How Much Should We Trust Regional-Exposure Designs?*

Jeremy Majerovitz[†] Karthik A. Sastry[‡]

July 1, 2023

Abstract

Many prominent studies in macroeconomics, labor, and trade use panel data on regions to identify the local effects of aggregate shocks. These studies construct *regional-exposure instruments* as an observed aggregate shock times an observed regional exposure to that shock. We argue that the most economically plausible source of identification in these settings is uncorrelatedness of observed and unobserved aggregate shocks. Even when the regression estimator is consistent, we show that inference is complicated by cross-regional residual correlations induced by unobserved aggregate shocks. We suggest two-way clustering, two-way heteroskedasticity- and autocorrelation-consistent standard errors, and randomization inference as options to solve this inference problem. We also develop a feasible optimal instrument to improve efficiency. In an application to the estimation of regional fiscal multipliers, we show that the standard practice of clustering by region generates confidence intervals that are too small. When we construct confidence intervals with robust methods, we can no longer reject multipliers close to zero. The feasible optimal instrument more than doubles statistical power; however, we still cannot reject low multipliers. Our results underscore that the precision promised by regional data may disappear with correct inference.

Keywords: Applied Econometrics, Regional Data, Shift-Share Instruments

JEL: C12, C18, C21, C23, C26, F16, R12

---

*We are grateful to Alberto Abadie, Daron Acemoglu, Isaiah Andrews, George-Marios Angeletos, David Atkin, Martín Beraja, Gabriel Chodorow-Reich, Bill Dupor, Peter Hull, Michal Kolesár, Daniel Lewis, Karel Mertens, Anna Mikusheva, Emi Nakamura, Ben Olken, Jón Steinsson, and to seminar participants at MIT, KU Leuven, and the Barcelona Summer Forum for helpful comments. The views expressed in this paper are those of the authors and do not necessarily reflect the views of the Federal Reserve Bank of St. Louis or the Federal Reserve System.

†Federal Reserve Bank of St. Louis; Email: jeremy.majerovitz@gmail.com

‡Department of Economics, Princeton; Email: sastry.karthik@gmail.com

1 Introduction

One of the most popular research designs in economics exploits locations' heterogeneous exposure to aggregate shocks to measure the local effects of those shocks. Concretely, consider a setting with time periods indexed by t, regions indexed by i, an aggregate shock vector S_t, and region-specific exposure vectors η_i. In the *regional-exposure design*, researchers construct an instrument $Z_{it} = \eta_i' S_t$ and use it to estimate how endogenous variable X_{it} affects outcome Y_{it} in a linear model. This empirical strategy is ubiquitous—for example, it has been used to estimate the regional fiscal multiplier (Nakamura and Steinsson, 2014), the inverse labor supply elasticity (Bartik, 1991), the effects of falling home prices on employment (Mian and Sufi, 2014), the effects of import competition from China (Autor et al., 2013) and immigration (Card, 2001) on local labor markets, and the effects of foreign aid (Nunn and Qian, 2014) and commodity price shocks (Dube and Vargas, 2013) on conflict. Literatures across fields, from macroeconomics to labor to political economy, rely on results from this research design.

Studies using the regional-exposure design often construct standard errors clustered by region. This approach to inference presumes that regression model residuals are uncorrelated across regions. But when regions are heterogeneously affected by aggregate shocks, the assumption of uncorrelated residuals across regions is unlikely to hold. Moreover, practitioners are often unclear about what assumptions underlie their identification strategy.

This paper studies how identification and inference in regional-exposure designs is affected by unobserved aggregate shocks. We show that clustering by region substantially understates true uncertainty both in theory and in practice. In a placebo test based on the study of Nakamura and Steinsson (2014), which estimates the regional fiscal multiplier in an annual panel of US states, a state-clustered 5% test falsely rejects the null more than 25% of the time. We provide alternative standard errors that are robust to cross-regional correlation as well as a randomization inference approach that provides exact coverage in finite samples. Since true statistical uncertainty is often high in these settings, we also develop a feasible optimal instrument that reweights data to improve efficiency in light of correlated residuals across regions. In our application, this new estimator more than doubles statistical power. Our results highlight the importance of accounting for the correlation of residuals across regions in regional-exposure settings.

Framework. We use the idea of an *approximate factor structure* to the residual to capture the notion that the residual contains aggregate shocks that have heterogeneous effects across regions. Under this structure, the residual contains a factor component, reflecting heterogeneous regional loadings of an aggregate shock, and an idiosyncratic component. More formally, we write the residual as $u_{it} = \lambda_i' F_t + \varepsilon_{it}$, where λ_i is the unobserved factor loading, F_t is the vector of unobserved factors, and ε_{it} is the idiosyncratic component.

We use the factor structure to clarify that identification relies on either as-good-as-random assignment of the aggregate shock or as-good-as-random assignment of the regional exposures to that shock. Given the structure of the instrument and the residual, there are two leading sufficient conditions for instrument exogeneity: (i) the aggregate shock, S_t, being orthogonal to the factor shock, F_t, or (ii) the exposure, η_i, being orthogonal to the unobserved factor loading, λ_i. We view this latter possibility as unlikely because it is easily contradicted by the data: the regional exposure is typically strongly correlated with several other important regional variables, which themselves may be factor loadings. Practitioners thus need to argue why the shocks are quasi-randomly assigned.

We next show that the validity of clustering by region depends critically on the source of identification. If identification were to come from as-good-as-random assignment of shares, then clustering by region would yield valid confidence intervals, although we view this as unlikely in practice. Otherwise, the standard practice of clustering by region will typically yield invalid confidence intervals. Intuitively, two regions with similar unobserved factor loadings, λ_i, will face common shocks, F_t. For example, Boston and San Francisco both have a large concentration of educated technology workers, are therefore exposed to aggregate shocks to the "high-tech" sector, and as a result may have correlated residuals. If η_i, the observed exposure to the aggregate shock, is correlated with the factor loadings, then two regions with similar exposures to the observed shock will have correlated residuals. This invalidates the typical approach of clustering by region.

Proposed Solutions. We next show how to construct valid confidence intervals using methods that are robust to correlated shocks across regions.

We first suggest more robust clustered standard errors. If the factors are uncorrelated across time, then two-way clustering is valid. If factors are correlated across time, but that correlation dies out asymptotically, researchers can use the method of Thompson (2011) that combines two-way clustering with a heteroskedasticity and autocorrelation correction à la Driscoll and Kraay (1998). We also highlight the potential importance of pairing valid clustering methods with weak-instrument robust tests, such as that of Anderson and Rubin (1949), since the first stage may be weaker than suggested by invalid inference techniques.

We next introduce a randomization inference approach. In randomization inference, we hold the residuals fixed and instead consider alternative draws of the shocks, S_t. Because this method makes no assumptions about the residual, it can accommodate an arbitrary correlation structure, including the factor structure we study. Randomization inference instead requires specification of the shock process. The researcher can then simulate the exact distribution of the test statistic under the null, and thus construct confidence intervals that have exact coverage even in finite samples.

These methods often reveal a lack of statistical power once coverage is corrected by accounting for the residuals' cross-regional correlation. We thus also propose a method to construct a feasible optimal instrument, in the spirit of Chamberlain (1987, 1992) and Borusyak and Hull (2021a). The optimal instrument reweights the original instrument based on the inverse covariance matrix of the residuals. Our feasible analog models this covariance via the factor structure. We show how to estimate this structure via principal components analysis and provide a method to select tuning parameters to maximize power.

Application: Regional Fiscal Multipliers. We show that these issues are quantitatively important in an application to the estimation of regional fiscal multipliers. Nakamura and Steinsson (2014) use the interaction of growth in national defense spending with individual states' exposure to that spending as an instrument to estimate the *regional fiscal multiplier*. We first show that there is a factor structure to the residual: the first two principal components explain over 60% of the variance. To study the performance of inference strategies in practice, we conduct a placebo simulation with fake spending shocks. Consistent with our results, we find that conventional tests at the 5% level based on clustering by state falsely reject the null more than 25% of the time. We find that the more robust clustering (two-way clustering or two-way HAC) gives substantially better coverage. Our randomization inference procedure, by construction, gives exact size.

In the data, valid confidence intervals cannot rule out low values of the regional fiscal multiplier with high precision. At the 5% level, in our preferred specification, randomization inference cannot rule out fiscal multipliers as low as 0.1 and robust confidence intervals cannot rule out 0. Both methods provide evidence for a multiplier greater than 0.46 at the 10% level and greater than 1 at the 32% level.

Implementing the feasible optimal instrument substantially improves power. We find in a power simulation that a test based on the optimal instrument is up to 2.75 times more likely to correctly reject the null of zero multiplier against a calibrated alternative in which the multiplier is 1.5, compared to a test based on the original IV. In principle, the optimal instrument can provide a much sharper estimate of the regional fiscal multiplier. In practice, the optimal instrument provides tighter confidence intervals but lowers the point estimate, and so we are still unable to reject values near zero.

Our results contrast with those of the original paper, which conducts asymptotic inference clustered by state and reports that 95% confidence intervals rule out multipliers below 0.6. Interpreting these results via a model, the authors argue that the data are more consistent with a New Keynesian model than with a "plain-vanilla Neoclassical model." Our analysis suggests that the data favor the New Keynesian model, but with considerably less precision than state-clustered standard errors would suggest.

Three Recommendations for Practice. First, we caution against clustering standard errors by region. This can lead to severe distortions in inference in the likely case that identification comes from aggregate shocks and regions are affected by other unobserved common shocks. We demonstrate severe under-coverage in our empirical example.

Second, for valid inference, we recommend two options. Researchers can use valid clustering methods, such as two-way clustering or two-way HAC. Alternatively, researchers can use randomization inference to obtain exact finite-sample coverage at the cost of needing to model the data-generating process for shocks.

Third, to improve precision, we suggest considering the feasible optimal-instrument procedure. As we showed in the application, this method can significantly improve statistical power. Implementing this method requires reasoning about the relevant null and alternative hypotheses that may be specific to the researcher's setting. In practice, it may be most useful in settings in which researchers have informed priors over the parameters of interest, such as the fiscal multipliers setting.

Related Literature. Our work relates to a growing literature on inference and estimator design in regional-exposure settings, of which the "shift-share" design is a special case (e.g., Adao et al., 2019; Goldsmith-Pinkham et al., 2020; Borusyak et al., 2022). We discuss the detailed relationship to this work in Section 3.4.

Our focus on randomization inference and efficient estimation as useful tools in settings with non-random exposure to aggregate shocks connects with, respectively, Borusyak and Hull (2021b) and Borusyak and Hull (2021a). We use the idea of a factor structure to clarify the value of these tools. Moreover, our implementation of the feasible optimal instrument focuses on reweighting the data to account for cross-observation covariance in the residual, an issue on which Borusyak and Hull (2021a) do not focus.

The most closely related work to ours is Arkhangelsky and Korovkin (2019). These authors also study regional-exposure settings in which identification comes from aggregate shocks and observe that a critical confounding force is unobserved aggregate shocks with heterogeneous exposure. They propose a split-sample estimator that minimizes the effects of these shocks to improve efficiency, inspired by the synthetic controls literature (Abadie and Gardeazabal, 2003; Abadie et al., 2010). We, by contrast, focus more closely on inference issues with the standard IV estimator and also propose a randomization inference approach. The new estimator that we propose is inspired by the optimal-instrument literature (Chamberlain, 1987, 1992; Borusyak and Hull, 2021a). We view our results as highly complementary to theirs. Together, they comprise an improved toolkit for estimation and inference in the regional-exposure setting.

Our fiscal-multipliers application relates to a growing literature on estimating cross-regional spending multipliers (reviewed by Chodorow-Reich, 2019) and, more broadly, connecting macroeconomic theory to econometric practice in similar settings (e.g., Chodorow-Reich, 2020; Guren et al., 2021). Most prior work in this area has focused on the economic interpretation of estimates. Our focus is instead on accurately reporting the precision of regional estimates and improving their efficiency, holding fixed their interpretation. Insofar as our results suggest that regional multiplier estimates are relatively imprecise, our results further highlight the importance of cross-study meta-analysis (e.g., as in Chodorow-Reich, 2019) for obtaining reliable estimates.

Finally, our analysis fits into a literature that gives practical guidance to researchers about selecting an appropriate level of standard-error clustering (e.g., Bertrand et al., 2004; MacKinnon et al., 2022; Abadie et al., 2023). Compared to these general analyses, our analysis uses a plausible economic structure, the regional factor structure, to propose and evaluate variance estimators in our setting.

Outline. Section 2 introduces the model with a residual factor structure and highlights identification and inference issues that arise. Section 3 proposes econometric solutions. Section 4 studies an application to estimating regional fiscal multipliers. Section 5 concludes.

2 Model and Econometric Issues

We first formally describe the *regional-exposure design*, which uses the interaction of observed aggregate shocks with observed regional heterogeneity as an instrument to estimate the relationship between an endogenous regressor and an outcome. To capture the possibility that other, unobserved shocks also have regionally heterogeneous effects on the outcome, we assume that the residual of the structural equation has an approximate factor structure. Under this structure, we clarify the assumptions under which the model is identified and the assumptions under which standard econometric practice of clustering standard errors by region yields correct inference. We argue that, in most applications, regional exposures are endogenous to economic conditions while aggregate shocks may be as-good-as-randomly assigned. In this case, clustering by region is invalid.

2.1 Set-up: The Regional-Exposure Model

There is a set of regions $i \in \{1, \ldots, N\}$ and a set of time periods $t \in \{1, \ldots, T\}$.[1] In each period there is a vector-valued aggregate shock $S_t \in \mathbb{R}^K$, for $K \geq 1$. Each region has an exposure $\eta_i \in \mathbb{R}^K$ to each dimension of the shock. We define the *regional-exposure instrument*

$$Z_{it} = \eta_i' S_t \tag{1}$$

There is an endogenous outcome $Y_{it} \in \mathbb{R}$ and an endogenous regressor $X_{it} \in \mathbb{R}$.

We study the two-equation instrumental-variables model

$$Y_{it} = \alpha_t + \gamma_i + \beta \cdot X_{it} + u_{it} \tag{2}$$

$$X_{it} = \omega_t + \zeta_i + \pi \cdot Z_{it} + e_{it} \tag{3}$$

We refer to these equations, respectively, as the "structural equation" and the "first-stage equation." The parameter of interest is $\beta \in \mathbb{R}$, the marginal effect of X_{it} on Y_{it}. The parameter $\pi \in \mathbb{R}$ is the first-stage coefficient and $(\alpha_t, \omega_t)_{t=1}^T$ and $(\gamma_i, \zeta_i)_{i=1}^N$ are fixed effects. The variables u_{it} and e_{it} are defined as residuals, which have zero mean in each time period and in each region. We also define variables $\tilde{X}_{it}$, $\tilde{Y}_{it}$, $\tilde{Z}_{it}$, $\tilde{u}_{it}$, and $\tilde{e}_{it}$ as the double-demeaned counterparts to the original variables.[2] For simplicity of exposition, we assume that X_{it}, Y_{it}, and Z_{it} have zero mean across regions and time-periods and that the econometrician observes a balanced panel of these variables. For technical simplicity, we assume that all moments of the form $\mathbb{E}[X_{it}^{a_X} Y_{js}^{a_Y} Z_{kr}^{a_Z}]$, for $(a_X, a_Y, a_Z) \geq 0$ and indices $(i, j, k) \in \{1, \ldots N\}$ and $(t, s, r) \in \{1, \ldots, T\}$, exist and are finite.

To illustrate this set-up, we describe Nakamura and Steinsson (2014)'s study of regional fiscal multipliers in our language. In their setting, β is the regional fiscal multiplier; Y_{it} is the two year growth rate in state GDP per capita; and X_{it} is the two year change in local military procurement spending per capita, divided by the two year lagged state GDP. In defining the instrument $Z_{it} = \eta_i S_t$, S_t is national military procurement spending growth and η_i is military procurement spending as a share of state GDP at the start of the sample.

[1] Although most applications of the regional-exposure instrument rely on regional data, our results could also be applied in settings with other types of cross-sectional units, such as firms, households, or individuals.

[2] That is, for each variable $W \in \{X, Y, Z, u, e\}$, $\tilde{W}_{it} := W_{it} - \bar{W}_i - \bar{W}_t + \bar{W}$, where $\bar{W}$ denotes the sample average, and $\bar{W}_i, \bar{W}_t$ denote the within-region and within-time-period sample averages respectively.

