobserved H_t. Indeed, if $\{H_t\}_{t=1}^T$ were known then one could have treated each unit separately and the cross-sectional dimension would have been redundant for identification purposes. At the same time, in applications, it is natural to assume that we cannot fully control for aggregate uncertainty that affects the outcome of interest. For example, in the application that we described in Section 2.3, W_{it} is the amount of food aid from the U.S. that a country i receives, Y_{it} is local conflict and Z_t is the amount of wheat produced in the U.S. in the previous year. Here H_t can be another aggregate shock that affects conflict (e.g., climate change) and is also correlated with production in the United States.

In the panel regressions H_t typically enters as a time fixed effect. In our framework, we treat H_t as a a random variable. To understand why it is important, consider an alternative formulation where it is treated as fixed. In this case, we need to model the conditional distribution of $\{Z_t\}_{t=1}^T$ given $\{H_t\}_{t=1}^T$, which can be quite complicated, e.g., the mean of Z_t will depend on the unobserved $\{H_l\}_{l=1}^T$. This can easily render identification impossible. These problems do not arise if we treat H_t as random and assume that the *marginal* distribution of the observed stochastic process is simple. This logic motivates our next identification assumption, which is a high-level restriction on what is known about Z_t. First, we define the conditional mean of Z_t

for all $t > l$ (some fixed period):

$$\mu_{t|l} := \mathbb{E}[Z_t | Z_l, Z_{l-1}, \dots, Z_1] \tag{4.4}$$

and make the following assumption:

Assumption 4.3. (Policy process)
For any fixed l $\mu_{t|l}$ is a known function, and Z_t is not completely predicted by Z^l:

$$\mathbb{E}[(Z_t - \mu_{t|l})^2 | \mathcal{I}_l] = \mathbb{E}[(Z_t - \mu_{t|l})^2] > 0. \tag{4.5}$$

We can construct a de-meaned process $\tilde{Z}_{t|l}$ for all $t > l$:

$$\tilde{Z}_{t|l} := Z_t - \mu_{t|l}. \tag{4.6}$$

Assumption 4.3 is another restriction on the design that we need for identification (along with Assumption 4.2). Conceptually, it is similar to assumptions on the form of propensity-score model that one would make in observational studies. One can think of time and past values of Z_t as covariates that we need to control for. The second condition in Assumption 4.3 (positive variance) is similar to the standard overlap condition one needs for identification in observational studies (see Imbens and Rubin [2015]).

So far, we left the potential outcomes defined in Assumption 4.1 completely unrestricted. Since H_t is not observed, one has to make an assumption that distinguishes it from Z_t thus restricting the potential outcomes. In the standard regression model, one assumes that all units have the same exposure to unobserved aggregate shocks ("parallel trends") and that treatment effects are constant. In our model, we can allow for much more general heterogeneity.

Assumption 4.4. (Restricted heterogeneity)
The process for π_{it} satisfies the following restrictions for all t, t':

$$cov[\pi_{it}, \pi_{it'}] > c > 0. \tag{4.7}$$

There exists a known l such that: (a) the process $(\theta_{it}, \gamma_{it})$ has the following representation:

$$\begin{aligned} &(\theta_{it}, \gamma_{it}) = (\theta_t, \gamma_t) + \varepsilon_{it} \\ &\mathbb{E}[\varepsilon_{it} | \mathcal{I}_{t-l}] = 0, \end{aligned} \tag{4.8}$$

and (b) treatment effects are conditionally independent of innovations ε_{it}:

$$\tau_{it} \perp\!\!\!\perp \varepsilon_{it} | \mathcal{I}_{t-l}, \tag{4.9}$$

where $\mathcal{I}_t$ includes all information up time t.

This assumption restricts the dependence over t in $(\theta_{it}, \gamma_{it}), \pi_{it}$, and τ_{it}. We assume that the process for π_{it} is persistent and its autocorrelations for all lags are bounded from below. In contrast, we assume that the process for $(\theta_{it}, \gamma_{it})$ has a short memory, and its innovations are orthogonal to a sufficiently distant past[2]We also require treatment effects to be independent of innovations as long as we condition on the past. Examples below show that Assumption 4.4 is flexible enough to capture multiple practically relevant models. Heuristically, the main tradeoff with this assumption is that the conditional independence (4.8) is more plausible for l as small as possible, whereas the conditional orthogonality (4.9) is more plausible for l as large as possible.

Example 4.1. (Factor model) Potential outcomes have the following structure:

$$\begin{aligned} &(\alpha_{it}, \beta_{it}, \pi_{it}, \tau_{it}) = (\alpha_i \mu_t^{\alpha}, \beta_i \mu_t^{\beta}, \pi_i, \tau) \\ &(\theta_{it}, \gamma_{it}) = (\theta_t, \gamma_t) + \varepsilon_{it} \\ &\varepsilon_{it} = \sum_{k=1}^{l-1} \nu_{it-k} \rho_k + \nu_{it}, \end{aligned} \tag{4.10}$$

where ν_{it} is a sequence of independent innovations. The permanent coefficients have a factor representation, and (π_i, τ_i) do not change over time. With these restrictions, we can express

[2]This approach is similar to one Griliches and Hausman [1986] use to deal with the measurement error.

observed outcomes as a two-way model:

$$
\begin{aligned}
Y_{it} &= \tilde{\alpha}_i + \tilde{\theta}_t + \tau_i W_{it} + u_{it}^{(1)}, \\
W_{it} &= \tilde{\beta}_i + \tilde{\gamma}_t + \pi_i Z_t + u_{it}^{(2)},
\end{aligned}
\tag{4.11}
$$

where parameters have the following definition:

$$
\begin{aligned}
&u_{it}^{(1)} := (\alpha_i - \mathbb{E}[\alpha_i])(\mu_t^{\alpha} - \overline{\mu^{\alpha}}) + \varepsilon_{it}(H_t, 0)^{\top}, \\
&u_{it}^{(2)} := (\beta_i - \mathbb{E}[\beta_i])(\mu_t^{\beta} - \overline{\mu^{\beta}}) + \varepsilon_{it}(0, H_t)^{\top} \\
&\tilde{\alpha}_i := \alpha_i \overline{\mu^{\alpha}}, \quad \tilde{\theta}_t := \theta_t H_t + \mathbb{E}[\alpha_i](\mu_t^{\alpha} - \overline{\mu^{\alpha}}), \\
&\tilde{\beta}_i := \beta_i \overline{\mu^{\beta}}, \quad \tilde{\theta}_t := \theta_t H_t + \mathbb{E}[\beta_i](\mu_t^{\beta} - \overline{\mu^{\alpha}}).
\end{aligned}
\tag{4.12}
$$

Standard OLS has two problems in this model: first, components of ε_{it} can be correlated between each other — the endogeneity problem. On top of that, we have a misspecification problem: because potential outcomes have a factor structure, errors $u_{it}^{(k)}$ are correlated as long as α_i, β_i are correlated. This model would satisfy Assumption 2.1 if $\pi_i = \pi D_i$, and π_i is uncorrelated with (α_i, β_i). In general, there is no reason to expect that π_i is not correlated with (α_i, β_i), and thus this model satisfies Assumption 3.1.

4.2 Construction of Unit Weights

In this section, we construct unit weights that filter out the unobserved aggregate shocks. We show that under Assumption 4.4, our algorithm successfully deals with H_t. We fix a period T_0 and define the following random variable:

$$
D_i := \frac{1}{T_0} \sum_{t=1}^{T_0} \tilde{Z}_{t|0} W_{it} \tag{4.13}
$$

Under Assumption 4.3, this construction involves only observable quantities and thus is readily available. We define unit weights ω_i as centered versions of D_i:

$$
\omega_i := D_i - \mathbb{E}[D_i | \{Z_t\}_{t=1}^{T_0}], \tag{4.14}
$$

Using these weights, we define the aggregated outcomes for $t > T_0$:

$$
\begin{aligned}
W_t &:= \mathbb{E}[\omega_i W_{it} | Z_1, \dots, Z_T] \\
Y_t &:= \mathbb{E}[\omega_i Y_{it} | Z_1, \dots, Z_T]
\end{aligned}
\tag{4.15}
$$

and aggregate parameters:

$$
\begin{aligned}
\alpha_{t|t'} &:= \mathbb{E}[\omega_i(\alpha_{it} + \tau_{it}\beta_{it}) | \{Z_l\}_{l=1}^{t'}], \qquad & \beta_{t|t'} &:= \mathbb{E}[\omega_i \beta_{it} | \{Z_l\}_{l=1}^{t'}], \\
\pi_{t|t'} &:= \mathbb{E}[\omega_i \pi_{it} | \{Z_l\}_{l=1}^{t'}], & \delta_{t|t'} &:= \mathbb{E}[\tau_{it}\omega_i \pi_{it} | \{Z_l\}_{l=1}^{t'}].
\end{aligned}
\tag{4.16}
$$

The properties of aggregated outcomes are summarized in the following proposition:

Proposition 4.1. *Suppose Assumptions 4.1, 4.2, 4.3, and 4.4 hold. Then aggregate outcomes have the following representation for all $t \geq T_0 + l$:*

$$
\begin{aligned}
Y_t &= \alpha_{t|0} + \mathbb{E}[\delta_{t|T_0}\mu_{t|T_0}] + \delta_{t|0}\tilde{Z}_{t|T_0} + \epsilon_{y,t}, \\
W_t &= \beta_{t|0} + \mathbb{E}[\pi_{t|T_0}\mu_{t|T_0}] + \pi_{t|0}\tilde{Z}_{t|T_0} + \epsilon_{w,t},
\end{aligned}
\tag{4.17}
$$

where errors have the following form:

$$
\begin{aligned}
\epsilon_{y,t} &:= \left(\alpha_{t|T_0} - \alpha_{t|0}\right) + \left(\delta_{t|T_0}\mu_{t|T_0} - \mathbb{E}[\delta_{t|T_0}\mu_{t|T_0}]\right) + \left(\delta_{t|T_0} - \delta_{t|0}\right)\tilde{Z}_{t|T_0}, \\
\epsilon_{w,t} &:= \left(\beta_{t|T_0} - \beta_{t|0}\right) + \left(\pi_{t|T_0}\mu_{t|T_0} - \mathbb{E}[\pi_{t|T_0}\mu_{t|T_0}]\right) + \left(\pi_{t|T_0} - \pi_{t|0}\right)\tilde{Z}_{t|T_0}.
\end{aligned}
\tag{4.18}
$$

Also $\pi_{t|0} > 0$ for all $t \geq T_0 + l$.

This proposition shows that once we aggregate the outcomes using appropriate weights, the problem is reduced to a time-series IV regression with time-dependent deterministic parameters and stochastic errors. The key feature of the weights ω_i is that they integrate the unobserved shocks H_t away but keep $\tilde{Z}_{t|T0}$ (all $\pi_{t|0}$ are positive).

4.3 Aggregate Model

Given the results of Proposition 4.1 together with Assumption 4.3, we end up with the following time-series model for $t \geq T_0 + l$:

$$\begin{aligned} &Y_t = \alpha_{t|0} + \mathbb{E}[\delta_{t|T_0}\mu_{t|T_0}] + \delta_{t|0}\tilde{Z}_{t|T_0} + \epsilon_{y,t}, \\ &W_t = \beta_{t|0} + \mathbb{E}[\pi_{t|T_0}\mu_{t|T_0}] + \pi_{t|0}\tilde{Z}_{t|T_0} + \epsilon_{w,t}, \\ &\mathbb{E}\left[\tilde{Z}_{t|T_0} \times (\alpha_{t|0}, \beta_{t|0}, \mathbb{E}[\delta_{t|T_0}\mu_{t|T_0}], \mathbb{E}[\pi_{t|T_0}\mu_{t|T_0}], \epsilon_{y,t}, \epsilon_{w,t})\right] = 0. \end{aligned} \tag{4.19}$$

Using (4.19), we construct aggregate moments:

$$\begin{aligned} &\pi_{t|0}\mathbb{E}[\tilde{Z}^2_{t|T_0}] = \mathbb{E}[W_t\tilde{Z}_{t|T_0}] \\ &\delta_{t|0}\mathbb{E}[\tilde{Z}^2_{t|T_0}] = \mathbb{E}[Y_t\tilde{Z}_{t|T_0}]. \end{aligned} \tag{4.20}$$

Summing up these moments and taking the ratio, we define the following parameter:

$$\tau := \frac{\sum_{t=T_0+l}^{T} \mathbb{E}[Y_t\tilde{Z}_{t|T_0}]}{\sum_{t=T_0+l}^{T} \mathbb{E}[W_t\tilde{Z}_{t|T_0}]}. \tag{4.21}$$

Our next proposition summarizes the properties of τ:

Proposition 4.2. *Suppose Assumptions 4.1, 4.2, 4.3, and 4.4 hold. Then τ has the following representation:*

$$\tau = \sum_{t=T_0+l}^{T} \varsigma_t \left(\frac{\delta_{t|0}}{\pi_{t|0}}\right), \tag{4.22}$$

where $\varsigma_t > 0$ and τ is an affine combination of treatment effects. If, in addition, for all $t \geq T_0+l$ τ_{it} is independent of $\{\pi_{il}\}_{l=1}^{t}$, then $\frac{\delta_{t|0}}{\pi_{t|0}} = \mathbb{E}[\tau_{it}]$ and τ is a convex combination of treatment effects.

This proposition tells us that τ can be interpreted as a weighted average of treatment effects τ_{it} with the weights being non-negative under an additional independence assumption. We observe that the time weights ν_t are always positive. If there is no heterogeneity in treatment effects across units and $\tau_{it} \equiv \tau_t$, then τ is a convex combination of τ_t. This might not be the case if there is cross-sectional heterogeneity in treatment effects.. Intuitively, this happens because we use $\{\pi_{it}\}_{t=1}^{T_0}$ as instruments and they are not randomly assigned. If negative weights is a first-order concern in a given application, then one has to introduce additional application-specific structure to deal with this problem.

5 Discussion

5.1 Alternative Weights

In this section, we briefly discuss alternative weighting schemes that are currently used in empirical practice, compare them with the one proposed in Section 4.2, and illustrate their potential problems both in theory and in simulations. Below, we focus on a simple model for the potential outcomes, but the insights apply to more general models as well. We make the following restrictions on the potential outcomes:

$$
\begin{aligned}
&(\alpha_{it}, \beta_{it}, \pi_{it}, \tau_{it}) = (\alpha_i, \beta_i, \pi_i, \tau_i) \\
&(\theta_{it}, \gamma_{it}) = (\theta, \gamma) + \varepsilon_{it},
\end{aligned}
\tag{5.1}
$$

where ε_{it} is a stationary white-noise process independent of $(\alpha_i, \beta_i, \pi_i, \tau_i)$. For simplicity, we also assume that τ_i is independent of π_i. Observed variables thus have the following representation:

$$
\begin{aligned}
Y_{it} &= \alpha_i + \theta H_t + \tau_i W_{it} + \varepsilon_{it}(H_t, 0)^\top \\
W_{it} &= \beta_i + \gamma H_t + \pi_i Z_t + \varepsilon_{it}(0, H_t)^\top.
\end{aligned}
\tag{5.2}
$$

5.1.1 Many Fixed Effects

The first scheme that we focus on is based on estimating the following regression model by TSLS:

$$
\begin{aligned}
Y_{it} &= \alpha_i + \theta_t + \tau W_{it} + \epsilon_{it} \\
W_{it} &= \beta_i + \gamma_t + \pi_i Z_t + u_{it}
\end{aligned}
\tag{5.3}
$$

Treating $(\alpha_i, \beta_i, \theta_t, \gamma_t, \pi_i)$ as fixed parameters. For example, this approach has been used in Nakamura and Steinsson [2014]. Given the structure (5.2), this approach makes perfect sense and bypasses explicit construction of unit weights ω_i.

At the same time, running this regression is equivalent to using unit weights $\tilde{\omega}_i$ that have the following form:

$$\tilde{\omega}_i = (\pi_i - \mathbb{E}[\pi]) + \frac{\sum_{t=1}^T \varepsilon_{it,2} H_t \left(Z_t - \frac{1}{T}\sum_{l=1}^T Z_l \right)}{\sum_{t=1}^T Z_t \left(Z_t - \frac{1}{T}\sum_{l=1}^T Z_l \right)}. \tag{5.4}$$

This can lead to a problem, especially if π_i are small (and Z_t is a weak instrument). In this case, $\tilde{\omega}_i$ is dominated by the noise, and when we construct a product of $\tilde{\omega}_i$ with W_{it} and Y_{it} part of this noise is squared. In the limit when n goes to infinity, one can show that the following is true for any fixed T:

$$\begin{aligned}
&\hat{\tau}_{TSLS} = \mathbb{E}[\tau_i] + \nu \frac{\mathbb{E}[\varepsilon_{it,1}\varepsilon_{it,2}]}{\mathbb{E}[\varepsilon_{it,2}^2]} \\
&\nu := \frac{\mathbb{E}[\varepsilon_{it,2}^2] \frac{\sum_{t=1}^T H_t^2 \left(Z_t - \frac{1}{T}\sum_{l=1}^T Z_l\right)^2}{\left(\sum_{t=1}^T Z_t\left(\tilde{Z}_t - \frac{1}{T}\sum_{l=1}^T Z_l\right)\right)^2}}{\mathbb{V}[\pi_i] + \mathbb{E}[\varepsilon_{it,2}^2] \frac{\sum_{t=1}^T H_t^2 \left(Z_t - \frac{1}{T}\sum_{l=1}^T Z_l\right)^2}{\left(\sum_{t=1}^T Z_t\left(\tilde{Z}_t - \frac{1}{T}\sum_{l=1}^T Z_l\right)\right)^2}}.
\end{aligned} \tag{5.5}$$

For any fixed T, the TSLS estimator converges to a convex combination of the true parameter of interest and a bias term. Parameter α of this combination is random in general but is always less than 1. In other words, the estimator is always inconsistent. In particular, if $\pi_i = 0$ then $\nu = 0$ and the estimator converges to a fixed quantity (for any fixed T).

This behavior is quite different from the well-understood behavior under a single weak instrument, where the estimator converges to a random variable which is centered at the OLS limit (e.g., Bekker [1994]). From the practical point of view, the difference is crucial, because it means that the estimator is very stable, or, in other words, the first stage is very strong. In fact, this is a manifestation of the standard overfitting problem that arises because all the fixed effects are estimated (Angrist and Krueger [1992], Stock and Yogo [2002], Hahn and Hausman [2003]). Our procedure bypasses this problem using sample splitting — we estimate the weights on one part of the sample and then apply it to the second one. To summarize, TSLS estimation based on (5.3) is a very natural procedure that nevertheless suffers from certain econometric problems and thus should be used with care.

5.1.2 Functions of W_{it}

A different weighting scheme that is used in some applications starts by constructing the measure of exposure of W_{it} to Z_t in the following way:

$$D_i := \frac{1}{T_0}\sum_{t=1}^{T_0} f(W_{it}), \tag{5.6}$$

where $f(\cdot)$ is some fixed function. Nunn and Qian [2014] use this approach with an indicator function $f(x) := \{x > 0\}$.[3] With $\tilde{D}_i$, one can construct the corresponding unit weights in the same way as before. This is a flexible approach that can be useful in many settings, in particular if there is a conceptual model that justifies it. Nunn and Qian [2014] interpret $\tilde{D}_i$ as a probability of receiving treatment, which is then multiplied by the size of the aggregate shock.

There are two concerns with this algorithm. First, it uses W_{it} in levels and thus $\tilde{D}_i$ is strongly correlated with unit-level intercepts. This is not a problem *per se*, but it potentially makes $\tilde{D}_i$ only weakly correlated with π_i. Of course, in a given application the levels can be strongly correlated with the slopes, and $\tilde{D}_i$ will work relatively well.

Second, the construction above does not use Z_t in any way and thus does not directly capture the relationship between the aggregate instrument and the endogenous variable. This appears to be wasteful, because the main goal of $\tilde{D}_i$ is to measure the strength of the effect of Z_t on unit i. Hence, as we illustrate below, not using all of the information on the relationship between the instrument and the treatment will yield a less efficient estimator.

5.2 Shift-share Designs

In this section, we want to discuss a relationship between our model and models from the shift-share, or "Bartik" instruments, literature (Adão et al. [2018], Borusyak et al. [2018], Goldsmith-Pinkham et al. [2018]). We start by considering an extension of our original framework. Assume that instead of a single aggregate shock, we have $|S|$ of them. In a typical application, these will correspond to industry-level shocks. Potential outcomes are now determined by the following

[3] In fact, Nunn and Qian [2014] do not stop at T_0; and they use all the periods for construction of $\tilde{D}_i$. This potentially creates an overifitting problem similar to one that we have discussed above.

equations:

$$
\begin{aligned}
Y_{it} &= \alpha_{it} + \tau_{it} W_{it} + \sum_{s \in S} \theta_{its} H_{ts} \\
W_{it} &= \beta_{it} + \sum_{s \in S} \pi_{its} \omega_{its} Z_{ts} + \sum_{s \in S} \gamma_{its} H_{ts},
\end{aligned}
\tag{5.7}
$$

where s is a generic industry, and we observe $\{\omega_{its}\}_{i,t,s}$, $\{W_{it}, Y_{it}\}_{it}$, $\{Z_{ts}\}_{t,s}$, and $\sum_i \omega_{its} = 1$. It is straightforward to see that our model is a special case of this with $|S| = 1$.

The model considered in the shift-share literature is a special case of (5.7) with $T = 1$, and two additional assumptions: (a) for every s, H_{ts} is uncorrelated with $\{Z_{ts}\}_{s \in S}$, and (b) $\mathbb{E}[Z_{ts}] = \mu$ and Z_{ts} are uncorrelated over s. Identification is now achieved exploiting variation over industries (see Borusyak et al. [2018]). In applications, T is usually not equal to 1, and often the model in differences is considered. At the same time, the identification argument does not exploit the time dimension and focuses on the variation over industries.

One can immediately see that these two models are non-nested, both formally and conceptually: we are focusing on the case, with a single aggregate shock, motivated by the applications reviewed in Section 2.3. In these applications, correlation between observed and unobserved aggregate shocks is the key problem one has to deal with to make causal claims. Shift-share literature, on the other hand, focuses on the case where the main source of endogeneity is the cross-sectional correlation between α_{it} and β_{it} that typically arises because of simultaneity issues (e.g., when Y_{it} is wage and W_{it} is a labor supply).