2.2 The Residual Factor Structure

We assume that the residual u_{it} of Equation 2 has an approximate factor structure. To capture this, we define a factor shock vector $F_t \in \mathbb{R}^J$, with $J \geq 1$, a collection of factor loadings $\lambda_i \in \mathbb{R}^J$ for each region i, and an idiosyncratic component ε_{it} which is independent from λ_i and F_t, and has zero mean in each time period and in each region. We define λ_i and F_t to each have mean zero. We write u_{it} as

$$u_{it} = \lambda_i' F_t + \varepsilon_{it} \tag{4}$$

We introduce the approximate factor structure because it parsimoniously captures the notion that different regions may comove in response to aggregate conditions. For example, regions with a similar industrial mix may comove in response to certain trade shocks, certain regions may be more sensitive to fiscal and monetary policy, or urban areas may comove as the returns to agglomeration rise or fall.

Moreover, assuming a residual factor structure is arguably the only way to be internally consistent with constructing a regional-exposure instrument. If we are to take seriously the various studies relying on the interaction between observed shocks and exposures as part of their research design, then we must believe that the residual contains the many such regionally heterogeneous shocks studied in other papers. If we believe Nakamura and Steinsson (2014), who find regionally heterogeneous effects of national military procurement spending on output through its effect on local defense procurement, and we believe Autor et al. (2013), who find regionally heterogeneous effects of rising trade with China, then the regional-exposure instrument of one study is in the residual of the other, and vice-versa.

Although the approximate factor structure is more flexible than the typical assumption of i.i.d. errors (or errors that are independent across regions), it is not entirely unrestrictive for the covariance of the error term across regions. However, our proposed solutions will typically not rely on the factor structure: the confidence intervals we recommend will be robust to a broader set of covariance structures in the residual, and our improved estimator will still offer efficiency improvements even if the residual does not truly have an approximate factor structure.

2.3 Identification: "From Shares" or "From Shocks"

As a prelude to our analysis, we recast the conditions under which an instrumental variables strategy with Z_{it} identifies the structural parameter β in Equations 2 and 3. We argue that this helps separate two logical paths to identification, one via the assignment of exposures and the other via the assignment of shocks.

To do this, we will maintain two assumptions for the remainder of the analysis. The first assumption is that the cross-sectional variables are independent from the time series variables.

Assumption 1. $(\eta_i, \lambda_i) \perp\!\!\!\perp (S_t, F_t)$

In essence, the properties of the regions that are drawn cannot affect the time series shocks, and vice-versa. This assumption might be violated, for example, if a financial crisis will only occur if certain regions are very indebted. The second assumption is that the idiosyncratic residual component is uncorrelated with the instrument.

Assumption 2. $\mathbb{E}\left[Z_{it}\varepsilon_{it}\right] = 0$.

This is without loss of generality. If the idiosyncratic component of the residual were correlated with the instrument, then it could be decomposed into the projection of ε_{it} onto $Z_{it} := \eta_i' S_t$ and the residual of that projection, which would be uncorrelated with Z_{it}. The projection of ε_{it} onto Z_{it} would have a factor structure by construction. Thus, any component of the residual which is correlated with the regressor can be represented as having a factor structure.

We next use these assumptions to unpack the exogeneity condition $\mathbb{E}\left[Z_{it}u_{it}\right] = 0$. In particular, we first use Equations 1 and 4 to write

$$\mathbb{E}\left[Z_{it}u_{it}\right] = \mathbb{E}\left[\eta_i' S_t \left(\lambda_i' F_t + \varepsilon_{it}\right)\right] = \mathbb{E}\left[\eta_i' S_t \cdot \lambda_i' F_t + \eta_i' S_t \cdot \varepsilon_{it}\right] \tag{5}$$

By Assumption 2, the second term is zero. We next manipulate the first term to write

$$\begin{aligned}
\mathbb{E}\left[Z_{it}u_{it}\right] &= \mathbb{E}\left[\eta_i' S_t \cdot \lambda_i' F_t\right] \\
&= \mathbb{E}\left[S_t'(\eta_i\lambda_i')F_t\right] \\
&= \mathbf{tr}\left(\mathbb{E}\left[S_t'(\eta_i\lambda_i')F_t\right]\right) \\
&= \mathbf{tr}\left(\mathbb{E}\left[(\eta_i\lambda_i')(F_t S_t')\right]\right) = \mathbf{tr}\left(\mathbb{E}\left[\eta_i\lambda_i'\right]\mathbb{E}\left[F_t S_t'\right]\right)
\end{aligned} \tag{6}$$

where **tr** denotes the trace of a matrix and, in the last line, we use the cyclic property and Assumption 1. Observe that $\mathbb{E}\left[\eta_i\lambda_i'\right]$ is a $K \times J$ matrix and $\mathbb{E}\left[F_t S_t'\right]$ is a $J \times K$ matrix, so the trace is over a $K \times K$ matrix.

Using this simplification, we argue there are two primary *sufficient conditions* for the identification condition $\mathbb{E}\left[Z_{it}u_{it}\right] = 0$. We state each below.

Condition 1 (Identification from Shares)**.** The regional exposures are uncorrelated with the factor loadings, or $\mathbb{E}[\eta_i\lambda_i'] = 0$.

Condition 2 (Identification from Shocks)**.** The aggregate shocks are uncorrelated with the factor shocks, or $\mathbb{E}[F_t S_t'] = 0$.

The first condition is natural if the exposures, η_i, are as-good-as-randomly assigned. We refer to this condition as *identification from shares*, reflecting its connection to the literature on shift-share instruments. In the shift-share setting, η_i is a vector of industrial employment shares, and this condition is equivalent to assuming that the industry shares are as-good-as-randomly assigned. In the shift-share literature, this is the route to identification assumed by Goldsmith-Pinkham et al. (2020).

The second condition is natural if we assume that the shocks, S_t, are as-good-as-randomly assigned. In the shift-share literature, this is the route to identification assumed by Adao et al. (2019) and Borusyak et al. (2022).

Of these two paths to identification, we view identification from shocks as more plausible. In typical applications, it is easy to show that the exposures, η_i, are correlated with other variables (these variables themselves being plausible potential factor loadings, λ_i), and thus are clearly not as-good-as-randomly assigned. Of course, the fact that identification from shares is dubious does not imply that identification from shocks is necessarily any more plausible. Regardless, we believe that if either of these identification approaches works, it is likely to be identification from shocks.

We moreover view these two sufficient conditions as the main routes to identification, since the others that are possible in principle are harder to justify economically. Mathematically, there exist many matrices $\mathbb{E}\left[\eta_i \lambda_i'\right] = Q$ and $\mathbb{E}\left[F_t S_t'\right] = R$ such that $\mathbf{tr}(QR) = 0$, but neither $Q = 0$ (Condition 1) nor $R = 0$ (Condition 2). For instance, we could also mix-and-match conditions (e.g. $\eta_{i1} \perp \lambda_{i1}$ and $S_{i2} \perp F_{i2}$). But this is unappealing in practice, since it requires a just-so combination of orthogonality conditions. An especially unappealing path to identification would be to assume that the individual bias terms $\mathbb{E}\left[\eta_i^k S_t^k \lambda_i^h F_t^h\right]$ do not equal zero, but that they happen to cancel out, so that their sum is zero. This would require an extraordinary coincidence.

Both identification from shares and identification from shocks are sufficient, when combined with appropriate conditions on dependence and second moments, for the IV estimate to converge in probability to the true β. More specifically, the former relies on a weak law of large numbers in the many-regions limit, and the latter relies on a weak law of large numbers in the many-time-periods limit. We state this formally below. The proof of this and all subsequent results is in Appendix A.

Proposition 1 (Convergence of the IV Estimator)**.** *Assume that $\mathbb{E}[\tilde{Z}_{it}\tilde{X}_{it}']$ is finite and full rank (instrument relevance). The following are true:*

1. *If Condition 1 holds and* $\left(\eta_i, \lambda_i, (\varepsilon_{it})_{t=1}^T, (e_{it})_{t=1}^T\right)$ *are drawn i.i.d. across regions, then* $\hat{\beta} \xrightarrow{p} \beta$ *as* $N \to \infty$.
2. *If Condition 2 holds and* $\left(S_t, F_t, (\varepsilon_{it})_{i=1}^N, (e_{it})_{i=1}^N\right)$ *are stationary and strongly mixing across time, then* $\hat{\beta} \xrightarrow{p} \beta$ *and* $T \to \infty$.

2.4 From Identification to Inference

We now turn our focus to inference and ask: how do assumptions about the sources of identification affect the validity of common strategies for inference? We show that clustering by region is likely invalid in settings where identification comes from shocks. We will use this finding to motivate our analysis of econometric solutions in Section 3.

Unpacking The Asymptotic Variance of $\hat{\beta}$

Whichever of our two routes to identification we rely on, the instrumental variables estimator will have an asymptotic variance of the familiar "sandwich" form:[3]

$$\text{AVAR}\left(\sqrt{N} \cdot \hat{\beta}\right) = \mathbb{E}\left[\tilde{Z}_{it}\tilde{X}_{it}'\right]^{-1} \text{AVAR}\left(\frac{1}{\sqrt{N}} \cdot \frac{1}{T} \sum_{i,t} \tilde{Z}_{it}\tilde{u}_{it}\right) \mathbb{E}\left[\tilde{X}_{it}\tilde{Z}_{it}'\right]^{-1} \tag{7}$$

The "bread" of this expression, $\mathbb{E}\left[\tilde{Z}_{it}\tilde{X}_{it}'\right]^{-1}$, is straightforward to estimate. We are primarily concerned with the middle, "meat" term, which we denote as

$$\Omega := \text{AVAR}\left(\frac{1}{\sqrt{N}} \cdot \frac{1}{T} \sum_{i,t} \tilde{Z}_{it}\tilde{u}_{it}\right) \tag{8}$$

We next use Equation 4, or the factor structure of $\tilde{u}_{it}$, to simplify this term:[4]

$$\begin{aligned} \Omega = &\ \text{AVAR}\left(\frac{1}{\sqrt{N}} \cdot \frac{1}{T} \sum_{i,t} \tilde{Z}_{it}\tilde{\lambda}_i'\tilde{F}_t\right) + \text{AVAR}\left(\frac{1}{\sqrt{N}} \cdot \frac{1}{T} \cdot \sum_{i,t} \tilde{Z}_{it}\tilde{\varepsilon}_{it}\right) \\ &+ 2 \cdot \frac{1}{T} \sum_{i,t} \sum_{j,s} \mathbb{E}\left[\tilde{Z}_{it}\tilde{Z}_{js}\tilde{\lambda}_i'\tilde{F}_t\tilde{\varepsilon}_{js}\right] \end{aligned} \tag{9}$$

[3] Our analysis of the asymptotic variance will assume that $N \to \infty$, and will rely on $\sqrt{N}$ asymptotics, consistent with the assumptions behind clustering by region. When we move to the two-way clustering setting, we will also require $T \to \infty$.

[4] Note that the double-demeaning "passes through" the factor structure. That is, $\tilde{u}_{it} = \tilde{\lambda}_i'\tilde{F}_t + \tilde{\varepsilon}_{it}$, where $\tilde{\lambda}_i = \lambda_i - \bar{\lambda}$ and $\tilde{F}_t = F_t - \bar{F}$.

For the remaining results in this paper, we will strengthen Assumption 2 to the following:

Assumption 3. *For all i, $(\varepsilon_{it})_{t=1}^{T} \perp\!\!\!\perp \left((\eta_j)_{j=1}^{N}, (\lambda_j)_{j=1}^{N}, (S_t)_{t=1}^{T}, (F_t)_{t=1}^{T}\right)$.*

This strengthens the interpretation of ε_{it} as an *idiosyncratic* component of the residual. For example, without this assumption, it would be possible for ε_{it} to be equal to the factor component, $\lambda_i' F_t$, as long as $\mathbb{E}\left[\lambda_i' F_t Z_{it}\right] = 0$. Assumption 3 allows us to highlight how the factor component complicates inference, relative to a more traditional model with idiosyncratic residual shocks. Note, however, that this assumption is still compatible with cross-sectional or time-series dependence in ε_{it} of other forms.

An implication of Assumption 3 is that $\mathbb{E}\left[\tilde{\lambda}_i' \tilde{F}_t \tilde{\varepsilon}_{js} \mid Z\right] = 0$. The third term in Equation 9 is zero, and we can therefore write Ω as the sum of two terms,

$$\Omega = \underbrace{\text{AVAR}\left(\frac{1}{\sqrt{N}} \cdot \frac{1}{T} \sum_{i,t} \tilde{Z}_{it} \tilde{\lambda}_i' \tilde{F}_t\right)}_{\text{Factor component}} + \underbrace{\text{AVAR}\left(\frac{1}{\sqrt{N}} \cdot \frac{1}{T} \sum_{i,t} \tilde{Z}_{it} \tilde{\varepsilon}_{it}\right)}_{\text{Idiosyncratic component}} \tag{10}$$

Separating the asymptotic variance of $\hat{\beta}$ into a factor component and an idiosyncratic component helps provide intuition about how clustering by region might fail. Clustering by region will be valid if it is valid for both the factor component and the idiosyncratic component.[5] If the idiosyncratic component, $\tilde{Z}_{it}\tilde{\varepsilon}_{it}$, is uncorrelated across regions, and if the factor component, $\tilde{Z}_{it}\tilde{\lambda}_i'\tilde{F}_t$ is also uncorrelated across regions, then clustering by region will yield consistent standard errors, under appropriate regularity conditions. If the factor component is not uncorrelated across regions, then clustering by region will typically be invalid.

We now examine how our identification assumptions will affect inference. Whether we get identification from shocks or from shares will determine which of the factor component covariance terms can be treated as zero. The following result demonstrates the critical role played by the identification assumption in this context:[6]

Lemma 1. *Let $\omega(i,j,t,s) = \mathbb{E}\left[Z_{it} \cdot \lambda_i' F_t \cdot Z_{js} \cdot \lambda_j' F_s\right]$ be the factor component covariance between units (i,t) and (j,s). The following statements are true:*

1. *If identification comes from shares (Condition 1) and (η_i, λ_i) is independent across regions, then $\omega(i,j,t,s) = 0$ for all $i \neq j$.*

[5] Mirroring our discussion of identification, there is also a knife-edge case in which non-zero covariances in each term cancel out that seems unlikely to arise in practice.

[6] In the Appendix, we prove a similar lemma for the demeaned objects.

2. *If identification comes from shocks (Condition 2) and* (S_t, F_t) *is independent across time, then* $\omega(i, j, t, s) = 0$ *for all* $t \neq s$.

The first part of the result gives a sufficient condition for the factor component not to induce correlation across regions: a combination of identification from shares and the assumption that shares are drawn independently across regions. Intuitively, the presence of a common factor does not induce cross-regional correlation *on average* if regions' characteristics are independently drawn. The lack of covariances across regions moreover suggests that, under the conditions of Part 1, clustering by region is valid.

The second part of the result gives a sufficient condition for the factor component not to induce correlation across time: a combination of identification from shares and the assumption that common factors are not autocorrelated. The lack of covariances across time moreover suggests that, under the conditions of Part 2, clustering by time is valid.

When is Clustering by Region Valid?

We now apply the logic of Lemma 1 to evaluate the standard econometric practice of clustering standard errors by region. Part 1 of that result suggested that this practice may be valid under the combination of identification from shares and independent draws of regional exposures as $N \to \infty$. We formalize this below.

Proposition 2 (Clustering by Region is Valid Under Identification from Shares). *Assume Condition 1 (Identification from Shares) and that* $\left(\eta_i, \lambda_i, (\varepsilon_{it})_{t=1}^{T}, (e_{it})_{t=1}^{T}\right)$ *is independently and identically drawn across regions. Then clustering by region consistently estimates* $AVAR\left(\sqrt{N} \cdot \hat{\beta}\right)$ *as* $N \to \infty$.