We believe that models of the type (5.7) can be promising, because they allow for a combination of two identification arguments: one that is based on the variation over time, and one that is based on the variation over s. In applications, both s and t tend to be small (especially, if we want shocks to be independent over s), and thus it is natural to use both sources of variation. Also, using a time-series dimension, one can estimate correlations between Z_s and adapt inference to this case. Full development of these arguments is beyond the scope of this paper: we leave it for future research.

5.3 Monte Carlo Simulations

In this section, we consider the performance of our test from Section 2 and our estimator from Section 3 in Monte Carlo simulations that are based on real data. Simulations confirm our theoretical results, as well as the intuition from Section 5.1.

5.3.1 Performance of the Test

We take the data from the baseline specification in Nunn and Qian [2014] and estimate the following two regressions by OLS:

$$\begin{aligned} Y_{it} &= \alpha_i + \theta_t + \delta D_i Z_t + \epsilon_{it} \\ W_{it} &= \beta_i + \gamma_t + \pi D_i Z_t + u_{it} \end{aligned} \tag{5.8}$$

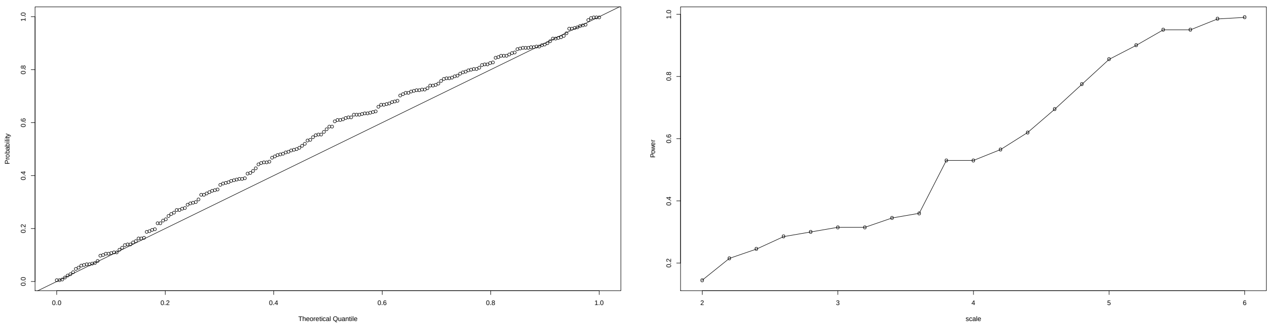
and keep all the parameters $\{\hat{\alpha}_i, \hat{\beta}_i, \hat{\theta}_t, \hat{\gamma}_t, \hat{\delta}, \hat{\pi}\}_{i,t}$. We also construct an empirical covariance matrix:

$$\hat{\Sigma}_{\epsilon,u} := \frac{1}{n} \sum_{i \le n} \begin{pmatrix} \hat{\epsilon}_i \\ \hat{u}_i \end{pmatrix} (\hat{\epsilon}_i^\top, \hat{u}_i^\top). \tag{5.9}$$

For each simulation draw s, we construct the outcomes as follows:

$$\begin{aligned} Y_{it}^{(s)} &:= \hat{\alpha}_i^{(s)} + \hat{\theta}_t + \hat{\delta} D_i^{(s)} Z_t + \epsilon_{it}^{(s)} \\ W_{it}^{(s)} &:= \hat{\beta}_i^{(s)} + \hat{\gamma}_t + \hat{\pi} D_i^{(s)} Z_t + u_{it}^{(s)}, \end{aligned} \tag{5.10}$$

where $(\hat{\alpha}_i^{(s)}, \hat{\beta}_i^{(s)}, D_i^{(s)})$ are sampled independently from $(\hat{\alpha}_i, \hat{\beta}_i, D_i)$, and $\epsilon_i^{(s)}, u_i^{(s)}$ are sampled from $\mathcal{N}(0, \hat{\Sigma}_{\epsilon,u})$. Panel A in Figure 5 shows the cumulative distribution function of the $\hat{p}$-value (2.15) versus the CDF of the uniform distribution across $S = 200$ simulations. The CDFs are close, and overall the approximation works well. It is not surprising, given that the errors are normally distributed.



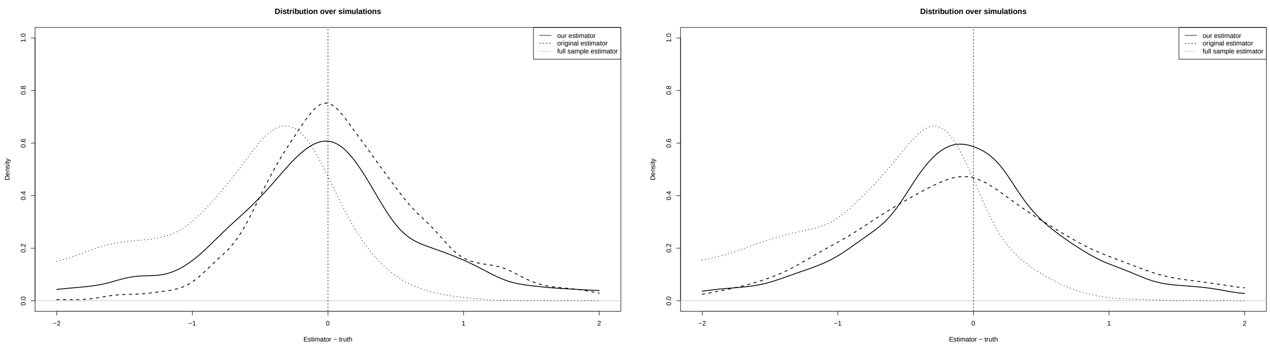
Panel A: QQ Plot Panel B: Local Power

Figure 5: Performance of the Test

Next, we examine power properties of our test and consider the following design:

$$
\begin{aligned}
Y_{it}^{(s)} &:= \hat{\alpha}_i^{(s)} + \hat{\theta}_t + \hat{\delta} D_i^{(s)} Z_t + \lambda D_i^{(s)} Z_t^{\perp} + \epsilon_{it}^{(s)} \\
W_{it}^{(s)} &:= \hat{\beta}_i^{(s)} + \hat{\gamma}_t + \hat{\pi} D_i^{(s)} Z_t + \lambda D_i^{(s)} Z_t^{\perp} + u_{it}^{(s)},
\end{aligned}
\tag{5.11}
$$

where $Z^{\perp}$ is a vector orthogonal to $(1, Z_t)$ (first vector of QR decomposition) such that $\frac{\|Z^{\perp}\|_2}{\sqrt{T}} = 1$. We set $\lambda = c\sqrt{\frac{\|\hat{\Sigma}_{\epsilon,u}\|}{nT}}$ and vary c. Panel B in Figure 5 shows the power of the test from $S = 200$ simulations (independent across points) for the size of 0.1. We see that once $c > 5$, the test perfectly detects the alternatives.



Panel A: Design I, (5.13) Panel B: Design II, (5.14)

Figure 6: Performance of the Estimator—Distributions

5.3.2 Performance of the Estimator

First, we estimate the ARMA model for Z_t:

$$Z_t := \mu + \rho_1 Z_{t-1} + \nu_t, \tag{5.12}$$

and we simulate $Z_t^{(s)}$ from this model using normal distribution of innovations. We consider two designs. In the first one, the outcomes are generated as follows:

$$\begin{aligned} Y_{it}^{(s)} &:= \hat{\alpha}_i^{(s)} + \hat{\theta}_t + \hat{\delta} D_i^{(s)} Z_t^{(s)} + 0.15 \times \hat{\epsilon}_{it}^{(s)} \\ W_{it}^{(s)} &:= \hat{\beta}_i^{(s)} + \hat{\gamma}_t + \hat{\pi} D_i^{(s)} Z_t^{(s)} + 0.15 \times \hat{u}_{it}^{(s)}. \end{aligned} \tag{5.13}$$

In the second design, the outcomes are generated as follows:

$$\begin{aligned} Y_{it}^{(s)} &:= \hat{\alpha}_i^{(s)} + \hat{\theta}_t + \hat{\delta} D_i^{(s)} (Z_t^{(s)} - 0.5\mathbb{E}^{(s)}[Z_t]) + 0.15 \times \hat{\epsilon}_{it}^{(s)} \\ W_{it}^{(s)} &:= \hat{\beta}_i^{(s)} + \hat{\gamma}_t + \hat{\pi} D_i^{(s)} (Z_t^{(s)} - 0.5\mathbb{E}^{(s)}[Z_t]) + 0.15 \times \hat{u}_{it}^{(s)}. \end{aligned} \tag{5.14}$$

Note that we do not simulate the errors from the normal distribution and instead sample the original residuals. The difference between these two designs is the level of information in the mean of Z_t. As we discussed before, this information is important for the original estimator used in Nunn and Qian [2014], so we want to investigate how it affects the performance. We compute three estimators: the one that was used by Nunn and Qian [2014], the one that was used by Nakamura and Steinsson [2014], and ours. We need to scale the original noise to 15%, since otherwise it completely overwhelms any signal in the data. In Figure 6, we plot the densities of three estimators over $S = 2,000$ simulations. We also report the quantiles of $\frac{|\hat{\tau}^{(s)} - \tau|}{|\tau|}$ for three estimators for both designs in Table 2.

The results confirm the intuition from Section 5.1: the estimator based on the full sample is biased and is dominated by our estimator in terms of quantiles of absolute bias. As expected, the original estimator wins if we do not change the mean of Z, but it is dominated by our estimator once we scale the mean by 50%. If we de-mean the process for $Z_t^{(s)}$ and use $Z_t^{(s)} - \mathbb{E}^{(s)}[Z_t]$ in 5.14, our estimator will substantially outperform the one of Nunn and Qian [2014].

Table 2: Estimator Performance–Quantiles of $\frac{|\hat{\tau}^{(s)}-\tau|}{|\tau|}$

	25%	50%	75%	25%	50%	75%
	Design I, (5.13)			*Design II, (5.14)*		
Nakamura and Steinsson [2014] estimator	0.07	0.16	0.34	0.07	0.16	0.35
Nunn and Qian [2014] estimator	0.04	0.10	0.19	0.07	0.17	0.31
Our estimator	0.05	0.13	0.27	0.06	0.13	0.27

Note: $\tau = 3.7$, $n = 98$, $S = 2,000$, $T = 36$.

6 Conclusion

Aggregate shocks are a natural source of exogenous variation for unit-level outcomes. As a result, many researchers use them to evaluate local-level policies. In this paper, we argue that this exercise has two critical steps: aggregation of unit-level variation and time-series analysis of the aggregated data. We illustrate that both of these steps should be done correctly. In essence, aggregation should deal with unobserved shocks, while time-series analysis should acknowledge that aggregate shocks, while exogenous, are not randomly assigned and thus should be modeled. We show how this can be done under particular assumptions. Of course, many questions remain open, both conceptually and practically: (a) How does dynamics affect these derivations? (b) What is the best way for conducting the inference? (c) What is the most efficient estimator in the current setup? We will explore these questions in future research.

References

Alberto Abadie and Javier Gardeazabal. The economic costs of conflict: A case study of the basque country. American Economic Review, 93(-):113–132, 2003.

Alberto Abadie, Alexis Diamond, and Jens Hainmueller. Synthetic control methods for comparative case studies: Estimating the effect of californias tobacco control program. Journal of the American statistical Association, 105(490):493–505, 2010.

Rodrigo Adão, Michal Kolesár, and Eduardo Morales. Shift-share designs: Theory and inference. Technical report, National Bureau of Economic Research, 2018.

Isaiah Andrews, James H Stock, and Liyang Sun. Weak instruments in instrumental variables regression: Theory and practice. Annual Review of Economics, 11:727–753, 2019.

Joshua D Angrist and Alan B Krueger. The effect of age at school entry on educational attainment: an application of instrumental variables with moments from two samples. Journal of the American statistical Association, 87(418):328–336, 1992.

Manuel Arellano. Panel data econometrics. Oxford university press, 2003.

Susan Athey and Guido W Imbens. Design-based analysis in difference-in-differences settings with staggered adoption. Technical report, National Bureau of Economic Research, 2018.

Timothy J Bartik. Who benefits from state and local economic development policies? 1991.

Paul A Bekker. Alternative approximations to the distributions of instrumental variable estimators. Econometrica: Journal of the Econometric Society, pages 657–681, 1994.

Alexandre Belloni, Victor Chernozhukov, Denis Chetverikov, Christian Hansen, and Kengo Kato. High-dimensional econometrics and regularized gmm. arXiv preprint arXiv:1806.01888, 2018.

Kirill Borusyak, Peter Hull, and Xavier Jaravel. Quasi-experimental shift-share research designs. Technical report, National Bureau of Economic Research, 2018.

Ivan A Canay, Joseph P Romano, and Azeem M Shaikh. Randomization tests under an approximate symmetry assumption. Econometrica, 85(3):1013–1030, 2017.

David Card and Alan B Krueger. Minimum wages and employment: A case study of the fast food industry in new jersey and pennsylvania. Technical report, National Bureau of Economic Research, 1993.

Oeindrila Dube and Juan F Vargas. Commodity price shocks and civil conflict: Evidence from colombia. The Review of Economic Studies, 80(4):1384–1421, 2013.

Graham Elliott, Ulrich K Müller, and Mark W Watson. Nearly optimal tests when a nuisance parameter is present under the null hypothesis. Econometrica, 83(2):771–811, 2015.

Paul Goldsmith-Pinkham, Isaac Sorkin, and Henry Swift. Bartik instruments: What, when, why, and how. Technical report, National Bureau of Economic Research, 2018.

Zvi Griliches and Jerry A Hausman. Errors in variables in panel data. Journal of econometrics, 31(1):93–118, 1986.

Jinyong Hahn and Jerry Hausman. Weak instruments: Diagnosis and cures in empirical econometrics. American Economic Review, 93(2):118–125, 2003.

Rustam Ibragimov and Ulrich K Müller. t-statistic based correlation and heterogeneity robust inference. Journal of Business & Economic Statistics, 28(4):453–468, 2010.

Guido W Imbens and Donald B Rubin. Causal inference in statistics, social, and biomedical sciences. Cambridge University Press, 2015.

Edward Miguel, Shanker Satyanath, and Ernest Sergenti. Economic shocks and civil conflict: An instrumental variables approach. Journal of political Economy, 112(4):725–753, 2004.

Ulrich K Müller. Hac corrections for strongly autocorrelated time series. Journal of Business & Economic Statistics, 32(3):311–322, 2014.

Ulrich K Müller and Mark W Watson. Testing models of low-frequency variability. Econometrica, 76(5):979–1016, 2008.

Ulrich K Müller and Mark W Watson. Low-frequency econometrics. Technical report, National Bureau of Economic Research, 2015.

Emi Nakamura and Jon Steinsson. Fiscal stimulus in a monetary union: Evidence from us regions. American Economic Review, 104(3):753–92, 2014.

Whitney K Newey and Kenneth D West. A simple, positive semi-definite, heteroskedasticity and autocorrelationconsistent covariance matrix, 1986.

Nathan Nunn and Nancy Qian. Us food aid and civil conflict. American Economic Review, 104 (6):1630–66, 2014.

Martin Raič et al. A multivariate berry–esseen theorem with explicit constants. Bernoulli, 25 (4A):2824–2853, 2019.

Mark Rudelson, Roman Vershynin, et al. Hanson-wright inequality and sub-gaussian concentration. Electronic Communications in Probability, 18, 2013.

James H Stock and Motohiro Yogo. Testing for weak instruments in linear iv regression. Technical report, National Bureau of Economic Research, 2002.

Roman Vershynin. High-dimensional probability: An introduction with applications in data science, volume 47. Cambridge University Press, 2018.

Martin J Wainwright. High-dimensional statistics: A non-asymptotic viewpoint, volume 48. Cambridge University Press, 2019.

Appendices

A Assumptions

Assumption A.1. *Errors have the following representation:*

$$\begin{pmatrix} \epsilon_i \\ u_i \end{pmatrix} = \begin{pmatrix} \mu_\epsilon \\ \mu_u \end{pmatrix} \nu_i^{(1)} + \begin{pmatrix} \Omega_\epsilon \\ \Omega_u \end{pmatrix} \nu_i^{(2)}, \tag{1.1}$$

where $\mu_\epsilon, \mu_u \in \mathbb{R}^{T\times 2k}$, $\nu_i^{(1)} \in \mathbb{R}^{2k}$, $\nu_i^{(2)} \in \mathbb{R}^{2T}$, *and* $\Omega_\epsilon, \Omega_u$ *are fixed symmetric* $T \times 2T$ *matrices; vector* $\nu_i := (\nu_i^{(1)}, \nu_i^{(2)})$ *has independent components,* $\mathbb{E}[\nu_i] = 0$, $\mathbb{E}[\nu_i\nu_i^\top] = \mathcal{I}_{2(k+T)}$ *and* $\|\nu_i\|_{\psi_2} \le K$; *let* $\mu := \begin{pmatrix} \mu_\epsilon \\ \mu_u \end{pmatrix}$ *and assume that* $\|\mu\| \le \bar{c}_\mu\sqrt{T}$; *let* $\Omega := \begin{pmatrix} \Omega_\epsilon \\ \Omega_u \end{pmatrix}$ *and assume that* $\max\{\|\Omega\|, \|\Omega^{-1}\|\} \le K$.

Assumption A.2. $\|D_i - \mathbb{E}[D_i]\|_{\psi_2} \le K$, $(D_i, \nu_i^{(1)})$ *is independent of* $\nu_i^{(2)}$, *the covariance matrix* $\Sigma := cov((D_i - \mathbb{E}[D_i])\nu_i)$ *is invertible, and trace*$(\Sigma) \lesssim \sqrt{(T+k)}$.

Assumption A.3. *The following moment restriction holds:*

$$\mathbb{E}[\nu_i^{(1)} D_i] = 0. \tag{1.2}$$

Assumption A.4. *The following tails conditions are satisfied:* $\max_t |H_t| \le C$, $\|\tilde{Z}\|_{\psi_2} \le K$, $\|D_i - \mathbb{E}[D_i]\|_{\psi_2} \le K$. *Errors have the following structure:*

$$\begin{pmatrix} \epsilon_i \\ u_i \end{pmatrix} = \begin{pmatrix} \mu_\epsilon \\ \mu_u \end{pmatrix} \nu_i^{(1)} + \begin{pmatrix} \mathbf{H}\Omega_\epsilon \\ \mathbf{H}\Omega_u \end{pmatrix} \nu_i^{(2)}, \tag{1.3}$$

where $\mu_\epsilon, \mu_u \in \mathbb{R}^{T\times 2k}$, $\nu_i^{(1)} \in \mathbb{R}^{2k}$, $\nu_i^{(2)} \in \mathbb{R}^{2T}$, $\Omega_\epsilon, \Omega_u$ *are fixed symmetric* $T \times 2T$ *matrices and* $\mathbf{H} := diag\{H_1, \dots, H_t\}$; *vector* $\nu_i := (\nu_i^{(1)}, \nu_i^{(2)})$ *has independent components,* $\mathbb{E}[\nu_i] = 0$, $\mathbb{E}[\nu_i\nu_i^\top] = \mathcal{I}_{2(k+T)}$ *and* $\|\nu_i\|_{\psi_2} \le K$; *let* $\mu := \begin{pmatrix} \mu_\epsilon \\ \mu_u \end{pmatrix}$ *and assume that* $\|\mu\| \le \bar{c}_\mu\sqrt{T}$; *let* $\Omega := \begin{pmatrix} \Omega_\epsilon \\ \Omega_u \end{pmatrix}$ *and assume that* $\max\{\|\Omega\|, \|\Omega^{-1}\|\} \le K$. $(D_i, \nu_i^{(1)})$ *is independent of* $\nu_i^{(2)}$.

B Formal Description of the Testing Procedure

B.1 Formal Statement of the Testing Problem

We assume that outcomes are generated by the following model:

$$
\begin{aligned}
Y_{it} &= \alpha_i + \theta_t + \delta D_i Z_t + \epsilon_{it} \\
W_{it} &= \beta_i + \gamma_t + \pi D_i Z_t + u_{it},
\end{aligned} \tag{2.1}
$$

where $(D_i, \epsilon_{it}, u_{it})$ satisfy Assumptions A.1, A.2, and $\{(D_i, \epsilon_i, u_i)\}_{i \le n}$ is an i.i.d. sequence conditional on $\{Z_t\}_{t \le T}$. We are testing Assumption A.3; formally $\mathbb{H}_0 : \mathbb{E}[\nu_i^{(1)} D_i] = 0$ and $\mathbb{H}_1 : \|\mathbb{E}[\nu_i^{(1)} D_i]\|_2 > 0$.

B.2 Bootstrap Details

Let $\mathbf{Z}$ be $T \times 2$ matrix, such that $(\mathbf{Z})_t = (1, Z_t)$, and let $\mathbf{P}_{Z^\perp}$ be the orthogonal projector on the subspace orthogonal to columns of $\mathbf{Z}$. We generate B bootstrap replications; for each of them, we generate n i.i.d. binary random variables $e_i \in \{-1, 1\}$, such that $\mathbb{E}[e_i] = 0$, and we construct the bootstrap version of the statistic:

$$
\begin{aligned}
&\hat{\xi}_\epsilon^{(b)} := \mathbf{P}_{Z^\perp} \left(\frac{1}{n} \sum_{i \le n} (e_i - \overline{e})(D_i - \overline{D}) \hat{\epsilon}_i \right), \; \hat{\xi}_u^{(b)} := \mathbf{P}_{Z^\perp} \left(\frac{1}{n} \sum_{i \le n} (e_i - \overline{e})(D_i - \overline{D}) \hat{u}_i \right), \\
&\hat{\xi}^{(b)} := \begin{pmatrix} \hat{\xi}_\epsilon \\ \hat{\xi}_u \end{pmatrix}, \; \hat{L}_n^{(b)}(\alpha) := \sqrt{\frac{n}{(T-2)}} \sqrt{\alpha \|\hat{\xi}_\epsilon^{(b)}\|_2^2 + (1-\alpha) \|\hat{\xi}_u^{(b)}\|_2^2}.
\end{aligned} \tag{2.2}
$$

The oracle test is defined as follows:

$$
\begin{aligned}
&\xi_\epsilon := \mathbf{P}_{Z^\perp} \left(\frac{1}{n} \sum_{i \le n} (D_i - \mathbb{E}[D_i]) \epsilon_i \right), \; \xi_u := \mathbf{P}_{Z^\perp} \left(\frac{1}{n} \sum_{i \le n} (D_i - \mathbb{E}[D_i]) u_i \right) \\
&\xi := \begin{pmatrix} \xi_\epsilon \\ \xi_u \end{pmatrix}, \; L_n(\alpha) := \sqrt{\frac{n}{(T-2)}} \sqrt{\alpha \|\xi_\epsilon\|_2^2 + (1-\alpha) \|\xi_u\|_2^2}, \\
&F_\alpha(x) := \mathbb{E}[\{L_n \le x\}].
\end{aligned} \tag{2.3}
$$

C Auxiliary Lemmas

Everywhere below, c refers to a universal constant that does not depend on the underlying DGP.