Clustering by region, however, is generally *not* valid under identification from shocks. This is because a setting with non-random assignment of shares allows different regions to predictably move together in response to unobserved aggregate shocks. Below, we formalize this point and describe the asymptotic bias in the region-clustered standard error estimator:

Proposition 3 (Clustering by Region is Biased Under Identification from Shocks). *Assume Condition 2 (Identification from Shocks), that* $\left(\eta_i, \lambda_i, (\varepsilon_{it})_{t=1}^{T}\right)$ *is independently and identically drawn across regions, that* ε *is independent of* Z*, and that* (S_t, F_t) *is independently and identically drawn across time. Define*

$$\Omega^{CR} := \mathbb{E}\left[\left(\frac{1}{T}\sum_t \tilde{Z}_{it}\tilde{u}_{it}\right)\left(\frac{1}{T}\sum_t \tilde{Z}_{it}\tilde{u}_{it}\right)'\right] \tag{11}$$

(i.e., the asymptotic limit of the region-clustered estimator, when such a limit is well-defined), and assume this expectation exists and is finite. Then, as $N \to \infty$, the asymptotic bias of the clustered estimate of Ω is given by:[7]

$$\frac{1}{N}(\Omega^{CR} - \Omega) \to -\frac{1}{T}\mathbb{E}\left[(\tilde{S}_t'\mathbb{E}\left[\tilde{\eta}_i\tilde{\lambda}_i'\right]\tilde{F}_t)^2\right] - O\left(\frac{1}{T^2}\right) \tag{12}$$

In the scalar case $J = K = 1$, this reduces to

$$\frac{1}{N}(\Omega^{CR} - \Omega) \to -\frac{1}{T}\left(\mathbb{E}[\tilde{\eta}_i\tilde{\lambda}_i]\right)^2 \ \mathbb{E}\left[\left(\tilde{S}_t\tilde{F}_t\right)^2\right] - O\left(\frac{1}{T^2}\right) \tag{13}$$

If we have identification from shocks rather than identification from shares, then clustering by region will give invalid standard errors. The bias is such that the confidence intervals will typically be too tight (that is, ignoring the $O(1/T^2)$ term arising from finite sample estimation of the fixed effects). Moreover, this bias is proportional to the number of regions, N. Clustering by region will falsely suggest that the standard errors shrink to zero as N grows large, but with small T the true standard errors will remain large. In such settings, researchers may believe that the data have spoken clearly, when in fact their results are mostly noise.

3 Proposed Econometric Solutions

Having cast doubt on conventional inference techniques, we now discuss potential solutions. We first discuss methods for confidence intervals that practitioners can feel confident in. We argue that two-way clustering and a combination of two-way clustering with an autocorrelation correction can be valid for settings in which clustering by region fails. We also propose a randomization inference method. Finally, we propose a method to construct a feasible optimal instrument à la Chamberlain (1987, 1992), which reweights data based on the factor structure to obtain a potentially more efficient estimator.

3.1 Better Standard Errors for Asymptotic Inference

Although clustering by region does not yield valid standard errors if identification comes from shocks, various existing methods yield valid standard errors in this setting. In this subsection, we discuss two options: two-way clustering and a combination of two-way clustering with an

[7] Because we are double-demeaning our variables, Ω^{CR} depends on N (as well as T). We consider the case where $N \to \infty$ because it ensures convergence of cross-sectional means.

autocorrelation correction. Regardless of whether identification comes from shares or shocks, two-way clustering yields valid standard errors if shocks are uncorrelated across time. We also discuss a method that enriches two-way clustering to allow for autocorrelation of shocks. We also comment on how these issues interaction with weak identification.

Two-way Clustering. Two-way clustering is an extension of one-way clustering that allows for both arbitrary correlation of the error term within region and arbitrary correlation of the error term within time period.[8] Although this imposes weaker restrictions on the correlation structure of the error term than one-way clustering, it still imposes that the error term is uncorrelated for observations that are in both different regions and different time periods.

Two-way clustering is implemented by combining clustering by region with clustering by time. To estimate the "meat" Ω (Equation 8), two-way clustering proposes the following estimator and considers its properties as $N \to \infty$ and $T \to \infty$:

$$\hat{\Omega}^{TWC} = \frac{1}{NT^2} \sum_{i,t} \sum_{j,s} \mathbf{1}\left(i = j \text{ OR } t = s\right) \tilde{u}_{it} \tilde{u}_{js} \tilde{Z}_{it} \tilde{Z}_{js} \tag{14}$$

Essentially, two-way clustering allows for arbitrary within-region and within-time correlation of the residual by setting $\mathbf{1}(i = j \text{ OR } t = s)$ equal to one within-region or within-time, and estimating the appropriate covariance. That indicator is still set to zero, however, for observations that are in different regions *and* different times, and so those covariances are assumed to be zero. To illustrate: clustering by region imposes that the error term in New York is uncorrelated with the error term in California, while two-way clustering imposes that the error term in New York in 2005 is uncorrelated with the error term in California in 2006.

Like clustering by region, two-way clustering is valid under identification from shares, under appropriate additional assumptions about dependence in the cross-section. If identification comes from shares and (η_i, λ_i) is drawn independently across regions, then Lemma 1 tells us that the factor component of the residual is uncorrelated across regions. If the idiosyncratic component is also uncorrelated across regions, then the whole error term is uncorrelated across regions, which allows us to either cluster by region or two-way cluster.

Unlike clustering by region, two-way clustering is also valid under identification from shocks, under appropriate assumptions about dependence across time. If identification comes from shocks and (S_t, F_t) is drawn independently across time, then Lemma 1 tells us that the factor component of the residual is uncorrelated across time. If the idiosyncratic component is uncorrelated across regions, then although the full error term has neither uncorrelatedness

[8]This method was introduced by Miglioretti and Heagerty (2007) and was further developed, independently, by Cameron et al. (2011) and Thompson (2011).

across region nor uncorrelatedness across time, it does have the property that observations that are from different regions *and* different time periods will have uncorrelated error terms.

Below, we formalize the logic that the asymptotic variance of $\sqrt{N} \cdot \hat{\beta}$ has a two-way clustering form:

Proposition 4. ***(Two-Way Clustering is Valid Under Either Identification Condition)*** *Assume that $(\varepsilon_{it})_{t=1}^{T}$ is drawn i.i.d. across regions. Assume further either of the following:*

1. *Condition 1 (Identification from Shares) holds and (η_i, λ_i) are i.i.d. across regions.*

2. *Condition 2 (Identification from Shocks) holds and (S_t, F_t) are i.i.d across time.*

Then, $\Omega = \Omega^{TWC}$, under the limit where $\frac{N}{T} \to C$, where C is a constant. That is, $AVAR\left(\sqrt{N} \cdot \hat{\beta}\right) = \lim_{N\to\infty, T\to\infty, \frac{N}{T}\to C} \frac{1}{NT^2} \sum_{i,t} \sum_{j,s} \mathbf{1}\left(i = j \;\; OR \; t = s\right) \mathbb{E}\left[\tilde{u}_{it}\tilde{u}_{js}\tilde{Z}_{it}\tilde{Z}_{js}\right]$.

An implication of this result is that if $\hat{\Omega}^{TWC}$ consistently estimates Ω^{TWC}, then it will consistently estimate the asymptotic variance. Note, however that providing conditions under which $\hat{\Omega}^{TWC}$ consistently estimates Ω^{TWC} is still an area of active research (see Davezies et al., 2021; MacKinnon et al., 2021; Menzel, 2021). We will later show Monte Carlo evidence, in our application, on the performance of two-way clustering.

Autocorrelation-Robust Clustered Standard Errors. Although two-way clustering allows for arbitrary correlation within-region or within-time, it imposes that observations that are both from different regions and different time periods (e.g., New York in 2005 and California in 2006) have uncorrelated error terms. Under identification from shocks, this requires shocks to be uncorrelated across time: if New York and California are affected by factor shocks, and those shocks are persistent over time, then California in 2006 will still be affected by the shock that affected both it and New York in 2005.

Thompson (2011) proposes an estimator that augments two-way clustering with additional terms that model cross-regional, cross-time period correlation.[9] In this "two-way HAC" method, one estimates the "meat" Ω as

$$\hat{\Omega}^{TWHAC} = \frac{1}{NT^2} \sum_{i,t} \sum_{j,s} \max\{K(t,s), \mathbf{1}(i=j)\} \hat{u}_{it}\hat{u}_{js}\tilde{Z}_{it}\tilde{Z}_{js} \tag{15}$$

where $K(t,s) = \max\left\{1 - \frac{|t-s|}{L+1}, 0\right\}$ is a kernel weight (here, the Bartlett kernel), parameterized by a bandwidth L. The use of the kernel allows for some persistence of the shock,

[9] This method builds on prior work by Driscoll and Kraay (1998) which introduced such corrections to the "one-way," regionally clustered estimate.

although the autocovariance must eventually die off. If the bandwidth, L, is selected in a way that increases with the number of time periods, then as $T \to \infty$ we also have $L \to \infty$. At the other extreme, if $L = 1$, this formula reduces to the two-way clustered standard errors considered earlier.

To our knowledge, there are no results about the asymptotic consistency of these standard errors in the literature.[10] Nonetheless, we derive confidence from our own simulation results (Section 4) that these methods can provide a good approximation to true uncertainty about estimates.

One downside of two-way clustered standard errors, with and without HAC corrections, is that they may be less efficiently estimated than those clustered just by region. If a researcher is confident that identification comes from shares and not shocks, then she may favor simple clustering by region. However, as we will see later in our application, identification from shares is unlikely to hold in practice. As a result, using a more robust formula for computing confidence intervals is crucial, and, in our application, will substantially change the results.

Weak Identification. We have focused so far on constructing valid confidence intervals under the assumption of a strong first stage, highlighting that the true uncertainty may be larger than what is suggested by clustering by region. The same concerns apply to the strength of the first stage. A first-stage relationship that appears strong according to an F-statistic that clusters by region may, in truth, be weak under a valid F-statistic. A standard solution is to construct confidence intervals based on weak-instrument robust tests. While Proposition 4 focused on the asymptotic variance of $\hat{\beta}$, our recommendations for consistently estimating Ω also can be used to compute test statistics such as the Anderson and Rubin (1949) statistic.

3.2 Randomization Inference for Finite-Sample-Valid Inference

An alternative method for constructing confidence intervals is to use randomization inference, as suggested in Borusyak and Hull (2021b). Randomization inference has two advantages over traditional asymptotic inference in our settings. First, randomization inference is valid in finite samples. This may be especially relevant in settings where the number of time periods is small and thus asymptotic approximations may be poor. Second, our randomization inference procedure will be weak-instrument robust. The main cost is that one must take a stand on the data-generating process for the instrument.

[10] We conjecture that it would be possible to prove such a result if one assumed that the shock and factor processes were α-mixing and applied a central limit theorem for α-mixing random fields, as in Driscoll and Kraay (1998).

Randomization inference inverts the logic of traditional inference. In traditional inference, the thought experiment is to redraw the residuals: we attempt to determine the variance of $\hat{\beta}$ by imagining that the residuals could have come out differently in a different draw. In contrast, randomization inference holds the residuals fixed and, instead, redraws the shocks. If we believe that identification comes from shocks, and we believe we know the underlying data generating process for the observable shocks S_t, then we can redraw S_t. To construct a hypothesis test, we compute the test statistic under the null hypothesis in the actual data, and compare this with the distribution of the test statistic under the counterfactual draws. To generate confidence intervals, we run the hypothesis test for each value of β_0 under consideration, and define the confidence interval as the set of β_0 for which the test fails to reject the null.

To implement this procedure, we need to assume a data-generating process for the shocks S_t and define a test statistic. Below, we describe the procedure that we will use in our application in Section 4. This procedure is defined without regional and time fixed effects, and we implement the procedure as randomization for the data after those fixed effects have been partialed out.

Algorithm 1. ***(Randomization Inference with One-Dimensional Shock)*** *To test a null hypothesis $\beta = \beta_0$,*

1. *Estimate a Gaussian AR(1) process for S_t:*

$$S_t = \mu + \rho S_{t-1} + \sigma \xi_t \tag{16}$$

where $\xi_t \sim N(0, 1)$.

2. *Simulate S_t using Equation 16 using estimates $(\hat{\mu}, \hat{\rho}, \hat{\sigma})$ and random shocks $\{\xi_t^{sim}\}_{t=1}^{T}$.*

3. *For each draw $\{S_t^{sim}\}_{t=1}^{T}$, construct the simulated instrument $Z_{it}^{sim} = \eta_i \cdot S_t^{sim}$.*

4. *Compare in-sample test statistic,*

$$\mathcal{T} := \frac{1}{NT} \sum_{i,t} Z_{it} \left(Y_{it} - X_{it}\beta_0\right) \tag{17}$$

to the simulated distribution of $\mathcal{T}^{sim} = \frac{1}{NT} \sum_{i,t} Z_{it}^{sim} \left(Y_{it} - X_{it}\beta_0\right)$, rejecting β_0 at the α level if $|\mathcal{T}|$ is above the $1 - \alpha$ quantile of $|\mathcal{T}^{sim}|$ (two-sided test).

Borusyak and Hull (2021b) show that randomization inference generates exact confidence intervals in a range of settings (including ours), as long as the underlying data-generating process for shocks is correctly specified. The need for correct specification raises two issues in our setting: the functional form assumptions must be correct, and the estimated parameters must be the true parameters. As T grows large, the estimated parameters of the shock process will converge to the true parameters, dealing with the second issue. The first is more difficult to solve: for realistic values of T, we will need to impose some parametric assumptions on the data generating process for shocks. In our application, we are reassured by the fact that a Gaussian AR(1) process seems to fit the data well. We also find that two richer data-generating processes, an AR(1) with Gaussian-mixture errors and a Gaussian AR(2), yield very similar randomization inference results.

Our analysis uses the test statistic (Equation 17) suggested by Borusyak and Hull (2021b). This statistic depends only on the instrument Z_{it} and the residual $Y_{it} - X_{it}\beta_0$. Conditional on those variables, it depends neither on the endogenous variable X_{it} nor the first-stage coefficient π. Thus, we do not need to specify a data-generating process for X_{it} or make an assumption about π to conduct inference. As a result, the statistical test will be weak-instrument robust.

3.3 Efficient Estimation with (Feasible) Optimal Instruments

Constructing valid confidence intervals may reveal that the standard instrumental variables estimator is too imprecise. How can we improve statistical power, while maintaining correct size?

We propose using the factor structure of the residuals to construct the optimal instrument. If we know the factor structure of the residuals, then we can improve the standard instrument through reweighting, in a process similar to generalized least squares (GLS).

The following result, adapted from Borusyak and Hull (2021a), gives an expression for the optimal instrument (Chamberlain, 1987, 1992) that minimizes the asymptotic variance of the IV estimator:

Proposition 5 (**Borusyak and Hull (2021a)**). *Suppose that the shocks, S, are independent of the error term, u, conditional on the shares, η. That is, $S \perp u \mid \eta$. Also, suppose that $\mathbb{E}[uu' \mid \eta]$ is almost-surely invertible. Consider the instrument*

$$Z^* = \mathbb{E}\left[uu' \mid \eta\right]^{-1} \left(\mathbb{E}\left[X \mid S, \eta\right] - \mathbb{E}\left[X \mid \eta\right]\right) \tag{18}$$

Then if the associated IV estimator $\beta^* = Z^{*\prime}Y/Z^{*\prime}X$ is regular, it has the smallest asymptotic variance of all regular recentered IV estimators.[11]

[11] Borusyak and Hull (2021a) define a regular IV estimator as follows: "We say that $\tilde{\beta}$ [$= \tilde{Z}'Y/\tilde{Z}'X$] is

Note that, in our setting, assuming $S \perp u \mid \eta$ implies that we are relying on identification from shocks.

The result of Borusyak and Hull (2021a) is quite general, and it simplifies substantially in our setting. First, note that since η_i is constant over time, the $\mathbb{E}[X \mid \eta]$ term will be absorbed by the region fixed effect. Thus, we can simply use $\mathbb{E}[X \mid S, \eta] - \mathbb{E}[X \mid \eta] = \pi \cdot \eta_i' S_t$, relying on the region fixed effect to residualize the instrument appropriately.