Lemma C.1. *Suppose Assumptions A.1,A.2 holds. Then conditionally on $\{e_i\}_{i=1}^n$, we have the following with probability at least $1-\frac{1}{n}$:*

$$\begin{aligned}
&\|\mathbb{P}_n\eta_i\|_2 \lesssim \frac{\|\mu\|_{HS}+\|\Omega\|_{HS}+\|\mu\|\sqrt{\log(n)}}{\sqrt{n}}\\
&|\mathbb{E}[D_i]-\mathbb{P}_nD_i| \lesssim \sqrt{\frac{\log(n)}{n}}\\
&\|\mathbb{P}_n\eta_iD_i\|_2 \lesssim \|\mu\|\sqrt{\frac{k+\log(n)}{n}}+\|\Omega\|\sqrt{\frac{T+\log(n)}{n}}\\
&\|\mathbb{P}_ne_i(D_i-\mathbb{E}[D_i])| \lesssim \sqrt{\frac{\log(n)}{n}}\\
&\|\mathbb{P}_ne_i\eta_i\|_2 \lesssim \frac{\|\mu\|_{HS}+\|\Omega\|_{HS}+\|\mu\|\sqrt{\log(n)}}{\sqrt{n}}\\
&\mathbb{P}_n\|\nu_i(D_i-\mathbb{E}[D_i])\|_2^p \lesssim 2^p\left(\sqrt{(T+k)}\log(n)\right)^p\\
&\left|\mathbb{P}_n(D_i-\mathbb{E}[D_i])^2-\mathbb{V}[D_i]\right| \lesssim \sqrt{\frac{\log(n)}{n}}.
\end{aligned} \tag{3.1}$$

Also, conditionally on $\{(D_i,\eta_i)\}_{i=1}^n$ we have the following probability at least $1-\frac{1}{n}$:

$$|\mathbb{P}_ne_i| \lesssim \sqrt{\frac{\log(n)}{n}}. \tag{3.2}$$

Proof. The proofs are straightforward applications of standard results for sug-Gaussian and sub-exponential random vectors (e.g., Vershynin [2018]), Hanson-Wright inequality (Theorem 1.1 in Rudelson et al. [2013]). □

Lemma C.2. *Suppose Assumptions A.1,A.2 holds. Then the following inequality holds conditionally on $\{D_i\}_{i=1}^n$ with probability at least $1-\frac{1}{n}$:*

$$\begin{aligned}
&\left|\left\|\mathbf{P}_{Z^\perp}\Omega\mathbb{P}_n\nu_i^{(2)}(D_i-\mathbb{E}[D_i])\right\|_2^2-\frac{\mathbb{P}_n(D_i-\mathbb{E}[D_i])^2}{n}\|\mathbf{P}_{Z^\perp}\Omega^2\mathbf{P}_{Z^\perp}\|_{HS}^2\right| \lesssim \frac{\sqrt{\log(n)}}{n}\|\mathbf{P}_{Z^\perp}\Omega^2\mathbf{P}_{Z^\perp}\|_{HS}\|D_i-\mathbb{E}[D_i]\|_\infty^2\\
&\left|\left(\mathbf{P}_{Z^\top}\mu\mathbb{P}_n\nu_i^{(1)}(D_i-\mathbb{E}[D_i])\right)^\top\left(\mathbf{P}_{Z^\top}\Omega\mathbb{P}_n\nu_i^{(2)}(D_i-\mathbb{E}[D_i])\right)\right| \lesssim \|\mu\|\|D_i-\mathbb{E}[D_i]\|_\infty\frac{\sqrt{\log(n)(\log(n)+k)}}{n}
\end{aligned} \tag{3.3}$$

Proof. The first result follows from conditional application of Hanson-Wright inequality, along with the following two simple facts:

$$\begin{aligned}
&\mathbb{E}\left[\left\|\mathbf{P}_{Z^\perp}\Omega\mathbb{P}_n\nu_i^{(2)}(D_i-\mathbb{E}[D_i])\right\|_2^2|\{D_i\}_{i\le n}\right] = \frac{\mathbb{P}_n(D_i-\mathbb{E}[D_i])^2}{n}\|\mathbf{P}_{Z^\perp}\Omega^2\mathbf{P}_{Z^\perp}\|_{HS} \\
&\|\mathbb{P}_n\nu_i^{(2)}(D_i-\mathbb{E}[D_i])|\{D_i\}_{i\le n}\|_{\psi_2} \le \frac{\|\nu_i^{(2)}\|_{\psi_2}\|D_i-\mathbb{E}[D_i]\|_\infty}{\sqrt{n}}.
\end{aligned} \tag{3.4}$$

The second result follows from first conditioning on $\left(D_i,\nu_i^{(1)}\right)_{i\le n}$ and then applying standard inequalities. □

Lemma C.3. *Suppose Assumptions A.1,A.2 are satisfied. Then we have the following with probability at least* $1-\frac{1}{n}$*:*

$$\begin{aligned}
&\sup_{A\in\mathcal{A}}\left|\mathbb{E}^{(b)}\left[\{\mathbb{P}_n\nu_i(D_i-\mathbb{E}[D_i])e_i\in A\}\right]-\mathbb{E}\left[\left\{\frac{1}{\sqrt{n}}\mathcal{N}(0,\Sigma_n)\in A\right\}\right]\right| \lesssim \frac{(k+T)^{\frac{7}{4}}}{\sqrt{n}}\log^3(n) \\
&\sup_{A\in\mathcal{A}}\left|\mathbb{E}^{(b)}\left[\left\{\mathbb{P}_n\nu_i^{(1)}(D_i-\mathbb{E}[D_i])e_i\in A\right\}\right]-\mathbb{E}\left[\left\{\frac{1}{\sqrt{n}}\mathcal{N}(0,\Sigma_{n,1})\in A\right\}\right]\right| \lesssim \frac{k^{\frac{7}{4}}}{\sqrt{n}}\log^3(n),
\end{aligned} \tag{3.5}$$

where $\mathcal{A}$ is a set of all measurable convex sets of relevant dimension.

Proof. Define $\Sigma_1 := \mathbb{E}[\nu_i^{(1)}(\nu_i^{(1)})^\top(D_i-\mathbb{E}[D_i)^2]$. We have the following inequality (using Theorem 4.7.1 in Vershynin [2018]) with probability at least $1-\frac{1}{n}$:

$$\begin{aligned}
&\|\mathbb{P}_n\nu_i\nu_i^\top(D_i-\mathbb{E}[D_i])^2-\Sigma\| \lesssim \|D_i-\mathbb{E}[D_i]\|_\infty^2\left(\sqrt{\frac{T+k+\log(n)}{n}}+\frac{T+k+\log(n)}{n}\right) \\
&\|\mathbb{P}_n\nu_i^{(1)}(\nu_i^{(1)})^\top(D_i-\mathbb{E}[D_i]^2)-\Sigma_1\| \lesssim \|D_i-\mathbb{E}[D_i]\|_\infty^2\left(\sqrt{\frac{k+\log(n)}{n}}+\frac{k+\log(n)}{n}\right).
\end{aligned} \tag{3.6}$$

We define the following covariance matrices:

$$\begin{aligned}
&\Sigma_n := \mathbb{P}_n(\nu_i(\nu_i)^\top(D_i-\mathbb{E}[D_i)^2) \\
&\Sigma_{n,1} : \mathbb{P}_n(\nu_i^{(1)}(\nu_i^{(1)})^\top(D_i-\mathbb{E}[D_i)^2).
\end{aligned} \tag{3.7}$$

Using multidimensional Berry-Esseen theorem conditionally on $\{(\nu_i,D_i)\}_{i\le n}$ and Lemma C.1, we get the following:

$$\begin{aligned}
&\sup_{A\in\mathcal{A}}\left|\mathbb{E}^{(b)}\left[\{\mathbb{P}_n\nu_i(D_i-\mathbb{E}[D_i])e_i\in A\}\right]-\mathbb{E}\left[\left\{\frac{1}{\sqrt{n}}\mathcal{N}(0,\Sigma_n)\in A\right\}\right]\right| \lesssim \frac{(k+T)^{\frac{7}{4}}}{\sqrt{n}}\log^3(n) \\
&\sup_{A\in\mathcal{A}}\left|\mathbb{E}^{(b)}\left[\left\{\mathbb{P}_n\nu_i^{(1)}(D_i-\mathbb{E}[D_i])e_i\in A\right\}\right]-\mathbb{E}\left[\left\{\frac{1}{\sqrt{n}}\mathcal{N}(0,\Sigma_{n,1})\in A\right\}\right]\right| \lesssim \frac{k^{\frac{7}{4}}}{\sqrt{n}}\log^3(n).
\end{aligned} \tag{3.8}$$

We also have the following inequality for multivariate normals with the same mean and different variances:

$$\sup_{A\in\mathcal{A}}\left|\mathbb{E}\left[\{\mathcal{N}(0,\Sigma)\in A\}\right]-\mathbb{E}\left[\{\mathcal{N}(0,\Sigma')\in A\}\right]\right| \lesssim \|\Sigma^{-1}\|_{HS}\|\Sigma-\Sigma'\|. \tag{3.9}$$

Applying this inequality to Σ, Σ_n, and $\Sigma_1, \Sigma_{1,n}$ we get the following bounds with probability at least $1 - \frac{c}{n}$:

$$\sup_{A\in\mathcal{A}} |\mathbb{E}[\{\mathcal{N}(0,\Sigma_n) \in A\}] - \mathbb{E}[\{\mathcal{N}(0,\Sigma) \in A\}]| \lesssim \sqrt{T+k}\|D_i - \mathbb{E}[D_i]\|_\infty^2 \left(\sqrt{\frac{T+k+\log(n)}{n}} + \frac{T+k+\log(n)}{n}\right)$$
$$\sup_{A\in\mathcal{A}} |\mathbb{E}[\{\mathcal{N}(0,\Sigma_{1,n}) \in A\}] - \mathbb{E}[\{\mathcal{N}(0,\Sigma_1) \in A\}]| \lesssim \sqrt{k}\|D_i - \mathbb{E}[D_i]\|_\infty^2 \left(\sqrt{\frac{k+\log(n)}{n}} + \frac{k+\log(n)}{n}\right). \tag{3.10}$$

As a result, we get the following:

$$\sup_{A\in\mathcal{A}} \left|\mathbb{E}^{(b)}\left[\left\{\sqrt{n}\mathbb{P}_n\nu_i(D_i - \mathbb{E}[D_i])e_i \in A\right\}\right] - \mathbb{E}\left[\left\{\Sigma^{\frac{1}{2}}\mathcal{N}(0,\mathcal{I}_{2(T+k)}) \in A\right\}\right]\right| \lesssim \frac{(k+T)^{\frac{7}{4}}}{\sqrt{n}}\log^3(n)$$
$$\sup_{A\in\mathcal{A}} \left|\mathbb{E}^{(b)}\left[\left\{\sqrt{n}\mathbb{P}_n\nu_i^{(1)}(D_i - \mathbb{E}[D_i])e_i \in A\right\}\right] - \mathbb{E}\left[\left\{\Sigma_1^{\frac{1}{2}}\mathcal{N}(0,\mathcal{I}_{2(k)}) \in A\right\}\right]\right| \lesssim \frac{k^{\frac{7}{4}}}{\sqrt{n}}\log^3(n), \tag{3.11}$$

which proves the claim. □

D Proofs for Section 2

Proof of Proposition 2.1:

Proof. From Assumptions A.1,A.2 we have that the following holds with probability at least $1-\frac{1}{n}$ (using standard inequalities for sub-Gaussian vectors inequalities):

$$\begin{aligned}
&|\hat{\mathbb{V}}[D_i] - \mathbb{V}[D_i]| \lesssim \sqrt{\frac{\log(n)}{n}} \\
&|\mathbb{E}[D_i] - \mathbb{P}_n[D_i]| \lesssim \sqrt{\frac{\log(n)}{n}} \\
&\|\mathbb{P}_n\epsilon_i\|_\infty \leq \|\mu_\epsilon \mathbb{P}_n \nu_i^{(1)}\|_\infty + \|\Omega_\epsilon \mathbb{P}_n \nu_i^{(2)}\|_\infty \lesssim \sup_t \|\mu_{\epsilon,t}\| \sqrt{\frac{\log(nk)}{n}} + \sqrt{\frac{\log(nT)}{n}} \\
&\|\mathbb{P}_n u_i\|_\infty \lesssim \sup_t \|\mu_{\epsilon,t}\| \sqrt{\frac{\log(nk)}{n}} + \sqrt{\frac{\log(nT)}{n}} \\
&\|\mathbb{P}_n D_i u_i\|_\infty \lesssim \sup_t \|\mu_{\epsilon,t}\| \sqrt{\frac{\log(nk)}{n}} + \sqrt{\frac{\log(nT)}{n}}.
\end{aligned} \tag{4.1}$$

Combining these inequalities, we get the following with the same probability:

$$\begin{aligned}
&\|\check{\epsilon}\|_\infty \lesssim \sup_t \|\mu_{\epsilon,t}\| \sqrt{\frac{\log(nk)}{n}} + \sqrt{\frac{\log(nT)}{n}} \\
&\|\check{u}\|_\infty \lesssim \sup_t \|\mu_{u,t}\| \sqrt{\frac{\log(nk)}{n}} + \sqrt{\frac{\log(nT)}{n}},
\end{aligned} \tag{4.2}$$

which concludes the proof. □

Proof of Proposition 2.2: In this proof, we focus on the case where $\alpha = \frac{1}{2}$; the proof for the general case is the same.

Proof. Define the infeasible bootstrap variables:

$$\begin{aligned}
&\xi_\epsilon^{(b)} := \mathbf{P}_{Z^\perp}\left(\frac{1}{n}\sum_{i\le n} e_i(D_i - \mathbb{E}[D_i])\epsilon_i\right),\ \xi_u^{(b)} := \mathbf{P}_{Z^\perp}\left(\frac{1}{n}\sum_{i\le n} e_i(D_i - \mathbb{E}[D_i])u_i\right),\\
&\xi^{(b)} := \begin{pmatrix}\xi_\epsilon^{(b)}\\ \xi_u^{(b)}\end{pmatrix},\ L_n^{(b)} := \sqrt{\frac{n}{2(T-2)}}\|\xi^{(b)}\|_2.
\end{aligned} \tag{4.3}$$

Below, we will abuse the notation and let $\mathbf{P}_{Z^\perp}$ to denote the following matrix:

$$\begin{pmatrix}\mathbf{P}_{Z^\perp} & \mathbf{0}_{T\times T}\\ \mathbf{0}_{T\times T} & \mathbf{P}_{Z^\perp}.\end{pmatrix} \tag{4.4}$$

Step 1: Define $\hat{\lambda} := (\hat{\theta}_t, \hat{\gamma}_t)$. We have the following representation:

$$\begin{aligned}
&r := (\mathbb{E}[D_i] - \mathbb{P}_n D_i)\mathbf{P}_{Z^\perp}\mathbb{P}_n\eta_i\\
&r^{(b)} := -\mathbf{P}_{Z^\perp}\mathbb{P}_n\eta_i\left(\mathbb{P}_n e_i(D_i - \mathbb{E}D_i) + (\mathbb{E}D_i - \mathbb{P}_n D_i)\mathbb{P}_n e_i\right) + (\mathbb{E}D_i - \mathbb{P}_n D_i)\mathbf{P}_{Z^\perp}\mathbb{P}_n e_i\eta_i\\
&\hat{\xi} = \xi + r\\
&\hat{\xi}^{(b)} = \xi^{(b)} - \mathbb{P}_n e_i\hat{\xi} + r^{(b)}.
\end{aligned} \tag{4.5}$$

By triangular inequality, we have the following:

$$|\hat{L}_n^{(b)} - L_n^{(b)}| \le \sqrt{\frac{n}{2(T-2)}}(\|r^{(b)}\|_2 + |\mathbb{P}_n e_i|\|r\|_2 + |\mathbb{P}_n e_i|\|\xi\|_2). \tag{4.6}$$

Using lemma C.1, we get that the following inequality holds with probability $1 - \frac{1}{n}$:

$$\begin{aligned}
&|\hat{L}_n^{(b)} - L_n^{(b)}| \lesssim \frac{\|\mu\|}{\sqrt{T}}\left(\sqrt{k + \log(n)}\right)\sqrt{\frac{\log(n)}{n}} =: r_{1,n}\\
&|\hat{L}_n - L_n| \lesssim r_{1,n}.
\end{aligned} \tag{4.7}$$

Step 2: Our goal is to show that the random distribution $L_n^{(b)}$ and the fixed distribution $\hat{L}_n$ with high probability are close in Kolmogorov-Smirnov distance with high probability. To do this, we show that both of them are close to a convex function of a multivariate normal distribution. We consider two different normal distributions: one is based on fixed T approximation, and another one is based on a large T approximation.

Below we define these two normal distributions and bound relevant distances.

$$\begin{aligned}\zeta_{fixed} &:= \frac{1}{\sqrt{2(T-2)}}\left(\mathbf{P}_{Z^\perp}\mu\Sigma_1^{\frac{1}{2}}\zeta^{(1)} + \mathbf{P}_{Z^\perp}\Omega\sqrt{\mathbb{V}[\pi]}\zeta^{(2)}\right)\\ \zeta_{limit} &:= \frac{1}{\sqrt{2(T-2)}}\left(\mathbf{P}_{Z^\perp}\mu\Sigma_1^{\frac{1}{2}}\zeta^{(1)}\right),\end{aligned} \tag{4.8}$$

where $(\zeta^{(1)}, \zeta^{(2)}) \sim \mathcal{N}(0, \mathcal{I}_{2(T+k)})$ and $\Sigma_1 := \mathbb{E}\left[(D_i - \mathbb{E}[D_i])^2\nu_i^{(1)}(\nu_i^{(1)})^\top\right]$. The following inequalities are immediate consequences of a multivariate Berry-Essen central limit theorem (e.g., Raič et al. [2019]) :

$$\begin{aligned}\sup_{A\in\mathcal{A}}\left|\mathbb{E}\left[\left\{\sqrt{\frac{n}{2(T-2)}}\mathbf{P}_{Z^\perp}\mathbb{P}_n(D_i - \mathbb{E}[D_i])\eta_i \in A\right\}\right] - \mathbb{E}[\zeta_{fixed} \in A\}]\right| &\lesssim \frac{(2k+T)^{\frac{7}{4}}}{\sqrt{n}}\\ \sup_{A\in\mathcal{A}}\left|\mathbb{E}\left[\left\{\sqrt{\frac{n}{2(T-2)}}\mathbf{P}_{Z^\perp}\mu\mathbb{P}_n(D_I - \mathbb{E}[D_i])\nu_i^{(1)} \in A\right\}\right] - \mathbb{E}[\zeta_{limit} \in A\}]\right| &\lesssim \frac{(2k)^{\frac{7}{4}}}{\sqrt{n}},\end{aligned} \tag{4.9}$$

where $\mathcal{A}$ is a set of all measurable convex sets of relevant dimension. The following inequalities follow from Lemma C.3 with probability at least $1 - \frac{1}{n}$:

$$\begin{aligned}\sup_{A\in\mathcal{A}}\left|\mathbb{E}^{(b)}\left[\left\{\sqrt{\frac{n}{2(T-2)}}\mathbf{P}_{Z^\perp}\mathbb{P}_n e_i(D_i - \mathbb{E}[D_i])\eta_i \in A\right\}\right] - \mathbb{E}[\zeta_{fixed} \in A\}]\right| &\lesssim \frac{(2k+T)^{\frac{7}{4}}}{\sqrt{n}}\log^3(n)\\ \sup_{A\in\mathcal{A}}\left|\mathbb{E}^{(b)}\left[\left\{\sqrt{\frac{n}{2(T-2)}}\mathbf{P}_{Z^\perp}\mu\mathbb{P}_n e_i(D_I - \mathbb{E}[D_i])\nu_i^{(1)} \in A\right\}\right] - \mathbb{E}[\zeta_{limit} \in A\}]\right| &\lesssim \frac{(2k)^{\frac{7}{4}}}{\sqrt{n}}\log^3(n).\end{aligned} \tag{4.10}$$

Step 3: Using Lemma C.2, we get the following inequality with probability $1 - \frac{1}{n}$:

$$\left|(L_n)^2 - \frac{\mathbb{V}[\pi_i]\|\mathbf{P}_{Z^\perp}\Omega^2\mathbf{P}_{Z^\perp}\|_{HS}^2}{2(T-2)} - \frac{n\left\|\mathbf{P}_{Z^\top}\mu\mathbb{P}_n\nu_i^{(1)}(D_i - \mathbb{E}[D_i])\right\|_2^2}{2(T-2)}\right| \lesssim$$
$$\frac{\|D_i - \mathbb{E}[D_i]\|_\infty\sqrt{\log(n)}\left(\|\mu\|\sqrt{(\log(n)+k)} + \|D_i - \mathbb{E}[D_i]\|_\infty\|\mathbf{P}_{Z^\perp}\Omega^2\mathbf{P}_{Z^\perp}\|_{HS}\right)}{T} =: r_{2,n}^2 \tag{4.11}$$