The remaining relevant parameter is $\mathbb{E}[uu' \mid \eta]^{-1}$. We simplify this exprsesion under the approximate factor structure and four additional assumptions. The first, reintroduced from Section 2.4, is that the factors and idiosyncratic shocks are uncorrelated, conditional on the instrument. That is, $\mathbb{E}[\lambda_i' F_t \varepsilon_{js} \mid Z_{it}, Z_{js}] = 0$. The second, new to this subsection, is that idiosyncratic components are i.i.d. across observations. We write $\sigma_\varepsilon^2 = \mathbb{E}[\varepsilon_{it}^2]$. The third is that the factors F_t have the same covariance matrix in each period. We write $\Sigma_F = \mathbb{E}[F_t F_t']$. Finally, we treat λ_i as fixed, so that $\mathbb{E}[\lambda_i' \Sigma_F \lambda_j \mid \eta] = \lambda_i' \Sigma_F \lambda_j$. Under these assumptions, we can write:

$$\mathbb{E}[u_{it} u_{js} \mid \eta] = \begin{cases} \lambda_i' \Sigma_F \lambda_j + \sigma_\varepsilon^2 & \text{if } s = t, j = i \\ \lambda_i' \Sigma_F \lambda_j & \text{if } s = t, j \neq i \\ 0 & \text{otherwise} \end{cases} \tag{19}$$

Intuitively, Equation 19 states that residual correlations across regions depend solely on the factor component of the residuals. If two regions have similar factor loadings, λ_i, then their residuals will be positively correlated, and they will not provide independent information. The optimal instrument reweights the data so that the residuals are uncorrelated and homoskedastic. Each observation in the reweighted data provides independent information.

Feasible Implementation. In practice, we cannot implement the optimal instrument, because we do not know the true factor structure of the errors and instead must estimate it. We will thus implement a feasible version of the optimal instrument.[12] This entails estimating Σ_F, σ_ε^2, and $(\lambda_i)_{i=1}^N$.

To do this, we first construct the model residuals under the assumption that $\beta = B$, that is, $u_{it} = Y_{it} - X_{it} B$. We then estimate the approximate factor structure of these u_{it} using principal components analysis (PCA). As shown by Stock and Watson (2002), PCA

"regular" if it converges to β at some rate r_N, if it has an asymptotic first stage (i.e. $\frac{1}{N}\tilde{Z}'X \xrightarrow{p} M$ for some $M \neq 0$), and if the sequences of $\frac{1}{N}\tilde{Z}'X$ and $\left(r_N \frac{1}{N}\tilde{Z}'u\right)^2$ are uniformly integrable." Regularity is not implied by our earlier assumption of finite "cross-term" moments between X_{it}, Y_{it}, and Z_{it}. But it would be trivially implied, for example, were all random variables bounded.

[12]The fact that the feasible optimal instrument is itself estimated will affect the distribution of any tests or estimators based on it. We will provide randomization inference based confidence intervals that account for this instrument estimation step, and thus will retain correct coverage even in finite samples.

will give consistent estimates of the factors and the loadings as long as $N, T \to \infty$. We write $u_{it} = \lambda_i' F_t + \varepsilon_{it}$, where $\lambda_i' F_t$ contains the first J components estimated by PCA. Finally, we estimate the parameters

$$\begin{aligned}\hat{\sigma}^2_\varepsilon(B,J) &= \frac{1}{NT}\sum_{i,t}\hat{\varepsilon}^2_{it} \\ \hat{\Sigma}_F(B,J) &= \frac{1}{T}\sum_t \hat{F}_t\hat{F}_t'\end{aligned} \tag{20}$$

and take $\hat{\lambda}_i(B,J)$ as the PCA estimates. We write all three statistics as a function of tuning parameters (B,J); we will discuss how to select these parameters momentarily. We use these parameter estimates to construct a feasible analogue to Equation 19, and hence a feasible optimal-instrument reweighting matrix. We summarize these steps in the following algorithm:

Algorithm 2. ***(Feasible Optimal Instrument)*** *Given a value of B and J,*

1. *Back out residuals $u_{it} = Y_{it} - X_{it}B$.*
2. *Use PCA on u_{it}. Select the first J components to define $\hat{\lambda}_i$ and $\hat{F}_t$, and define $\hat{\varepsilon}_{it} = u_{it} - \hat{\lambda}_i'\hat{F}_t$.*
3. *Estimate $(\hat{\sigma}^2_\varepsilon(B,J), \hat{\Sigma}_F(B,J))$ using Equation 20.*
4. *Estimate the matrix $\mathbb{E}[uu' \mid \eta]$ using Equation 19.*
5. *Construct the new instrument $Z^* = \mathbb{E}[uu' \mid \eta]^{-1} Z$.*

We can use the feasible optimal instrument to generate a point estimate $\hat{\beta}^{\text{opt}}(B,J)$ and associated confidence intervals. Consistent with Proposition 5, the optimal instrument could improve efficiency substantially, insofar as the feasible instrument is close to the (infeasible) true optimal instrument. We can also use the feasible optimal instrument to perform more efficient randomization inference. These confidence intervals have the added advantage of accounting for the estimation of the feasible optimal instrument, as long as B is selected appropriately. We elaborate on this next.

Efficient Randomization Inference. To perform randomization inference using the optimal instrument, we combine Algorithms 1 and 2. We use the following steps:

Algorithm 3. ***(Efficient Randomization Inference with One-Dimensional Shock)*** *Fix a value of B and J. To test a null hypothesis $\beta = \beta_0$,*

1. *Perform Steps 1-3 of Algorithm 1 to obtain simulated instrument* Z^{sim}.

2. *Use Algorithm 2 to construct optimal instrument* $Z^{*sim}(B,J)$.[13]

3. *Compare in-sample test statistic,*

$$\mathcal{T} := \frac{1}{NT}\sum_{i,t} Z_{it}^{*}\left(Y_{it} - X_{it}\beta_0\right) \tag{21}$$

to the simulated distribution of $\mathcal{T}^{sim} = \frac{1}{NT}\sum_{i,t} Z_{it}^{*sim}\left(Y_{it} - X_{it}\beta_0\right)$*, rejecting* β_0 *at the* α *level if* $|\mathcal{T}|$ *is above the* $1-\alpha$ *quantile of* $|\mathcal{T}^{sim}|$ *(two-sided test).*

The feasible implementation of the optimal instrument requires approximating $\mathbb{E}\left[uu' \mid \eta\right]^{-1}$ with Equation 19. However, randomization inference is still finite-sample-valid if B is selected correctly, even though the instrument was estimated. This is because randomization inference holds the residuals fixed, and thus the estimated weighting matrix $\mathbb{E}\left[uu' \mid \eta\right]^{-1}$ is also fixed. By redrawing the shocks, S_t, rather than the residuals, randomization inference sidesteps the issue of an estimated instrument. Note however that this argument relies on selecting B correctly. If B corresponds to the true β, then the residuals $Y_{it} - X_{it}B$ will be the true residuals, and thus will be fixed. This is an argument in favor of using $B = \beta_0$: in this case, $Y_{it} - X_{it}B$ will be the true residuals under the null, and thus the weighting matrix is also fixed under the null.

In contrast, if $B \neq \beta_0$, then we will have $Y_{it} - X_{it}B = u_{it} + X_{it}\left(\beta_0 - B\right) = u_{it} + \left(\pi Z_{it} + e_{it}\right)\left(\beta_0 - B\right) = u_{it} + \left(\beta_0 - B\right)\pi\eta_i S_t + \left(\beta_0 - B\right)e_{it}$. Under the null, these mis-estimated residuals will depend partly on the shocks. Thus, if B is selected incorrectly, our randomization inference procedure will not fully account for the estimation of the feasible optimal instrument. This distortion will be small if B is close to the true β.

Our discussion so far has assumed that the tuning parameters, B and J, have already been selected. Picking these parameters sensibly is important to unlocking the efficiency benefits of the feasible optimal instrument. We next discuss how to select these tuning parameters.

Selecting Tuning Parameters. We propose two methods to select B. When testing the null $\beta = \beta_0$, a natural choice is to pick $B = \beta_0$. Under the null, this will yield the true residuals u_{it}. However, if the researcher believes that the true B is most likely not equal to β_0, then tests using $B = \beta_0$ may be less powerful than sensible alternatives. Moreover, the $B = \beta_0$ approach is not useful for generating a point estimate, $\hat{\beta}^{\text{opt}}(B,J)$, since β_0 is only defined in the context of hypothesis testing. An alternative approach is to select

[13]Note that the weights $\mathbb{E}\left[uu' \mid \eta\right]^{-1}$ will be the same in all simulations, as they do not depend on Z.

B based on the researcher's priors. If the researcher's priors are close to the true β, then selecting such a B is likely to yield a better approximation to the true optimal instrument, maximizing power. This approach also allows the researcher to generate point estimates. We demonstrate both approaches in practice in our application to regional fiscal multipliers.

Once the researcher has selected B, we propose selecting J based on a power simulation. We construct this simulation to mirror randomization inference. First, the researcher generates simulated data under an alternative hypothesis, $\beta = \beta_a$ and $\pi = \pi_a$.[14] Then, for each simulation draw, the researcher conducts efficient randomization inference as in Algorithm 3, using a particular value of J and testing the null hypothesis $\beta = \beta_{\text{null}}$. The frequency with which efficient randomization inference rejects β_{null} gives the simulated power of the test under the alternative hypothesis. The researcher repeats this for each value of J under consideration, and then picks the value that maximizes power.

In summary, we use the following steps:

Algorithm 4. ***(Power Simulation to Select*** J***)*** *Fix values of* β_a, π_a, β_{null}, *and* B.

1. *Perform Steps 1-3 of Algorithm 1 to obtain simulated instrument* Z^{sim}.

2. *Back out the true residuals of the data under* β_a *and* π_a, *using*

$$\begin{aligned} e_{it} &= X_{it} - \pi_a Z_{it} \\ u_{it} &= Y_{it} - \beta_a X_{it} \end{aligned} \tag{22}$$

3. *Let* $Z_{it}^{sim,s}$ *denote the* s*-th simulation draw of the instrument. Simulate new draws of* $X^{sim,s}$ *and* $Y^{sim,s}$, *using:*

$$\begin{aligned} X_{it}^{sim,s} &= \pi_a Z_{it}^{sim,s} + e_{it} \\ Y_{it}^{sim,s} &= \beta_a X_{it}^{sim,s} + u_{it} \end{aligned} \tag{23}$$

4. *For a given value of* J, *use Algorithm 2 to construct optimal instrument* $Z^{*sim,s}(B, J)$ *for each simulation* s.

5. *Define the test statistic,*

$$\mathcal{T}^{r,s} := \frac{1}{NT} \sum_{i,t} Z_{it}^{*sim,r} \left(Y_{it}^{sim,s} - X_{it}^{sim,s} \beta_{null} \right) \tag{24}$$

where r *denotes the simulation draw from which* Z *is taken and* s *denotes the simulation draw from which* X *and* Y *are taken. Compute* $\mathcal{T}^{r,s}$ *for all* (r, s), *including* (s, s).

[14] It is necessary to specify a hypothesized first stage because we will simulate new values of X and Y, instead of just redrawing the instrument.

6. *For each simulation draw s, reject the null if $\mathcal{T}^{s,s}$ is above the $(1-\alpha)$ quantile of the simulated distribution of $\mathcal{T}^{r,s}$, where α is the desired size.*

7. *Compute the simulated power of the test as the share of simulation draws, s, that result in a rejection of β_{null}.*

8. *Repeat steps 4-7 for each value of J, to obtain simulated power as a function of J.*

9. *Select the value of J that maximizes the power to reject β_{null}.*

In this approach, the researcher thus must conduct a simulation within a simulation. However, because the process to draw the instrument Z_{it}^{sim} does not depend on X and Y, we only need to simulate a distribution of Z^{sim} once, substantially reducing the computational burden.

This approach allows us to select J to maximize power, for a given null $\beta = \beta_{\text{null}}$, and for a given alternative hypothesis, $\beta = \beta_a$ and $\pi = \pi_a$. The researcher must select these hypotheses. Selecting β_{null} is typically straightforward: there is a specific null hypothesis that the researcher is testing.[15] More difficult is selecting β_a and π_a. These should correspond to natural alternative hypotheses given the nature of the economic question, and/or to the researcher's priors. This is likely to be easier in some settings than others. In our regional fiscal multipliers example, a natural choice is $\beta_a = 1.5$, which corresponds to a common view about the size of the fiscal multiplier among many economists, and $\pi_a = 1$, which corresponds to the view that military procurement spending increases in each state in fixed proportion to the level of national spending. In settings where economists do not yet have well-formed priors, selecting the alternative hypothesis will be more difficult.

Selection of tuning parameters in a data-driven way can sometimes result in coverage distortions. However, we argue that we are somewhat insulated from that concern in this setting. Coverage distortions from pre-testing arise in classical inference due to conditioning on stochastic variables. In randomization inference, the residuals are held fixed, while the instrument is stochastic. Thus, if the alternative β_a and π_a are correctly specified, then our power simulation depends only on fixed variables.[16] Of course, the researcher does not know the true β_a and π_a *ex ante*. Thus, in practice, it is useful to examine how different values of the alternative affect the choice of J. We discuss this further in our application.

[15] In principle, a researcher constructing a confidence interval for β could simply set $\beta_{\text{null}} = \beta_0$, and select an optimal J for each point they are testing to generate their confidence interval. In practice, this is computationally burdensome, and using the same J for the whole confidence interval is much more practical.

[16] The power simulation will give correct power if π_a and β_a are correct, and will only depend on fixed variables. However, as discussed earlier, achieving tests with correct size will rely on correctly selecting B.

3.4 Further Issues

Before proceeding, we comment on two further issues: choosing the level at which to do the analysis and inference in shift-share designs. Although they are not the primary focus of our analysis, our framework is useful to better understand these topics.

Choosing the Unit of Analysis. Researchers interested in the effect of an aggregate shock, S_t, often must choose the level of aggregation at which to do the analysis. One key question is whether to do a time-series analysis (e.g., with national level data) or a panel analysis (e.g., with regional data). One issue is that the national and regional analyses may estimate different quantities: for example, the aggregate fiscal multiplier will typically be different from the regional fiscal multiplier, due to trade across regions and the response of monetary policy to fiscal shocks.

The national and regional approaches also rely on different identifying assumptions, which our framework helps to clarify. Suppose, for illustration, that the national variables are an unweighted average of the regional variables.[17] Aggregating the structural equation (2) gives the new regression

$$Y_t = \beta \cdot X_t + \underbrace{\alpha_t + \frac{1}{N} \sum_i \left(\varepsilon_{it} + \lambda_i' F_t\right)}_{u_t} \tag{25}$$

Consider estimating this model using the instrumental variable $Z_t = S_t$. The identifying assumption for the national regression is that Z_t is orthogonal to u_t. Since λ_i has mean zero and ε_{it} also has mean zero in each period, the identification condition reduces to $\mathbb{E}\left[Z_t u_t\right] = \mathbb{E}\left[S_t \alpha_t\right] = 0$. Thus, the national regression requires that the shock, S_t, is orthogonal from other aggregate *shocks that affect all regions equally*, α_t. This is in contrast to identification from shocks in the regional setting, which requires S_t to be orthogonal to aggregate *shocks that differentially affect regions*, F_t. Although conceptually related, these identifying assumptions are distinct and non-nested.

One potential advantage of the regional regression versus the national regression is that, in many cases, it may increase the amount of variation that the researcher can exploit, improving statistical power. This is application specific, and will depend both on the variation in the regressor and the cross-region correlation structure of the residual. We provide valid confidence intervals for the regional case, so that researchers can correctly assess how much more precision the regional regression has provided them.

A researcher who chooses to do a regional analysis must also choose the level at which

[17] A similar argument holds if the national variables are a weighted average with weights that are uncorrelated with the regional exposures λ_i.

to do the analysis: for example, whether to do an analysis at the state or county level. Using more granular data has the potential to offer additional precision, by exploiting finer variation in the treatment. However, granular data raises two issues. First, using more granular data may change the quantity being estimated, due to spillovers across adjacent regions. This is distinct from the issue of correctly estimating standard errors: researchers who use more granular data must either handle spillovers econometrically and/or have some model for how to map regional estimates to aggregate effects.

Second, to the extent that shocks are correlated across (granular) regions, increasing the number of regions may further distort inference. In fact, Proposition 3 shows that the size of the bias in the clustered-by-region standard error is proportional to the ratio of N to T. Here, our proposed methods for inference can help. By correctly handling the cross-region correlation structure of the residual, our methods can allow researchers to exploit more granular data without worrying that it will distort their confidence intervals.

Shift-Share Designs. Shift-share instruments are a special case of our setting. Shift-share, or "Bartik (1991)" instruments interact initial industry shares with national shocks to those industries. In our notation, the vector of initial industry shares is the vector of exposures η_i, and the vector of industry shocks is S_t.