The same holds conditionally on $\{e_i\}_{i\le n}$ for $L_n^{(b)}$. Specifically, with conditional probability at least $1 - \frac{c}{n}$:

$$\left|(L_n^{(b)})^2 - \frac{\mathbb{V}[\pi_i]\|\mathbf{P}_{Z^\perp}\Omega^2\mathbf{P}_{Z^\perp}\|_{HS}^2}{2(T-2)} - \frac{n\left\|\mathbf{P}_{Z^\top}\mu\mathbb{P}_n e_i\nu_i^{(1)}(D_i - \mathbb{E}[D_i])\right\|_2^2}{2(T-2)}\right| \lesssim r_{2,n}^2. \tag{4.12}$$

Using elementary inequality $|\sqrt{x}-\sqrt{y}| \le \sqrt{|x-y|}$, we can conclude the following with conditional probability at least $1-\frac{c}{n}$:

$$\begin{aligned}
&\left| L_n - \sqrt{\frac{\mathbb{V}[\pi_i]\|\mathbf{P}_{Z^\perp}\Omega^2\mathbf{P}_{Z^\perp}\|_{HS}^2}{2(T-2)} + \frac{n\left\|\mathbf{P}_{Z^\perp}\mu\mathbb{P}_n\nu_i^{(1)}(D_i-\mathbb{E}[D_i])\right\|_2^2}{2(T-2)}} \right| \lesssim r_{2,n} \\
&\left| L_n^{(b)} - \sqrt{\frac{\mathbb{V}[\pi_i]\|\mathbf{P}_{Z^\perp}\Omega^2\mathbf{P}_{Z^\perp}\|_{HS}^2}{2(T-2)} + \frac{n\left\|\mathbf{P}_{Z^\perp}\mu\mathbb{P}_n e_i\nu_i^{(1)}(D_i-\mathbb{E}[D_i])\right\|_2^2}{2(T-2)}} \right| \lesssim r_{2,n}.
\end{aligned} \tag{4.13}$$

Step 4: Now we can combine the bounds.. Let f_1 be the density function of $\|\zeta_{fixed}\|_2$. Note that $\|f_1\|_\infty \le C$. Using ζ_{fixed}, we get the following with probability at least $1-\frac{1}{n}$:

$$\sup_x \left| \mathbb{E}^{(b)}[\{\hat{L}_n^{(b)} \le x\}] - \mathbb{E}[\{\hat{L}_n \le x\}] \right| \lesssim \frac{1}{n} + \frac{(2k+T)^{\frac{7}{4}}}{\sqrt{n}} \log^3(n) + r_{1,n}\|f_1\|_\infty. \tag{4.14}$$

Let f_2 be the density of $\sqrt{\|\zeta_{limit}\|_2^2 + \frac{\mathbb{V}[\pi_i]\|\mathbf{P}_{Z^\perp}\Omega^2\mathbf{P}_{Z^\perp}\|_{HS}^2}{2(T-2)}}$. As long as $\|\mu\| \ge c\sqrt{T}$, we have that $\|f_2\|_\infty \le C$. Using ζ_{limit}, we get the following with probability at least $1-\frac{1}{n}$:

$$\sup_x \left| \mathbb{E}^{(b)}[\{\hat{L}_n^{(b)} \le x\}] - \mathbb{E}[\{\hat{L}_n \le x\}] \right| \lesssim \frac{1}{n} + \frac{(2k)^{\frac{7}{4}}}{\sqrt{n}} \log^3(n) + (r_{1,n}+r_{2,n})\|f_2\|_\infty. \tag{4.15}$$

The first inequality delivers the result when T is fixed, while the second one delivers the result when T is going to infinity and factors are bounded from below. □

Proof for Proposition 2.3

Proof. It is easy to see that $\hat{L}_n^{(b)}$ is invariant with respect to $\mathbb{E}[(D_i-\mathbb{E}D_i])\eta_i)$, and thus its approximation remains unchanged. At the same time, $|\hat{L}_n - L_n| \lesssim r_{1,n}$ (using the notation from the previous proof), and clearly L_n converges to inifinity as long as $\|\mathbb{E}[\nu_i^{(1)}(D_i-\mathbb{E}[D_i])]\|_2 > c$

□

E Proofs for Section 3

Proof. We define the following two matrices:

$$\begin{aligned}
\mathbf{P}_L &:= \begin{pmatrix} \mathcal{I}_T - \Psi\left(\Psi^\top \Psi\right)^{-1} \Psi^\top & \mathbf{0}_{T\times T} \\ \mathbf{0}_{T\times T} & \mathbf{0}_{T\times T} \end{pmatrix} \\
\mathbf{P}_R &:= \begin{pmatrix} \mathbf{0}_{T\times T} & \mathbf{0}_{T\times T} \\ \mathbf{0}_{T\times T} & \mathcal{I}_T - \Psi\left(\Psi^\top \Psi\right)^{-1} \Psi^\top. \end{pmatrix}
\end{aligned} \tag{5.1}$$

Under Assumption 3.1, the estimator has the following representation:

$$\hat{\tau}_{TS} - \tau = \frac{\tilde{Z}^\top \mathbf{P}_L \mathbb{P}_n (W_i - \mathbb{P}_n W_i)\epsilon_i^\top \mathbf{P}_R \tilde{Z}}{\tilde{Z}^\top \mathbf{P}_L \mathbb{P}_n (W_i - \mathbb{P}_n W_i) W_i^\top \mathbf{P}_R \tilde{Z}}. \tag{5.2}$$

We start the proof by stating several facts. For arbitrary fixed matrices A_1, A_2, and arbitrary random matrix A_3, we have the following with probability at least $1 - \frac{1}{T}$:

$$\begin{aligned}
&|\tilde{Z}^\top A_1 \tilde{Z} - \sigma_Z^2 \text{trace}(A)| \lesssim \sigma_Z^2 \|Z\|_{\psi_2}^2 \sqrt{\log(T)} \max\{\sqrt{\log(T)}\|A_1\|, \|A_1\|_{HS}\} \\
&\|\tilde{Z}^\top A_2\| \lesssim \sigma_Z \|Z\|_{\psi_2}^2 \|A_2\| \left(\sqrt{\text{rank}(A_2)} + C\sqrt{\log(T)}\right) \\
&|\tilde{Z}^\top A_3 \tilde{Z}| \lesssim \sigma_Z^2 \|A_3\| T,
\end{aligned} \tag{5.3}$$

where c_1 is a fixed constant. This follows directly from Assumption 3.3 and from the Hanson-Wright inequality. We define the sample covariance matrix of ν_i:

$$\Sigma_{n,\nu} := \mathbb{P}_n (\nu_i - \mathbb{P}_n \nu_i)\nu_i^\top. \tag{5.4}$$

The following inequality holds with probability at least $1 - \frac{1}{n}$:

$$\begin{aligned}
&\|\Sigma_{n,\nu} - \mathcal{I}_{2(T+K)}\| \lesssim \alpha := \frac{T}{n} \\
&|\mathbb{P}_n \pi_i (\pi_i - \mathbb{P}_n \pi_i) - \mathbb{V}[\pi_i]| \lesssim \sqrt{\frac{\log(n)}{n}} \|\pi_i\|_{\psi_2}^2.
\end{aligned} \tag{5.5}$$

This follows directly from Assumption A.4, Bernstein's inequality, and standard results on covariance matrices for sub-Gaussian random vectors (Theorem 6.5 in Wainwright [2019]). Using these facts, we get that the following inequalities hold with probability at least $1 - \frac{1}{n} - \frac{1}{T}$:

$$\begin{aligned}
&\tilde{Z}^\top \mathbf{P}_L \Sigma_u \mathbf{P}_R \tilde{Z} = \tilde{Z}^\top \mathbf{P}_L \mu_u \mu_u^\top \mathbf{P}_R \tilde{Z} + \tilde{Z}^\top \mathbf{P}_L H \Omega_u^2 H \mathbf{P}_R \tilde{Z} \lesssim \sigma_Z^2 \|Z\|_{\psi_2}^2 \log(T) \|\mu_u\|^2 + \sigma_Z^2 \|H\|^2 \|\Omega_u\|^2 T \\
&\tilde{Z}^\top \mathbf{P}_L \left(\hat{\Sigma}_u - \Sigma_u\right) \mathbf{P}_R \tilde{Z} \lesssim \alpha \sigma_Z^2 \left(\|Z\|_{\psi_2}^2 \log(T) \|\mu_u\|^2 + \|Z\|_{\psi_2}^4 \sqrt{T} \|\Omega_u\| \|H\|^2 \|\mu_u\| \sqrt{\log(T)} + \|H\|^2 \|\Omega_u\|^2 T \right) \\
&\tilde{Z}^\top \mathbf{P}_L \Sigma_{u,\epsilon} \mathbf{P}_R \tilde{Z} \lesssim \sigma_Z^2 \|Z\|_{\psi_2}^2 \log(T) \|\mu_u\| \|\mu_\epsilon\| + \sigma_Z^2 \|H\|^2 \|\Omega_u\| \|\Omega_\epsilon\| T \\
&\max\{|\tilde{Z}^\top \mathbf{P}_L \tilde{Z}|, |\tilde{Z}^\top \mathbf{P}_L \tilde{Z}|\} \lesssim \sigma_Z^2 T \\
&\max\{|\tilde{Z}^\top \mathbf{P}_L \mu_u \mathbb{E}[\nu_i^{(1)} \pi_i]|, |\tilde{Z}^\top \mathbf{P}_R \mu_u \mathbb{E}[\nu_i^{(1)} \pi_i]|\} \lesssim \sigma_Z \|Z\|_{\psi_2}^2 \|\mu_u\| \|\mathbb{E}[\nu_i^{(1)} \pi_i]\|_2 \sqrt{\log(T)} \\
&|\tilde{Z}^\top \mathbf{P}_R \mu_\epsilon \mathbb{E}[\nu_i^{(1)} \pi_i]| \lesssim \sigma_Z \|Z\|_{\psi_2}^2 \|\mu_\epsilon\| \|\mathbb{E}[\nu_i^{(1)} \pi_i]\|_2 \sqrt{\log(T)}
\end{aligned} \tag{5.6}$$

$$\begin{aligned}
\tilde{Z}^\top \mathbf{P}_L (\hat{\Sigma}_{u,\epsilon} - \Sigma_{u,\epsilon}) \mathbf{P}_R \tilde{Z} \lesssim{}& \alpha \sigma_Z^2 \|Z\|_{\psi_2}^2 \log(T) \|\mu_u\| \|\mu_\epsilon\| + \alpha \sigma_Z^2 \|Z\|_{\psi_2}^4 \sqrt{T} \|\Omega_u\| \|H\|^2 \|\mu_\epsilon\| \sqrt{\log(T)} + \\
& \alpha \sigma_Z^2 \|Z\|_{\psi_2}^4 \sqrt{T} \|\Omega_\epsilon\| \|H\|^2 \|\mu_u\| \sqrt{\log(T)} + \alpha \sigma_Z^2 \|H\|^2 \|\Omega_u\| \|\Omega_\epsilon\| T
\end{aligned} \tag{5.7}$$

$$\begin{aligned}
&|\mathbb{P}_n \pi_i (\pi_i - \mathbb{P}_n \pi_i) \tilde{Z}^\top \mathbf{P}_L \tilde{Z} \tilde{Z}^\top \mathbf{P}_R \tilde{Z} - \sigma_Z^4 \mathbb{V}[\pi_i] \mathrm{trace}(\mathbf{P}_R) \mathrm{trace}(\mathbf{P}_R)| \lesssim \\
&\qquad \mathbb{V}[\pi_i] \sigma_Z^4 \|Z\|_{\psi_2}^4 \log(T) \|\mathbf{P}_L\|_{HS} \|\mathbf{P}_R\|_{HS} + \sqrt{\alpha} \sigma_Z^4 \sqrt{\log(n)} \|\pi_i\|_{\psi_2}^2 T^{\frac{3}{2}}
\end{aligned} \tag{5.8}$$