The recent literature on shift-share instruments can also be understood in the context of our framework. Goldsmith-Pinkham et al. (2020) achieve identification through as-good-as-random assignment of shares, relying on identification from shares as in our Condition 1. Adao et al. (2019) and Borusyak et al. (2022) argue that identification from shocks (our Condition 2) is more plausible, and show that traditional clustering by region is invalid in that setting.

We differ from the shift-share literature in at least two key respects. First, our framework is more general: shift-share is a special case of the regional-exposure research design that we study, but there are many regional-exposure designs that do not fall under the shift-share category (e.g. Nakamura and Steinsson, 2014). Second, although we agree with Adao et al. (2019) and Borusyak et al. (2022) that identification from shocks is the most plausible path, we solve the inference problem differently. Whereas we rely on $T \to \infty$ asymptotics, Adao et al. (2019) and Borusyak et al. (2022) assume that shocks are independent across sectors, and rely on asymptotics in which the number of sectors grows large. Intuitively, their standard errors cluster by sector, rather than clustering by region.

We prefer our approach based on concerns that shocks may not be independent across sectors. There is an active literature in macroeconomics studying how macroeconomic fluctuations can arise from firm-level or sector-level shocks when the number of firms/sectors grows large. The leading solutions to this question include spillovers (Acemoglu et al., 2012;

Baqaee and Farhi, 2019), a heavy-tailed distribution of firm sizes (Gabaix, 2011) and a factor structure to sectoral shocks (Foerster et al., 2011). Each of these explanations would violate the assumptions necessary for many-industries asymptotics.

Unfortunately, many shift-share studies have relatively few time periods with which to do inference. The methods we provide to construct valid confidence intervals rely on large T, and thus cannot be used in these settings. We thus cannot compare our own estimates of the standard error in these settings to the standard errors provided by Adao et al. (2019) and Borusyak et al. (2022), and so we cannot easily determine how important these concerns are in practice. We believe this is an important area for future research.

4 Application: Regional Fiscal Multipliers

To study how important our theoretical concerns are in practice, we apply them to the estimation of regional fiscal multipliers in Nakamura and Steinsson (2014). These authors use variation over time in national defense procurement spending, interacted with differential exposure across US states, to construct an instrument that they use to estimate the regional fiscal multiplier. As we explain below, the empirical strategy of Nakamura and Steinsson (2014) fits within our framework as a regional-exposure design with identification from shocks.

We first illustrate that the concerns about conventional inference that we raise in Section 2 apply in this setting. There is a strong factor structure to the residual, with the first two principal components explaining over 60% of the variance. This suggests that residuals are not independent across states, and thus clustering standard errors by state will yield incorrect confidence intervals. We demonstrate this incorrect coverage using a placebo test, in which we randomly generate fake military spending shocks. Although $\beta = 0$ in this setting by construction, we incorrectly reject this null hypothesis more than 25% of the time when using standard errors clustered by state.

We then show that valid confidence intervals estimate the regional fiscal multiplier with considerable uncertainty. A 95% confidence interval constructed with randomization inference contains values as low as 0.1, while a 68% confidence interval contains values as low as 1.1. The feasible optimal instrument substantially improves statistical power, but we still cannot reject low multipliers at the 95% level.

4.1 Setting

Nakamura and Steinsson (2014) estimate the following equation:

$$\text{Output Growth}_{it} = \alpha_t + \gamma_i + \beta \cdot \text{Military Procurement Growth}_{it} + u_{it} \tag{26}$$

where Output Growth$_{it}$ is defined as $\frac{Y_{it}-Y_{i,t-2}}{Y_{i,t-2}}$, with Y_{it} being per capita output in state i and year t, and Military Procurement Growth$_{it}$ is defined as $\frac{G_{it}-G_{i,t-2}}{Y_{i,t-2}}$, with G_{it} being per capita military procurement spending in that state and year. In their main specification, the authors use data on fifty states, plus the District of Columbia, and use annual data from 1968-2006.

The authors use two instrumental variables strategies. In one strategy, they construct an instrument that interacts the growth rate of total national military procurement spending with the state's average level of spending, relative to state output, in the first five years of the sample. In our language, S_t is the national growth rate of military procurement spending, and η_i is the average of $\frac{G_{it}}{Y_{i,t}}$ in the first five years of the sample. We refer to this as Strategy 1 or the "Intial Share" strategy. In another strategy, Nakamura and Steinsson (2014) construct 51 instruments, one per state (plus the District of Columbia), as the interaction of state fixed effects with the growth rate of total national military procurement spending. In essence, the first stage regressions estimate a sensitivity $\hat{\eta}_i$ for each state using the data. We refer to this as Strategy 2 or the "State FE" strategy. This is actually the preferred specification in the original study.

While we report results below for both strategies, we focus more on the Initial Share strategy for two reasons. First, this strategy is nested exactly in our framework. Second, as we show below, the State FE strategy suffers from a many weak instruments problem that makes the IV estimator uninformative for a reason outside of this paper's main interest.

Identification from Shocks or Shares. We begin by asking whether they plausibly achieve identification from shocks or from shares. As we showed in Section 2, this is important both for assessing whether the identifying assumptions are plausible and for understanding whether clustering by region is likely to yield valid confidence intervals.

Although they do not use our paper's language, the authors themselves argue that identification in their setting comes from shocks. They write:

> Our identifying assumption is that the United States does not embark on military buildups—such as those associated with the Vietnam War and the Soviet invasion of Afghanistan—because states that receive a disproportionate amount of military spending are doing poorly relative to other states.

In our framework, S_t represents military buildups and F_t shifts the relative economic performance of high- and low-procurement states. Thus, the authors are arguing that $S_t \perp F_t$.

Moreover, Nakamura and Steinsson (2014) argue that identification from shares is implausible. They write:

> Military spending is notoriously political and thus likely to be endogenous to regional economic conditions (see, e.g., Mintz 1992).

If military spending is endogenous to regional economic conditions, it seems likely that any regional exposure variable, η_i, will be correlated with other factor loadings, λ_i. This is especially true given that the exposure variable is itself constructed as the state's average level of military spending, as a share of output, in the first five years of the sample.

To fully rule out identification from shares, we show that the initial share of military spending in state output, η_i in Strategy 1, is correlated with a variety of other important variables at the state level. The variables that we consider are the six control variables from Autor et al. (2013): the share of employment that is in manufacturing, the share of the population that has a college education, the share that is foreign born, the share of working-age women that are employed, the share of employment that is in routine occupations, and an offshorability index for the occupations in that state. We use the 1990 values of these variables.[18]

We show the correlation between the procurement share, η_i, and these variables in Table 1. Four of the six correlations are statistically significant: η_i is higher in places with more routine occupations, more offshorable occupations, a larger college-educated population share, and a larger foreign population share. This suggests that identification is unlikely to come from shares. States with different values of η_i are observably different in other ways; to the extent that these observables may themselves interact with aggregate shocks, we would thus think that η_i is not orthogonal to λ_i.

Factor Structure of the Residual. We next study whether the model's residual has a factor structure. To do this, we estimate the model from Nakamura and Steinsson and back out the estimated residuals, $\hat{u}_{it}$. We then use principal component analysis (PCA) on the estimated residuals to estimate the factor structure (i.e., the shocks $\hat{\lambda}_i$ and the loadings $\hat{F}_t$). We perform this procedure separately for each of the two instrumental variable strategies. We then calculate the cumulative share of variance explained by the first 10 factors. We plot the results in Figure 1, which also indicates an optimally selected number of factors according to the information criterion in Bai and Ng (2002).

[18] Autor et al. construct their data set at the commuting zone level, and exclude Alaska, Hawaii, and the District of Columbia. We aggregate their variables to the state level by taking the population-weighted average.

Table 1: Correlation with Initial Share of Military Spending in State Output

Variable	Correlation	Variable	Correlation
% Employment in Manufacturing	0.079 ($p = 0.451$)	% Employment Among Women	0.223 ($p = 0.129$)
% College Educated	0.302 ($p = 0.023$)	% Routine Occupations	0.467 ($p = 0.005$)
% Foreign Born	0.445 ($p < 0.001$)	Offshorability	0.507 ($p = 0.001$)
Observations	48		48

Notes: This table shows the correlation between the initial share of military spending in state output and six covariates from Autor et al. (2013). Each p-value is computed from a univariate regression of the initial military spending share on the covariate, using heteroskedasticity robust standard errors. Autor et al. (2013) provide their covariates at the commuting zone level and exclude Alaska, Hawaii, and the District of Columbia: we aggregate their covariates to the state level by taking the population-weighted average. The variables are the share of employment that is in manufacturing, the share of the population that has a college education, the share that is foreign born, the share of working-age women that are employed, the share of employment that is in routine occupations, and an offshorability index for the occupations in that state. The initial share variable is computed as the share of military procurement spending in state output, averaged over the first five years of the sample.

We find that a large component of both model residuals is explained by very few factors. In either specification, the first two principal components explain more than 60% of the variance of the residual (66% for the first specification, and 62% for the second specification), and the factor component explains 80% of the variance using the optimally selected number of factors. Not only is there a factor structure to the residual, but the factor component explains most of its variance. Given our earlier finding that identification is unlikely to come from shares, this strongly suggests that clustering by state will typically yield invalid confidence intervals, an issue we turn to next.

4.2 Placebo Test: Standard Methods Reject Too Often

Since the residual has a factor structure, and identification likely comes from shocks rather than shares, our results suggest that clustering by state is unlikely to yield valid confidence intervals. To explore how this and other methods perform in practice, we conduct a placebo test using fake military procurement shocks.

Methods. Our procedure follows the logic of randomization inference, in which the residuals are held fixed but the instrument is redrawn. First, we back out the first- and second-stage residuals (e_{it}, u_{it}) under a maintained null hypothesis $\beta = 0$ and $\pi = \hat{\pi}$, where $\hat{\pi}$

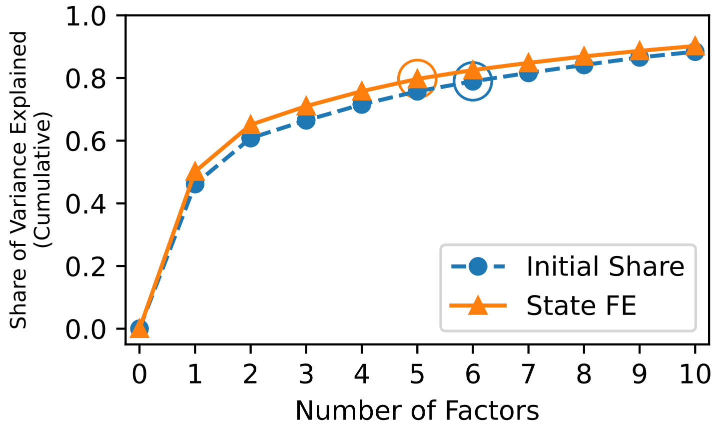


Figure 1: Share of Variance in Residual Explained by Factors

Notes: This figure shows the cumulative share of the residual's variance explained by each principal component (factor) in the residual. Residuals are based on the regression model in Nakamura and Steinsson (2014), estimated using two-stage least squares. Factors are estimated using PCA, and ordered by the share of the variance of the residual that they explain. The blue circles show results based on the first instrumental variable strategy, which interacts defense spending growth with the share of military procurement spending in state output, averaged over the first five years of the sample. The orange triangles show results based on the second instrumental variable strategy, which interacts (placebo) defense spending growth with state fixed effects to generate the instruments. The points that are circled correspond to the optimally selected number of factors, based on the information criterion in Bai and Ng (2002).

is our first-stage point estimate.[19] Next, for each of many simulation draws, we simulate placebo sequences of national military procurement growth, S_t^{sim}, using the first three steps of Algorithm 1. In particular, we model national military procurement spending growth as a Gaussian, AR(1) process, which we estimate in the data. In the data, our estimate for the shock persistence is $\rho = 0.66$. We then construct a placebo sequence of the endogenous variable, local procurement spending, as

$$X_{it}^{\text{sim}} = \omega_t + \zeta_i + \pi \eta_i S_t^{\text{sim}} + e_{it} \tag{27}$$

where (ω_t, ζ_i) are estimated fixed effects, π is the estimated first-stage coefficient, and η_i is the exposure variable.[20] We similarly construct $Y_{it}^{\text{sim}} = \alpha_t + \gamma_i + \beta X_{it}^{\text{sim}} + u_{it}$. Under our null hypothesis that $\beta = 0$, this reduces to $Y_{it}^{\text{sim}} = Y_{it}$.

[19] In Algorithm 2, we did not need to generate simulated values of X_{it} because our test statistic did not depend on X_{it} conditional on u_{it} and Z_{it}. In this exercise, by contrast, the test statistic does depend directly on X_{it}. We choose $\pi = \hat{\pi}$ for illustration so that the data-generating process of the placebo (and, in particular, the first stage correlation of X_{it}^{sim} and Z_{it}^{sim}) closely matches the observed data. Had we used a value of π closer to zero, we would introduce a (more severe) weak-instrument problem.

[20] For the "State FE $\times S_t$" strategy, these exposures are first-stage regression coefficients, and $\pi = 1$. For the "Initial Share $\times S_t$" strategy, these exposures are the observed pre-period spending shares.

As robustness checks, we also do four additional simulations with other data-generating processes for X_t. The first two explore the importance of autocorrelation by setting $\rho = 0$ and $\rho = 0.9$ in the Gaussian AR(1) model, holding fixed the unconditional variance of X_{it}^{sim}. The third explores the role of leptokurtic shocks—concretely, that the military procurement time series is characterized by a few large shocks of fast growth and draw-downs. To do this, we model the AR(1) residual as a Gaussian mixture with two components, which can differ in their mean and variance.[21] The fourth explores the role of richer dynamics by estimating an Gaussian AR(2) process for X_t.[22]

Finally, for each simulation draw, we estimate the model using two-stage least squares and perform hypothesis tests at the 5% level for $\beta_0 = 0$. We study 16 tests, which interact four "base tests" with four different clustering strategies. The clustering strategies are: clustering by state, clustering by year, two-way clustering, and two-way HAC clustering ($L = 3$).

The first two tests are based on conventional t-statistics of the form $t(\text{se}, \beta_0) = \frac{\hat{\beta}-\beta_0}{\text{se}}$, assumed to have an asymptotic $N(0, 1)$ distribution. We consider two standard error estimates.[23] The first, which we call $\text{se}_{\hat{\beta}}$, plugs in residuals evaluated at the point estimate, $\hat{\tilde{u}}_{it} = \tilde{Y}_{it} - \hat{\beta}\tilde{X}_{it}$. The second, which we call se_{β_0}, plugs in residuals evaluated at the null hypothesis of interest, $\hat{\tilde{u}}_{it}^0 = \tilde{Y}_{it} - \beta_0\tilde{X}_{it}$. As observed by Cameron and Miller (2015) and Adao et al. (2019), the former estimator for cluster-robust standard errors may be biased in small samples, and alternative estimators that impose the null hypothesis often perform significantly better.

The second two tests are versions of the weak-instrument robust test of Anderson and Rubin (1949) that are adapted to allow for clustering, as introduced by Finlay and Magnusson (2009) and Magnusson (2010). The first is the "Minimum Distance" test described in those references, which calculates a covariance matrix that conditions on the estimate $\hat{\beta}$, in analogy to $\text{se}_{\hat{\beta}}$. We refer to this test as "AR-MD." The second is a "Lagrange Multiplier" variant, which calculates a covariance matrix that uses β_0 in place of $\hat{\beta}$, in analogy to se_{β_0}, and potentially with similar advantageous properties for bias reduction when estimating the clustered covariances. We refer to this test as "AR-LM." In both cases, the test statistics

[21] Specifically, we assume that the pdf of the innovation ξ_t is $f(\xi) = \alpha\phi(\xi; \mu_1, \sigma_1) + (1-\alpha)\phi(\xi; \mu_2, \sigma_2)$, where $\phi(\cdot; \mu, \sigma)$ is a Gaussian pdf with mean μ and standard deviation σ and the free parameters are $\alpha \in (0, 1)$, $(\mu_1, \mu_2) \in \mathbb{R}^2$, and $(\sigma_1, \sigma_2) \in \mathbb{R}_+$. We estimate all parameters via maximum likelihood estimation. We estimate $\hat{\alpha} = 0.42$ and $\hat{\sigma}_1/\hat{\sigma}_2 = 3.47$, so there is a 42% chance of a "large shock" with 3.47 times the volatility of the "regular shock." This distribution has a kurtosis of 5.77 or an excess kurtosis of 2.77.

[22] That is, we model $X_t = \rho_1 X_{t-1} + \rho_2 X_{t-2} + \xi_t$ where $\xi_t \sim N(0, \sigma^2)$. We estimate $\hat{\rho}_1 = 0.96$ and $\hat{\rho}_2 = -0.36$. The AR(2) model is selected by both the AIC and BIC as the best-fitting AR(p) model for $p \in \{1, \ldots 5\}$.

[23] In both methods, we apply Stata's default small-sample correction to multiply the standard error estimate by $c = \sqrt{G(N-1)/((G-1)(N-K))}$,where G is the minimum number of clusters across the two dimensions, N is the sample size, and K is the number of regressors (including estimated fixed effects).

Table 2: False Rejection Rates for Placebo Test Based on Nakamura and Steinsson (2014)

Panel A: Initial Share Strategy

	Conventional		Weak-IV Robust	
	$\text{se}_{\hat{\beta}}$	se_{β_0}	AR-MD	AR-LM
Cluster by State	25.4%	19.8%	27.8%	19.8%
Cluster by Year	24.4%	20.8%	28.4%	20.8%
Two-way Cluster	21.1%	9.0%	20.9%	9.0%
Two-way HAC ($L = 3$)	20.3%	3.0%	20.7%	3.0%
Randomization Inference	5% (By Construction)			

Panel B: State FE Strategy

	Conventional		Weak-IV Robust	
	$\text{se}_{\hat{\beta}}$	se_{β_0}	AR-MD	AR-LM
Cluster by State	27.0%	17.6%	100.0%	0.0%
Cluster by Year	28.8%	21.8%	94.2%	0.0%
Two-way Cluster	20.2%	6.8%	15.0%	2.6%
Two-way HAC ($L = 3$)	20.5%	1.4%	8.2%	2.4%
Randomization Inference	5% (By Construction)			

Notes: This table shows the frequency at which the null hypothesis of $\beta_0 = 0$ is rejected at the 5% level in our placebo test based on Nakamura and Steinsson (2014). A correctly calibrated 5% test would reject 5% of the time. Panel A shows results based on the IV that interacts defense spending growth with the pre-period share of military procurement spending in state output. Panel B shows results based on the IV strategy that interacts defense spending growth with state fixed effects. The first four rows of each panel show results from tests that implement clustering by state, clustering by year, using two-way clustering (state and year), and using two-way HAC standard errors with a kernel bandwidth of three years. In each of these rows, we report results from conventional t-tests with standard error estimates $\text{se}_{\hat{\beta}}$ and se_{β_0} and weak-instrument-robust tests using the Anderson-Rubin Minimum Distance and the Anderson-Rubin Lagrange Multiplier statistics of Finlay and Magnusson (2009) and Magnusson (2010) (see Section 4.2 for details). In Panel A, the se_{β_0} and AR-LM tests exactly coincide. The fifth row reports that randomization inference rejects the null 5% of the time by construction, since the placebo test uses the same simulated shocks as randomization inference.

have an asymptotic $\chi^2(n_I)$ distribution where n_I is the number of instruments. In Appendix A.6, we give formulae for these test statistics. We also observe that, in the just-identified case (e.g., Strategy 1), the AR-LM test exactly coincides with the conventional se_{β_0} test.

The true β for the placebo regression is zero, by construction. Thus we expect all tests at the 5% level to falsely reject this null 5% of the time. Note that randomization inference rejects the null 5% of the time by construction, since the placebo test uses the same simulated shocks as randomization inference.

Results. We show our results in Table 2. We split the results into two panels corresponding to the Initial Share Strategy and the State FE Strategy.

For both strategies, clustering on one dimension performs pooorly. The preferred empirical method of Nakamura and Steinsson (2014), clustering by state in the State FE strategy and using $\text{se}_{\hat{\beta}}$ standard errors, spuriously rejects the null hypothesis 27.0% of the time. The same clustering strategy when applied to the Initial Share strategy spuriously rejects 25.4% of the time. Clustering by year does not perform much better, with rejection rates of 28.8% and 24.4%, respectively. In each case, calculating standard errors imposing $\beta_0 = 0$ improves this false rejection modestly. These findings are consistent with being in the Identification from Shocks case and having a significant factor structure in the residual.

Multi-way clustering performs better, particularly when combined with the standard error calculation that impose the null hypothesis. In particular, two-way clustering with the se_{β_0} calculation rejects 9.0% of the time for the Initial Share strategy and 6.8% of the time for the State FE strategy. Two-way HAC standard errors with the se_{β_0} calculation slightly *over*-reject in each case, at 3.0% and 1.4% rates, respectively. The rejection probabilities in all four cases with the $\text{se}_{\hat{\beta}}$ calculation are all smaller than the probabilities with one-way clustering, but larger than 20%. This drastic improvement in coverage is consistent with our reasoning in Section 3.1. Our finding that imposing the null hypothesis drastically improves performance is consistent with the findings of Adao et al. (2019) in a different context with cluster-robust inference.

We now consider weak-instrument robust inference. For the Initial Share strategy (Panel A), this gives consistent results to conventional, Wald-test inference. Multi-way and multi-way HAC clustering perform much better than single-way clustering. And the AR-LM test, which imposes the null hypothesis when constructing the test statistic, has considerably lower rejection probabilities than the minimum distance test, which does not. As observed earlier the AR-LM test in this single-variable context coincides exactly with the conventional se_{β_0} test.

For the State FE strategy (Panel B), Weak-IV robust methods behave erratically. The AR-LM test substantially under-rejects the null hypothesis—strikingly, the one-way clustering tests *never* reject the null hypothesis at the 5% level. The AR-MD test substantially over-rejcts the null hypothesis—strikingly, the one-way clustering tests almost always reject the null hypothesis at the 5% level. This is a first indication that the State FE strategy suffers from a severe many weak instruments issue.

In the Appendix, we explore the robustness of these results to the alternative data-generating processes described above. In light of the results above, we focus on the Initial Share strategy and on conventional inference. We present all results in Table 5. In the simulation with $\rho = 0$, clustering by region continues to over-reject (21.6% for with $\text{se}_{\hat{\beta}}$ and 16.0% with se_{β_0}). Clustering by year gives almost correct coverage (5.8% for with $\text{se}_{\hat{\beta}}$ and 3.6% with se_{β_0}). Because our $\rho = 0$ simulation makes S_t both serially uncorrelated and independent from the residual, the terms ignored by time-clustering (i.e., $\mathbb{E}[Z_{it}u_{it}Z_{js}u_{js}]$ for $t \neq s$) will in fact be zero in this environment regardless of the correlation structure for u_{it}.[24,25] The persistent, $\rho = 0.9$ simulation has slightly worse coverage compared to our baseline in all cases, consistent with the same intuition about the role of autocorrelated shocks. The leptokurtic, normal-mixture simulation and AR(2) simulation each have almost identical coverage to our baseline. Thus, while these simulations represent qualitatively different dynamics, they have similar implications for the performance of our statistical estimators. This is natural if, asymptotically, only second-moment properties matter for the estimators and if the AR(1) model captures these properties relatively well.

We draw two main conclusions from this exercise. First, cross-regional correlation in the residuals distorts inference in both conventional and weak-instrument-robust inference, especially when the instrument is persistent. Second, addressing this issue with two-way clustering or two-way HAC standard errors leads to valid inference when combined with estimators that impose the null hypothesis (i.e., the se_{β_0} standard errors or the Lagrange Multiplier test statistic).

As an alternative method, we suggest using randomization inference. By construction, randomization inference will reject the null hypothesis 5% of the time in this simulation. In the next subsection, we will show how the confidence interval in the regional fiscal multipliers example changes under different clustering methods and under randomization inference.

4.3 Valid Confidence Intervals Include Low Multipliers

We now re-estimate the confidence intervals from Nakamura and Steinsson (2014) using all of our methods. We show the results for 95% confidence intervals in Table 3, and in the Appendix we report the equivalent for 90% confidence intervals (Table 6) and 68% confidence intervals (Table 7). Panel A shows results for the first IV strategy (the initial share of military spending in state output during the first five years of the sample, interacted with the growth of national military procurment spending) and Panel B shows the results for the second IV strategy (state fixed effects interacted with the growth of national military procurement spending). We show results clustering by state (as in the original study), using two-way clustering, using two-way HAC standard errors (with a kernel bandwidth of three years), and using randomization inference. For each of the clustering options, we show conventional

[24] This argument uses the fact that $\mathbb{E}[Z_{it}u_{it}Z_{js}u_{js}] = \mathbb{E}[\mathbb{E}[Z_{it}Z_{js}|u_{it}, u_{js}]u_{it}u_{js}] = \mathbb{E}[0 \cdot u_{it}u_{js}] = 0$. This relies not just on uncorrelatedness of S_t, but uncorrelatedness conditional on the residual (itself implied by the stronger assumption of independence in the simulation).

[25] In a different regional-exposure study of stock-market wealth effects, Chodorow-Reich et al. (2021) argue that combining two-way clustering with a shock variable that is nearly uncorrelated over time (national stock returns) allays concerns about distorted inference cross-region and cross-time-period residual correlation.

Table 3: 95% Confidence Intervals for Conventional IV Estimate of Regional Fiscal Multiplier in Nakamura and Steinsson (2014)

Panel A: Initial Share Strategy
Point Estimate: 2.477

	Conventional		Weak-IV Robust	
	$\text{se}_{\hat{\beta}}$	se_{β_0}	AR-MD	AR-LM
Cluster by State	$(0.583, 4.371)$	$(0.906, \infty)$	$(0.784, 4.959)$	$(0.906, \infty)$
Two-way Cluster	$(0.370, 4.583)$	$(0.712, \infty)$	$(0.746, 5.440)$	$(0.712, \infty)$
Two-way HAC ($L = 3$)	$(0.045, 4.909)$	$(-\infty, \infty)$	$(0.498, 5.879)$	$(-\infty, \infty)$
Randomization Inference	$(0.08, 5.34)$			

Panel B: State FE Strategy
Point Estimate: 1.426

	Conventional		Weak-IV Robust	
	$\text{se}_{\hat{\beta}}$	se_{β_0}	AR-MD	AR-LM
Cluster by State	$(0.704, 2.149)$	$(0.76, 2.75)$	Empty	$(-\infty, \infty)$
Two-way Cluster	$(0.324, 2.528)$	$(0.40, \infty)$	$(-\infty, \infty)^*$	$(-\infty, \infty)^*$
Two-way HAC ($L = 3$)	$(0.032, 2.821)$	$(-\infty, \infty)$	$(-\infty, \infty)^*$	$(-\infty, \infty)^*$
Randomization Inference	$(-4.4, 8.5)$			

Notes: This table shows 95% confidence intervals for the regional fiscal multiplier, estimated in the setting of Nakamura and Steinsson (2014) using the IV estimator. Panel A shows results based on the instrumental variable strategy that interacts defense spending growth with the pre-period share of military procurement spending in state output. Panel B shows results based on the instrumental variable strategy that interacts defense spending growth with state fixed effects. The first four rows of each panel show results from tests that implement clustering by state, clustering by year, using two-way clustering (state and year), and using two-way HAC standard errors with a kernel bandwidth of three years. In each of these rows, we report results from conventional t-tests with standard error estimates $\text{se}_{\hat{\beta}}$ and se_{β_0} and weak-instrument-robust tests using the Anderson-Rubin Minimum Distance and the Anderson-Rubin Lagrange Multiplier statistics of Finlay and Magnusson (2009) and Magnusson (2010) (see Section 4.2 for details). In Panel A, the se_{β_0} and AR-LM tests exactly coincide. The fifth row of each panel reports results from randomization inference. *: All of the weak-instrument-robust confidence intervals in Column 1, marked with this symbol, contain "holes" (i.e., intervals of width close to 0.01 in which one can reject the null hypothesis).

t-statistic-based confidence intervals corresponding to each standard error estimator ($\text{se}_{\hat{\beta}}$ and se_{β_0}) and weak-instrument robust Anderson-Rubin confidence intervals corresponding to both the minimum distance and Lagrange Multiplier statistics.[26] Since Nakamura and Steinsson focus on the idea that a high regional fiscal multiplier provides evidence against a "plain-vanilla Neoclassical model," we center our discussion on the lower bound of each confidence interval.

Clustering strategies that account for cross-regional correlation of the residual yield

[26] Our randomization inference interval uses a weak-instrument robust test statistic, so we only have one type of confidence interval to show in that row.

substantially wider traditional confidence intervals. For the Initial Share strategy, the conventional confidence interval widens from $(0.583, 4.371)$ under clustering by state to $(0.370, 4.583)$ under two-way clustering and $(0.045, 4.909)$ under two-way HAC standard errors. The confidence intervals calculated with se_{β_0} are substantially wider in the two-way clustering and two-way HAC cases, which are precisely those in which the placebo test suggested that this debiased estimate provided better coverage. The strategy with closest to correct coverage in the placebo test, the Two-Way HAC confidence interval with se_{β_0}, cannot rule out any value of the fiscal multiplier at the 95% level. The 90% confidence interval from the same method rejects multipliers lower than 0.473 and the 68% confidence interval rejects multipliers lower than 1.462.

Randomization inference yields a confidence interval of $(0.08, 5.34)$. This is consistent with our findings from conventional inference with more conservative clustering. The same strategy yields a 90% confidence interval of $(0.46, 4.72)$, and a 68% confidence interval of $(1.34, 3.70)$. To probe the sensitivity of these results to alternative specifications of the data-generating process for military procurement growth, we re-calculate the randomization-inference confidence intervals with the alternative data-generating processes that we fit to the data. For the Gaussian-mixture (leptokurtic) simulation, we find 95% confidence intervals of $(-4.7, 9.1)$ for Strategy 1 and $(0.26, 5.10)$ for Strategy 2. For the AR(2) simulation, we find, respectively, $(-4.3, 8.6)$ and $(0.28, 5.24)$. All of these intervals are very similar to the findings in Table 3. These results suggest that our randomization inference results, like our placebo test results, are not unduly sensitive to the parameterization of the data-generating process.

For the the State FE strategy, we also see conventional confidence intervals widen for the clustering strategies that are more conservative and, according to the simulation, give more accurate inference. We also find that weak-instrument robust confidence intervals, regardless of clustering, are unbounded.[27] Like our simulation results, this is indicative of a severe problem with weak instruments. As such, we do not interpret the conventional inference results as conclusive.

The dramatic expansion of confidence intervals in Table 3 illustrates the practical importance of correctly accounting for the correlation structure of the residual in a setting with regional data. Since shares are non-randomly assigned, and since we find strong evidence of a factor structure to the residual, clustering by state will not yield valid confidence intervals. When we adjust the confidence intervals to allow for the factor structure of the residual, we

[27] We also replicate these results using the conditional likelihood ratio test of Moreira (2003), which Andrews et al. (2006) shows is nearly uniformly most powerful.. We again find that the confidence intervals range from $-\infty$ to ∞, for all forms of clustering.

can no longer rule out very low fiscal multipliers at the 95% level.

This is important for the interpretation of Nakamura and Steinsson (2014), who argue that their estimates of reasonably high fiscal multipliers provide evidence against a "plain-vanilla" Neoclassical model of the US economy. The authors' preferred estimate, clustering by state in the State FE strategy, corresponds to a lower bound in the 95% confidence interval of 0.704; the same number in the Initial Share strategy is 0.583. By contrast, at the 95% level, randomization inference cannot rule out multipliers as low as -4.4 in the State FE strategy or 0.08 in the Initial Share strategy.

4.4 Efficient Estimation Improves Power

We now implement efficient estimation with a feasible optimal instrument, as introduced in Section 3.3. Based on the results of the previous subsection, which highlighted the issues with "State FE $\times S_t$" empirical strategy, we focus on the "Initial Share $\times S_t$" empirical strategy.