$$\begin{aligned}
|\tilde{Z}^\top \mathbf{P}_R \left(\mathbb{P}_n \epsilon_i (\pi_i - \mathbb{P}_n[\pi_i]) - \mathbb{E}[\epsilon_i (\pi_i - \mathbb{E}[\pi_i])] \right)| \lesssim{}& k \sigma_Z \|Z\|_{\psi_2}^2 \|\mu_\epsilon\| \|\nu_i^{(1)}\|_{\psi_2} \|\|\pi_i - \mathbb{E}[\pi]\|_{\psi_2} \sqrt{\frac{\log(n)}{n}} \log(T) + \\
& \alpha \sigma_Z \|Z\|_{\psi_2}^2 \|H\| \|\Omega_\epsilon\| \|\nu_i^{(2)}\| \|\|(\pi_i - \mathbb{E}[\pi_i])\|_{\psi_2} \log(n)
\end{aligned} \tag{5.9}$$

$$\begin{aligned}
|\tilde{Z}^\top \mathbf{P}_L \left(\mathbb{P}_n u_i (\pi_i - \mathbb{P}_n[\pi_i]) - \mathbb{E}[u_i (\pi_i - \mathbb{E}[\pi_i])] \right)| \lesssim{}& k \sigma_Z \|Z\|_{\psi_2}^2 \|\mu_u\| \|\nu_i^{(1)}\|_{\psi_2} \|\|\pi_i - \mathbb{E}[\pi]\|_{\psi_2} \sqrt{\frac{\log(n)}{n}} \log(T) + \\
& \alpha \sigma_Z \|Z\|_{\psi_2}^2 \|H\| \|\Omega_u\| \|\nu_i^{(2)}\| \|\|(\pi_i - \mathbb{E}[\pi_i])\|_{\psi_2} \log(n).
\end{aligned} \tag{5.10}$$

We define two errors:

$$\begin{aligned}
&e_{num} := \tilde{Z}^\top \mathbf{P}_L \hat{\Sigma}_{W,Y} \mathbf{P}_R \tilde{Z} - \sigma_Z^2 \tilde{Z}^\top \mathbf{P}_R \mu_\epsilon \mathbb{E}[\nu_i^{(1)} \pi_i] \mathrm{trace}(\mathbf{P}_L) \\
&e_{den} := \tilde{Z}^\top \mathbf{P}_L \hat{\Sigma}_W \mathbf{P}_R \tilde{Z} - \sigma_Z^4 \mathbb{V}[\pi_i] \mathrm{trace}(\mathbf{P}_R) \mathrm{trace}(\mathbf{P}_L).
\end{aligned} \tag{5.11}$$

We can express the estimator as follows:

$$\hat{\tau}_{TS} - \tau = \frac{e_{num} + \sigma_Z^2 \tilde{Z}^\top \mathbf{P}_R \mu_\epsilon \mathbb{E}[\nu_i^{(1)} \pi_i] \mathrm{trace}(\mathbf{P}_L)}{e_{den} + \sigma_Z^4 \mathbb{V}[\pi_i] \mathrm{trace}(\mathbf{P}_R) \mathrm{trace}(\mathbf{P}_L)}. \tag{5.12}$$

Using the results above, we have the following bounds:

$$
\begin{aligned}
|e_{num}| &\lesssim \sigma_Z^2(1+\alpha)\log(T)T + \sigma_Z^3\alpha\log(T)\sqrt{\log(n)}T + \sigma_Z^3\log(T)T\|\mathbb{E}[\nu_i^{(1)}\pi\|_2 \\
|e_{den}| &\lesssim \sigma_Z^2(1+\alpha)\log(T)T + \sigma^3\sqrt{\log(T)}T(1+\alpha\sqrt{\log(n)}\log(T)) + \sqrt{\alpha}\sigma_Z^4 T^{\frac{3}{2}}.
\end{aligned} \tag{5.13}
$$

In the regime where α is constant, we get the following with probability at least $1 - \frac{1}{n} - \frac{1}{T}$:

$$
\begin{aligned}
&|\hat{\tau}_{TS} - \tau| \lesssim \frac{\log(T)}{\mathbb{V}[\pi]\sigma_Z^2 T} + \frac{\log(T)\sqrt{\log(n)}}{\sigma_Z T} + \frac{\log(T)\|\mathbb{E}[\nu_i^{(1)}\pi_i]\|_2}{\sigma_Z\mathbb{V}[\pi_i]\sqrt{T}} \\
&\frac{\sigma_z\mathbb{V}[\pi_i]\text{trace}(\mathbf{P}_R)}{\|\mathbf{P}_L\mu_\epsilon\mathbb{E}[\nu_i^{(1)}\pi_i]\|_2}\left|\hat{\tau}_{TS} - \tau - \frac{\tilde{Z}^\top\mathbf{P}_L\mu_\epsilon\mathbb{E}[\nu_i^{(1)}\pi_i]}{\sigma_Z^2\mathbb{V}[\pi_i]\text{trace}(\mathbf{P}_L)}\right| \lesssim \frac{\log(T)}{\sigma_Z\mathbb{V}[\pi]\sqrt{T}}\left(\sqrt{\log(n)} + \frac{1}{\|\mathbb{E}[\nu_i^{(1)}\pi_i]\|_2}\right).
\end{aligned} \tag{5.14}
$$

We define the following variables:

$$
\begin{aligned}
\xi_{est} &:= \frac{\tilde{Z}^\top\mathbf{P}_L\mu_\epsilon\mathbb{E}[\nu_i^{(1)}\pi_i]}{\sigma_Z^2\mathbb{V}[\pi_i]\text{trace}(\mathbf{P}_L)} \\
\xi_{lim} &\sim \mathcal{N}(0,1) \\
\sigma_{est} &:= \frac{\|\mathbf{P}_L\mu_\epsilon\mathbb{E}[\nu_i^{(1)}\pi_i]\|_2}{\sigma_z\mathbb{V}[\pi_i]\text{trace}(\mathbf{P}_L)}.
\end{aligned} \tag{5.15}
$$

The rest follows from standard arguments (Berry-Esseen's theorem):

$$
\begin{aligned}
&\sup_x\left|\mathbb{E}\left[\left\{\frac{\hat{\tau}_{TS}-\tau}{\sigma_{est}} \le x\right\}\right] - \mathbb{E}\left[\{\xi_{lim} \le x\}\right]\right| \le \\
&\frac{c_1}{T} + \frac{c_2}{n} + 3\sup_x\left|\mathbb{E}\left[\{\xi_{lim} \le x\}\right] - \mathbb{E}\left[\left\{\frac{\xi_{est}}{\sigma_{est}} \le x\right\}\right]\right| + \sup_x\left|\mathbb{E}\left[\{\xi_{lim} \le x\}\right] - \mathbb{E}\left[\{\xi_{lim} \le x + r_T\}\right]\right| \lesssim \\
&\qquad \frac{\log(T)}{\sigma_Z\mathbb{V}[\pi]\sqrt{T}}\left(\sqrt{\log(n)} + \frac{1}{\|\mathbb{E}[\nu_i^{(1)}\pi_i]\|_2}\right).
\end{aligned} \tag{5.16}
$$

□

F Proofs for Section 4

Proof for Proposition 4.1 We need to prove that H_t is not present in the aggregate model, but this follows directly from Assumption 4.4, because $\mathbb{E}[\varepsilon_{it}|\mathcal{I}_{t-l}] = 0$ and the weights average up to zero. We also need to prove that $\pi_{t|0} > 0$. This follows from the fact that $D_i - \mathbb{E}[D_i] = \frac{1}{T_0}\sum_{t \leq T_0}(\pi_{it} - \mathbb{E}[\pi_{it}])\tilde{Z}^2_{t|0} +$ noise, where $\mathbb{E}[\text{noise} \times \pi_{it}] = 0$. Then the result follows from the first part of Assumption 4.4.

Proof for Proposition 4.2 The result follows from the fact that $\alpha_{t|T_0}, \beta_{t|T_0}, \delta_{t|T_0}, \pi_{t|T_0}, \mu_{t|T_0}$ are functions of $\{Z_l\}_{l \leq T_0}$ and thus are uncorrelated with $\tilde{Z}_{t|T_0}$. Also $\mathbb{E}[\pi_{t|T_0}\tilde{Z}^2_{t|T0}] = \mathbb{E}[\pi_{t|T_0}\mathbb{E}[\tilde{Z}^2_{t|T0}|\mathcal{I}_{T_0}]] = \mathbb{E}[\pi_{t|T_0}\mathbb{E}[\tilde{Z}^2_{t|T0}]] = \pi_{t|0}\mathbb{E}[\tilde{Z}^2_{t|T0}]$.

Abstrakt

Navrhujeme obecný postup pro odhad kauzálního vlivu změn v kontextu, kde jediným zdrojem exogenní variace je časová řada agregátních šoků. Nejdříve argumentujeme, že běžně používané postupy odhadu ignorují klíčové vlastnosti časové dimenze dat. Dále vyvíjíme grafický nástroj a nový test k ilustraci problematiky naší metody s využitím dat z vlivných studií v oboru ekonomického rozvoje [Nunn and Qian, 2014] a makroekonomie [Nakamura a Steinsson, 2014]. Motivováni těmito studiemi, konstruujeme nový odhad, který je založen na modelu časových řad pro agregátní šoky. Analyzujeme statistické vlastnosti našeho odhadu v situacích relevantních v praxi, kdy průřezová i časová dimenze je podobné velikosti. Nakonec poskytujeme kauzální interpretaci pro náš odhad a analyzujeme nový kauzální model, který umožňuje jak rozsáhlou nepozorovanou heterogenitu v potenciálních výstupech, tak nepozorované agregátní šoky. ¨

Working Paper Series
ISSN 1211-3298
Registration No. (Ministry of Culture): E 19443

Individual researchers, as well as the on-line and printed versions of the CERGE-EI Working Papers (including their dissemination) were supported from institutional support RVO 67985998 from Economics Institute of the CAS, v. v. i.

Specific research support and/or other grants the researchers/publications benefited from are acknowledged at the beginning of the Paper.

Published by
Charles University, Center for Economic Research and Graduate Education (CERGE)
and
Economics Institute of the CAS, v. v. i. (EI)
CERGE-EI, Politických vězňů 7, 111 21 Prague 1, tel.: +420 224 005 153, Czech Republic.
Printed by CERGE-EI, Prague
Subscription: CERGE-EI homepage: http://www.cerge-ei.cz

Phone: + 420 224 005 153
Email: office@cerge-ei.cz
Web: http://www.cerge-ei.cz

Editor: Byeongju Jeong

The paper is available online at http://www.cerge-ei.cz/publications/working_papers/.

ISBN 978-80-7343-464-9 (Univerzita Karlova, Centrum pro ekonomický výzkum a doktorské studium)
ISBN 978-80-7344-533-1 (Národohospodářský ústav AV ČR, v. v. i.)