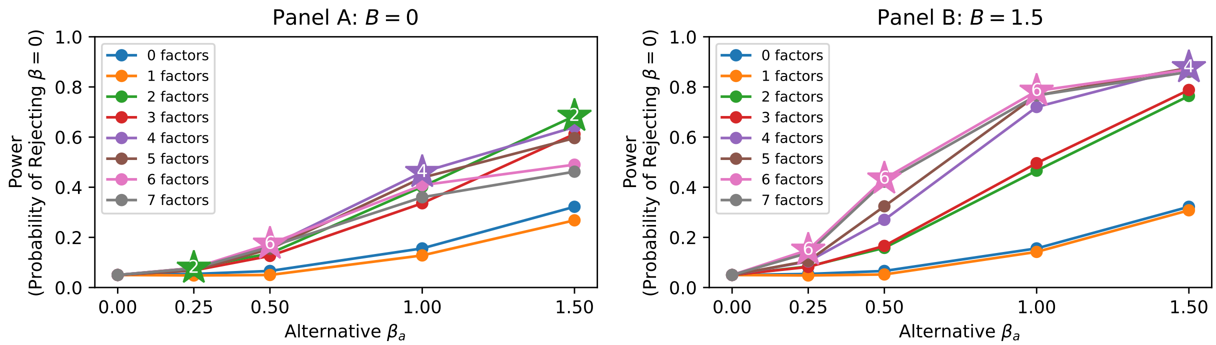
Power Simulation and Selection of (B, J). We begin with a power simulation, as described in Algorithm 4. We select our null hypothesis as $\beta_{\text{null}} = 0$, and we simulate data under the alternative hypothesis $\pi_a = 1$, $\beta_a = 1.5$, although we explore power under different values of β_a. The hypothesis of $\pi_a = 1$ corresponds to a first stage in which national military spending growth is allocated across states in exact proportion to each state's initial share of national military spending. We choose to focus on $\beta_a = 1.5$ because it represents a common view about the size of the fiscal multiplier in the US. We explore power under $B = 0$, which corresponds to the null hypotheses, and under $B = 1.5$, which corresponds to our alternative hypothesis and may be close to the priors of many researchers.

We show the results of the power simulation in Figure 2. The figure reveals very substantial power improvements from using the optimal instrument, for $J = 2$ and above. For $\beta_a = 1.5$, the power of the test using the unweighted instrument is 0.32. In the left panel ($B = 0$), the power increases to 0.68 using the optimal number of factors ($J = 2$). In the right panel, the power increases to 0.88 using the optimal number of factors ($J = 4$). Moreover, in both panels, there are substantial power gains for all choices that we examine except $J = 1$, which interestingly does not meaningfully improve power.

The power simulation reveals that tests based on the unweighted instrument have low power once size distortions have been corrected. For example, under a true regional fiscal multiplier of 1.5 and a first-stage coefficient of 1, our baseline randomization inference method would only reject the null hypothesis of $\beta = 0$ one third of the time.

However, the power simulation also demonstrates that our feasible optimal instrument substantially improves power. Under plausible parameter values, the optimal instrument

Figure 2: Power Simulation for Randomization Inference, Varying J



Notes: The figure plots the probability of rejecting the null hypothesis $\beta = 0$, under simulations based on $\pi_a = 1$, and various β_a. Panel A (left) shows simulation results under $B = 0$, and Panel B (right) shows simulation results under $B = 1.5$. Details of the simulation are described in Algorithm 4. Each curve corresponds to an optimal instrument, $Z^* (B, J)$, where B is as indicated in the title and J varies. The curve for $J = 0$ corresponds to the original, unweighted instrument, and is therefore the same in both plots. For all instruments, the power at $\beta_a = 0$ is 0.05 by construction. For each value β_a, we indicate with a star the J that maximizes power.

can increase power by a factor of 2.13 ($B = 0$) or 2.75 ($B = 1.5$). Inference based on this optimal instrument can provide a much sharper picture of the regional fiscal multiplier.

Results: Randomization Inference with Optimal Instruments. We now implement randomization inference with the feasible optimal instrument. We use the power-maximizing values for J identified above: $J = 2$ for $B = 0$ and $J = 4$ for $B = 1.5$. Our results are shown in Table 4.

In Panel A, we report point estimates and confidence intervals based on asymptotically valid methods. We estimate $\hat{\beta}^{\text{opt}}(B, J)$ as 1.276 for $(B, J) = (0, 2)$ and 1.156 for $(B, J) = (1.5, 4)$. Both estimates are substantially smaller than the unweighted point estimate, $\hat{\beta}^{2SLS} = 2.477$. However, the confidence intervals suggest substantial uncertainty.

In Panel B, we report confidence intervals based on randomization inference. In the first two rows, we report results for $B = 0$ and $B = 1.5$, using $J = 2$ and $J = 4$, respectively, based on the power simulation (see Figure 2). In the second two rows, we report results setting $B = \beta_0$; we show these results for $J = 2$ and $J = 4$. Across all four of these methods, we cannot rule out regional fiscal multipliers as low as -0.08 at the 5% level.

Despite the fact that the optimal instrument greatly increased statistical power, Table 4 shows that the data cannot rule out low values of the regional fiscal multiplier β. Following Nakamura and Steinsson's economic interpretation of β, we cannot rule out the "plain-vanilla Neoclassical model" with this evidence.

If our test is powerful, why can we not rule out low values of β? One explanation is

Table 4: 95% Confidence Intervals for Optimal IV Estimate of Regional Fiscal Multiplier in Nakamura and Steinsson (2014)

Panel A: Point Estimates and Asymptotic Inference

		Estimator 1: $B=0, J=2$		Estimator 2: $B=1.5, J=4$	
Point Estimate		1.276		1.156	
Cluster by State	$\text{se}_{\hat{\beta}}$	-0.052	2.606	0.052	2.093
	se_{β_0}	-0.103	∞	0.116	3.178
Cluster Two-Way	$\text{se}_{\hat{\beta}}$	0.216	2.337	0.219	2.093
	se_{β_0}	$-\infty$	∞	$-\infty$	∞
Two-way HAC, $L=3$	$\text{se}_{\hat{\beta}}$	0.366	2.187	0.325	1.987
	se_{β_0}	$-\infty$	∞	$-\infty$	∞

Panel B: Randomization Inference

Estimator 1:	$B=0, J=2$	-0.08	3.06
Estimator 2:	$B=1.5, J=4$	0.12	2.34
Estimator 3:	$B=\beta_0, J=2$	-0.08	>7.00
Estimator 4:	$B=\beta_0, J=4$	-0.28	5.54

Notes: This table shows 95% confidence intervals for the regional fiscal multiplier, estimated in the setting of Nakamura and Steinsson (2014) using efficient instruments. All results are based on the instrumental variable strategy that interacts defense spending growth with the pre-period share of military procurement spending in state output. In Panel A, we report the point estimates $\hat{\beta}^{\text{opt}}(B, J)$ and asymptotically based confidence intervals. We use two calibrations of the tuning parameters; in each case, J is chosen optimally based on a power test given the chosen B, the null hypothesis $\beta_0 = 0$, and the alternative hypothesis $(\beta_a, \pi_a) = (1.5, 1.0)$. The rows show results clustering by state, using two-way clustering (state and year), and using two-way HAC standard errors with a kernel bandwidth of three years. We provide confidence intervals constructed with conventional standard errors ($\text{se}_{\hat{\beta}}$) and standard errors that impose the null hypothesis β_0 to estimate the residuals (se_{β_0}); the latter confidence intervals coincide with those of the weak-instrument-robust AR-LM method. In Panel B, we report confidence intervals from optimal randomization inference. The method is described in Algorithm 3. In the first two rows, we fix B at the indicated value; the second two rows, we vary B to equal the tested null hypothesis β_0. The "> 7.00" in the third row indicates that the upper end-point exceeds the largest grid point considered.

that the true β is low. Another possibility is that β is high but we made a "false negative" error. Our power simulation suggests that, given a high value of the multiplier ($\beta_a = 1.5$), we would fail to reject the null of $\beta = 0$ between 12% and 32% of the time depending on B and J. Rejecting values of β that are above zero but still low is even more difficult.

5 Conclusion

Regional-exposure designs are ubiquitous in current empirical practice. Researchers use these designs in the hopes that regional data will provide them with more credible identification and that a greater number of observations will provide precise estimates.

We study how unobserved aggregate shocks affect regional-exposure econometric designs. We argue that the most plausible source of identification is the orthogonality of an observed aggregate shock from unobserved aggregate shocks, and that the presence of these unobserved shocks induces a factor structure to the model residual. We show that the standard econometric practice of clustering standard errors by cross-sectional units (e.g. regions) may understate uncertainty because it fails to account for the systematic correlations induced by heterogeneous responses to aggregate shocks. To remedy this issue with inference, we propose more robust asymptotic methods and finite-sample-valid randomization inference. To improve statistical power, we propose a feasible optimal instrument that reweights the data to account for units' exposure to common shocks. In an application to the study of Nakamura and Steinsson (2014), we show that standard confidence intervals give poor coverage, that corrected confidence intervals give correct coverage, and that the feasible optimal instrument substantially improves power.

We provide three recommendations for practice. First, we caution against clustering standard errors by region. Our results show, in theory and practice, that this method is not robust to the presence of cross-regional correlations in model residuals. Second, for correct inference, we provide two options. One option is to use two-way clustering (with or without HAC correction), to account for the data's correlation structure. Alternatively, researchers can use randomization inference, which is weak-instrument robust and accounts for the correlation structure of the data by modeling the shock process. Third, we suggest considering a feasible optimal instrument. We found that a method based on estimating a factor structure in the data substantially improved power.

An important issue that our paper did not address was how to correct inference with very few time periods. The econometric issues we identify could all arise in these settings, but solutions that rely on $T \to \infty$ may not be useful. A promising path is to implement randomization inference with a different procedure to estimate the data-generating process for the underlying shock, which does not rely on having many time periods of observation. We leave further study of small T settings to future work.

References

Abadie, A., S. Athey, G. W. Imbens, and J. M. Wooldridge (2023): "When should you adjust standard errors for clustering?" *The Quarterly Journal of Economics*, 138, 1–35.

Abadie, A., A. Diamond, and J. Hainmueller (2010): "Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program," *Journal of the American Statistical Association*, 105, 493–505.

Abadie, A. and J. Gardeazabal (2003): "The Economic Costs of Conflict: A Case Study of the Basque Country," *American Economic Review*, 93, 113–132.

Acemoglu, D., V. M. Carvalho, A. Ozdaglar, and A. Tahbaz-Salehi (2012): "The Network Origins of Aggregate Fluctuations," *Econometrica*, 80, 1977–2016.

Adao, R., M. Kolesár, and E. Morales (2019): "Shift-share designs: Theory and inference," *The Quarterly Journal of Economics*, 134, 1949–2010.

Anderson, T. W. and H. Rubin (1949): "Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations," *The Annals of Mathematical Statistics*, 20, 46–63.

Andrews, D. W. K., M. J. Moreira, and J. H. Stock (2006): "Optimal Two-Sided Invariant Similar Tests for Instrumental Variables Regression," *Econometrica*, 74, 715–752.

Arkhangelsky, D. and V. Korovkin (2019): "On Policy Evaluation with Aggregate Time-Series Shocks," Tech. Rep. arXiv:1905.13660, arXiv.

Autor, D. H., D. Dorn, and G. H. Hanson (2013): "The China Syndrome: Local Labor Market Effects of Import Competition in the United States," *American Economic Review*, 103, 2121–2168.

Bai, J. and S. Ng (2002): "Determining the Number of Factors in Approximate Factor Models," *Econometrica*, 70, 191–221.

Baqaee, D. R. and E. Farhi (2019): "The Macroeconomic Impact of Microeconomic Shocks: Beyond Hulten's Theorem," *Econometrica*, 87, 1155–1203.

Bartik, T. (1991): *Who benefits from state and local economic development policies*, Kalamazoo, Mich: W.E. Upjohn Institute for Employment Research.

Bertrand, M., E. Duflo, and S. Mullainathan (2004): "How much should we trust differences-in-differences estimates?" *The Quarterly Journal of Economics*, 119, 249–275.

Borusyak, K. and P. Hull (2021a): "Efficient Estimation with Non-Random Exposure to Exogenous Shocks," Working paper, Brown University.

——— (2021b): "Non-Random Exposure to Exogenous Shocks: Theory and Applications," Working Paper w27845, National Bureau of Economic Research.

Borusyak, K., P. Hull, and X. Jaravel (2022): "Quasi-experimental shift-share research designs," *The Review of Economic Studies*, 89, 181–213.

Cameron, A. C., J. B. Gelbach, and D. L. Miller (2011): "Robust inference with multiway clustering," *Journal of Business & Economic Statistics*, 29, 238–249.

Cameron, A. C. and D. L. Miller (2015): "A practitioner's guide to cluster-robust inference," *Journal of Human Resources*, 50, 317–372.

Card, D. (2001): "Immigrant Inflows, Native Outflows, and the Local Labor Market Impacts of Higher Immigration," *Journal of Labor Economics*, 19, 22–64.

Chamberlain, G. (1987): "Asymptotic efficiency in estimation with conditional moment restrictions," *Journal of Econometrics*, 34, 305–334.

——— (1992): "Efficiency Bounds for Semiparametric Regression," *Econometrica*, 60, 567.

Chodorow-Reich, G. (2019): "Geographic cross-sectional fiscal spending multipliers: What have we learned?" *American Economic Journal: Economic Policy*, 11, 1–34.

——— (2020): "Regional data in macroeconomics: Some advice for practitioners," *Journal of Economic Dynamics and Control*, 115, 103875.

Chodorow-Reich, G., P. T. Nenov, and A. Simsek (2021): "Stock market wealth and the real economy: A local labor market approach," *American Economic Review*, 111, 1613–1657.

Davezies, L., X. D'Haultfoeuille, and Y. Guyonvarch (2021): "Empirical process results for exchangeable arrays," *The Annals of Statistics*, 49, 845 – 862.

Driscoll, J. C. and A. C. Kraay (1998): "Consistent Covariance Matrix Estimation with Spatially Dependent Panel Data," *Review of Economics and Statistics*, 80, 549–560.

DUBE, O. AND J. F. VARGAS (2013): "Commodity Price Shocks and Civil Conflict: Evidence from Colombia," *The Review of Economic Studies*, 80, 1384–1421.

FINLAY, K. AND L. M. MAGNUSSON (2009): "Implementing weak-instrument robust tests for a general class of instrumental-variables models," *The Stata Journal*, 9, 398–421.

FOERSTER, A. T., P.-D. G. SARTE, AND M. W. WATSON (2011): "Sectoral versus Aggregate Shocks: A Structural Factor Analysis of Industrial Production," *Journal of Political Economy*, 119, 1–38.

GABAIX, X. (2011): "The Granular Origins of Aggregate Fluctuations," *Econometrica*, 79, 733–772.

GOLDSMITH-PINKHAM, P., I. SORKIN, AND H. SWIFT (2020): "Bartik instruments: What, when, why, and how," *American Economic Review*, 110, 2586–2624.

GUREN, A., A. MCKAY, E. NAKAMURA, AND J. STEINSSON (2021): "What do we learn from cross-regional empirical estimates in macroeconomics?" *NBER Macroeconomics Annual*, 35, 175–223.

MACKINNON, J. G., M. Ø. NIELSEN, AND M. D. WEBB (2021): "Wild Bootstrap and Asymptotic Inference With Multiway Clustering," *Journal of Business and Economic Statistics*, 39, 505–519.

——— (2022): "Cluster-robust inference: A guide to empirical practice," *Journal of Econometrics*.

MAGNUSSON, L. M. (2010): "Inference in limited dependent variable models robust to weak identification," *The Econometrics Journal*, 13, S56–S79.

MENZEL, K. (2021): "Bootstrap With Cluster-Dependence in Two or More Dimensions," *Econometrica*, 89, 2143–2188.

MIAN, A. AND A. SUFI (2014): "What Explains the 2007-2009 Drop in Employment?" *Econometrica*, 82, 2197–2223.

MIGLIORETTI, D. L. AND P. J. HEAGERTY (2007): "Marginal modeling of nonnested multilevel data using standard software," *American Journal of Epidemiology*, 165, 453–463.

MINTZ, A. (1992): *The Political Economy of Military Spending in the United States*, New York: Routledge.

Moreira, M. J. (2003): "A Conditional Likelihood Ratio Test for Structural Models," *Econometrica*, 71, 1027–1048.

Nakamura, E. and J. Steinsson (2014): "Fiscal Stimulus in a Monetary Union: Evidence from US Regions," *American Economic Review*, 104, 753–792.

Nunn, N. and N. Qian (2014): "US Food Aid and Civil Conflict," *American Economic Review*, 104, 1630–1666.

Stock, J. H. and M. W. Watson (2002): "Forecasting Using Principal Components From a Large Number of Predictors," *Journal of the American Statistical Association*, 97, 1167–1179.

Thompson, S. B. (2011): "Simple formulas for standard errors that cluster by both firm and time," *Journal of Financial Economics*, 99, 1–10.

A Omitted Proofs

A.1 Proof of Proposition 1

Proof. We start with Part 1. Here, we hold the number of time periods T fixed, while the number of regions $N \to \infty$. We thus have a fixed number of time fixed effects, while the region fixed effects are nuisance parameters. Note that estimation with time and region fixed effects is equivalent to double-demeaning the regressors and instruments. We will thus work with the double-demeaned instruments and regressors.

First, we will prove that $\frac{1}{NT}\sum_{i,t}\tilde{Z}_{it}u_{it} \xrightarrow{p} 0$. We start by proving that $\frac{1}{NT}\sum_{i,t} Z_{it}u_{it} \xrightarrow{p} 0$:

$$\begin{aligned}
\frac{1}{NT}\sum_{i,t} Z_{it}u_{it} &= \frac{1}{NT}\sum_{i,t}\eta_i' S_t \cdot \lambda_i' F_t + \frac{1}{NT}\sum_{i,t}\eta_i' S_t \cdot \varepsilon_{it} \\
&= \frac{1}{N}\sum_i \frac{1}{T}\sum_t \mathbf{tr}\left((\eta_i\lambda_i')(F_t S_t')\right) + \frac{1}{N}\sum_i \frac{1}{T}\sum_t \eta_i' S_t \cdot \varepsilon_{it} \\
&= \frac{1}{N}\sum_i \frac{1}{T}\mathbf{tr}\left((\eta_i\lambda_i')\sum_t (F_t S_t')\right) + \frac{1}{N}\sum_i \frac{1}{T}\sum_t \eta_i' S_t \cdot \varepsilon_{it} \qquad (28) \\
&\xrightarrow{p} \frac{1}{T}\mathbb{E}\left[\mathbf{tr}\left((\eta_i\lambda_i')\sum_t (F_t S_t')\right)\right] + \frac{1}{T}\sum_t \mathbb{E}\left[\eta_i' S_t \cdot \varepsilon_{it}\right] \\
&= 0
\end{aligned}$$

The second to last line applies the Weak Law of Large Numbers because $\left(\eta_i, \lambda_i, (\varepsilon_{it})_{t=1}^{T}\right)$ are drawn i.i.d. across regions. The last line uses $\mathbb{E}\left[\eta_i\lambda_i'\right] = 0$ from Condition 1 and $\mathbb{E}\left[\eta_i' S_t \cdot \varepsilon_{it}\right] = 0$ from Assmption 2.

Next, we prove that $\frac{1}{NT}\sum_{i,t}\bar{Z}_i u_{it} \xrightarrow{p} 0$. This proceeds similarly to the above. We have:

$$\begin{aligned}
\frac{1}{NT}\sum_{i,t}\bar{Z}_i u_{it} &= \frac{1}{NT}\sum_{i,t}\frac{1}{T}\left(\sum_s Z_{is}\right)u_{it} \\
&= \frac{1}{N}\sum_{i,t}\frac{1}{T^2}\eta_i'\left(\sum_s S_s\right)u_{it} \\
&= \frac{1}{N}\sum_i \frac{1}{T^2}\mathbf{tr}\left((\eta_i\lambda_i')\sum_t F_t\left(\sum_s S_s'\right)\right) + \frac{1}{N}\sum_i\frac{1}{T^2}\sum_t \eta_i'\left(\sum_s S_s\right)\cdot\varepsilon_{it} \qquad (29) \\
&\xrightarrow{p} \frac{1}{T^2}\mathbb{E}\left[\mathbf{tr}\left((\eta_i\lambda_i')\sum_t F_t\left(\sum_s S_s'\right)\right)\right] + \frac{1}{T^2}\sum_t\mathbb{E}\left[\eta_i'\left(\sum_s S_s\right)\cdot\varepsilon_{it}\right] \\
&= 0
\end{aligned}$$

The last two lines use the same logic as the last two lines of the previous derivation.

Next, we show that $\frac{1}{NT}\sum_{i,t}\left(\bar{Z}_t - \bar{Z}\right)u_{it} \xrightarrow{p} 0$. To do this, it is sufficient to show that $\frac{1}{N}\sum_i \bar{Z}_t u_{it} \xrightarrow{p} 0$ for each t. From there, since the number of time periods is fixed, we can add up to get our desired result. We have:

$$\begin{aligned}
\frac{1}{N}\sum_i \bar{Z}_t u_{it} &= \frac{1}{N}\sum_i\frac{1}{N}\left(\sum_j \eta_j' S_t\right)\left(\lambda_i' F_t + \varepsilon_{it}\right) \\
&= \frac{1}{N}\sum_i\frac{1}{N}\left(\sum_j \eta_j'\right)S_t\left(\lambda_i' F_t + \varepsilon_{it}\right) \\
&= \frac{1}{N}\left(\sum_j \eta_j\right)\cdot\frac{1}{N}\sum_i S_t\left(\lambda_i' F_t + \varepsilon_{it}\right) \qquad (30) \\
&\xrightarrow{p} \mathbb{E}\left[\eta_j\right]\cdot\mathbb{E}\left[S_t\left(\lambda_i' F_t + \varepsilon_{it}\right)\right] \\
&= 0
\end{aligned}$$

The second to last line uses the weak law of large numbers, and the last line uses $\mathbb{E}\left[\eta_j'\right] = 0$. Adding up, we thus have that $\frac{1}{NT}\sum_{i,t}\tilde{Z}_{it}u_{it} \xrightarrow{p} 0$.

Next, we show that there exists a finite and full rank matrix Q such that $\frac{1}{NT}\sum_{i,t}\tilde{Z}_{it}\tilde{X}_{it}' \rightarrow_p Q$ as $N \rightarrow \infty$. Note that, based on the assumptions we have made, (X_{it}, Z_{it}) is drawn i.i.d.

across regions. We thus have $\bar{Z}_t \xrightarrow{p} \mathbb{E}[Z_{it} \mid t]$, and similarly for $\bar{Z}$,$\bar{X}$, and $\bar{X}_t$. We thus have:

$$\begin{aligned}
\frac{1}{NT}\sum_{i,t}\tilde{Z}_{it}\tilde{X}_{it}' &= \frac{1}{NT}\sum_{i,t}\left(Z_{it}-\bar{Z}_i-\bar{Z}_t+\bar{Z}\right)\left(X_{it}-\bar{X}_i-\bar{X}_t+\bar{X}\right)' \\
&\xrightarrow{p} \frac{1}{N}\sum_i\frac{1}{T}\sum_t\left(Z_{it}-\bar{Z}_i-\mathbb{E}[Z_{it}\mid t]+\mathbb{E}[Z_{it}]\right)\left(X_{it}-\bar{X}_i-\mathbb{E}[X_{it}\mid t]+\mathbb{E}[X_{it}]\right)' \\
&\xrightarrow{p} \mathbb{E}\left[\frac{1}{T}\sum_t\left(Z_{it}-\bar{Z}_i-\mathbb{E}[Z_{it}\mid t]+\mathbb{E}[Z_{it}]\right)\left(X_{it}-\bar{X}_i-\mathbb{E}[X_{it}\mid t]+\mathbb{E}[X_{it}]\right)'\right] \\
&= \mathbb{E}\left[\tilde{Z}_{it}\tilde{X}_{it}'\right]
\end{aligned} \tag{31}$$

By assumption in the statement of Proposition 1, this expectation is finite and of full rank.

Finally, to show the desired result, we apply the continuous mapping theorem, relying on the fact that matrix inversion is continuous wherever the matrix is full rank. We have:

$$\begin{aligned}
\hat{\beta} &= \left(\frac{1}{NT}\sum_{i,t}\tilde{Z}_{it}\tilde{X}_{it}'\right)^{-1}\frac{1}{NT}\sum_{i,t}\tilde{Z}_{it}Y_{it} \\
&= \left(\frac{1}{NT}\sum_{i,t}\tilde{Z}_{it}\tilde{X}_{it}'\right)^{-1}\frac{1}{NT}\sum_{i,t}\tilde{Z}_{it}\left(\tilde{X}_{it}'\beta+u_{it}\right) \\
&= \beta+\left(\frac{1}{NT}\sum_{i,t}\tilde{Z}_{it}\tilde{X}_{it}'\right)^{-1}\frac{1}{NT}\sum_{i,t}\tilde{Z}_{it}u_{it} \\
&\xrightarrow{p} \beta+\mathbb{E}\left[\tilde{Z}_{it}\tilde{X}_{it}'\right]^{-1}\cdot 0 \\
&= \beta
\end{aligned} \tag{32}$$

This proves the claim of Part 1.

The proof of Part 2 is analogous to the proof for Part 1, but using a law of large numbers for stationary and strongly mixing time series rather than for i.i.d. regions. We omit this for brevity. □

A.2 Proof of Lemma 1

Proof. Using the definition of Z_{it}, we write

$$\omega(i,j,t,s) = \mathbb{E}[\eta_i' S_t \cdot \lambda_i' F_t \cdot \eta_j' S_s \cdot \lambda_j' F_s] \tag{33}$$

We first show Part 1. For $i \neq j$, we manipulate the inside of the expectation

$$\omega(i,j,t,s) = \mathbb{E}[\mathbf{tr}(\eta_i' S_t \cdot \lambda_i' F_t \cdot \eta_j' S_s \cdot \lambda_j' F_s)] = \mathbb{E}[\mathbf{tr}(\eta_i \lambda_i' \cdot F_t S_s' \eta_j \lambda_j' F_s S_t')] \tag{34}$$

using rearrangement and the cyclic property of the trace. We observe that, due to the linearity of the trace, we can write $\mathbb{E}[\mathbf{tr}[A]] = \mathbf{tr}[\mathbb{E}[A]]$ for a real-matrix-valued random variable A. We then use the assumed independence of (η_i, λ_i) from the vector $(\eta_j, \lambda_j, F_s, F_t, S_s, S_t)$, encapsulating independence across regions and indepence of cross-sectional from time-series variables, to write

$$\omega(i,j,t,s) = \mathbf{tr}(\mathbb{E}[\eta_i \lambda_i' \cdot F_t S_s' \eta_j \lambda_j' F_s S_t']) = \mathbf{tr}(\mathbb{E}[\eta_i \lambda_i'] \cdot \mathbb{E}[F_t S_s' \eta_j \lambda_j' F_s S_t']) \tag{35}$$

We then use the identification from shares condition to observe that $\mathbb{E}[\eta_i \lambda_i']$ is a $K \times J$ matrix of zeros, and hence $\omega(i,j,t,s) = 0$.

We next show Part 2. We manipulate terms to write

$$\omega(i,j,t,s) = \mathbb{E}[\eta_i'(S_t F_t')\lambda_i \cdot \eta_j'(S_s F_s')\lambda_j)] \tag{36}$$

We now condition down on the values of $(\lambda_i, \lambda_j, \eta_i, \eta_j)$ to write

$$\omega(i,j,t,s) = \mathbb{E}[\eta_i'(\mathbb{E}[S_t F_t' | \lambda_i, \lambda_j, \eta_i, \eta_j])\lambda_i \cdot \eta_j'(\mathbb{E}[S_s F_s' | \lambda_i, \lambda_j, \eta_i, \eta_j])\lambda_j)] \tag{37}$$

where we observe that $\mathbb{E}[S_t F_t' | \lambda_i, \lambda_j, \eta_i, \eta_j] = \mathbb{E}[S_t F_t'] = 0$ due to the assumed conditional independence and the identification from shocks condition; similarly, $\mathbb{E}[S_s F_s' | \lambda_i, \lambda_j, \eta_i, \eta_j] = \mathbb{E}[S_s F_s'] = 0$. Hence, in these cases, $\omega(i,j,t,s) = 0$. This proves Lemma 1 as stated.

We now show, additionally, that the zero covariances in Lemma 1 are consistently estimated. In particular: □

Lemma 2. *Let* $\tilde{\omega}(i,j,t,s) = \mathbb{E}\left[\tilde{Z}_{it} \cdot \tilde{\lambda}_i' \tilde{F}_t \cdot \tilde{Z}_{js} \cdot \tilde{\lambda}_j' \tilde{F}_s\right]$ *be the demeaned-factor-component covariance between units* (i,t) *and* (j,s)*, using the double demeaned instrument. If Assumptions 1 and 2 hold, then*

1. *If identification comes from shares (Condition 1) and* (η_i, λ_i) *is independent across regions, then* $\tilde{\omega}(i,j,t,s) = O(1/N^2)$ *for all* $i \neq j$*.*

2. *If identification comes from shocks (Condition 2) and* (S_t, F_t) *is independent across time, then* $\tilde{\omega}(i,j,t,s) = O(1/T^2)$ *for all* $t \neq s$*.*

Proof. To prove this, we first observe that the double-demeaned instrument is

$$\begin{aligned}
\tilde{Z}_{it} &= Z_{it} - \bar{Z}_i - \bar{Z}_t + \bar{Z} \\
&= \eta_i' S_t - \eta_i' \bar{S} - \bar{\eta}' S_t + \overline{\eta_i' S_t} \\
&= (\eta_i - \bar{\eta})' \left(S_t - \bar{S}\right) - \bar{\eta}' \bar{S} + \overline{\eta_i' S_t} \\
&= (\eta_i - \bar{\eta})' \left(S_t - \bar{S}\right)
\end{aligned} \tag{38}$$

where $\bar{\eta}'\bar{S} = \overline{\eta_i' S_t}$ because we have assumed a balanced panel. We will define $\tilde{\eta}_i := \eta_i - \bar{\eta}$ and $\tilde{S}_t := S_t - \bar{S}$. Note that an identical argument shows that $\tilde{u}_{it} - \tilde{\varepsilon}_{it} = \tilde{\lambda}_i' \tilde{F}_t$.

To prove case two, we re-write $\tilde{\omega}$ as

$$\tilde{\omega}(i,j,t,s) = \mathbb{E}[\tilde{\eta}_i' \tilde{S}_t \cdot \tilde{\lambda}_i' \tilde{F}_t \cdot \tilde{\eta}_j' \tilde{S}_s \cdot \tilde{\lambda}_j' \tilde{F}_s] \tag{39}$$

We can rewrite the above as a sum. Let k and k' index entries of the observed shock, S, and let l and l' index entries of the unobserved factor shock, F. We then have:

$$\begin{aligned}
\tilde{\omega}(i,j,t,s) &= \sum_k \sum_{k'} \sum_l \sum_{l'} \mathbb{E}\left[\tilde{\eta}_i^k \tilde{S}_t^k \tilde{\lambda}_i^l \tilde{F}_t^l \tilde{\eta}_j^{k'} \tilde{S}_s^{k'} \tilde{\lambda}_j^{l'} \tilde{F}_s^{l'}\right] \\
&= \sum_k \sum_{k'} \sum_l \sum_{l'} \mathbb{E}\left[\tilde{S}_t^k \tilde{F}_t^l \tilde{S}_s^{k'} \tilde{F}_s^{l'}\right] \cdot \mathbb{E}\left[\tilde{\eta}_i^k \tilde{\lambda}_i^l \tilde{\eta}_j^{k'} \tilde{\lambda}_j^{l'}\right] \\
&= \sum_k \sum_{k'} \sum_l \sum_{l'} \mathbb{E}\left[\left(S_t^k - \bar{S}^k\right)\left(S_s^{k'} - \bar{S}^{k'}\right)\left(F_t^l - \bar{F}^l\right)\left(F_s^{l'} - \bar{F}^{l'}\right)\right] \cdot \mathbb{E}\left[\tilde{\eta}_i^k \tilde{\lambda}_i^l \tilde{\eta}_j^{k'} \tilde{\lambda}_j^{l'}\right]
\end{aligned} \tag{40}$$

where the second line uses the assumed independence of cross-sectional and time-series variables.

We then consider the case where $t \neq s$. We first observe that $\mathbb{E}\left[S_{t_0}^k S_{t_1}^{k'} F_{t_2}^l F_{t_3}^{l'}\right] = 0$ if $t_0 \neq t_1$ or $t_2 \neq t_3$. To show this, we consider all the relevant cases. We observe that for any $r \neq t$, $\mathbb{E}\left[S_t^k S_s^{k'} F_t^l F_r^{l'}\right] = \mathbb{E}[S_s^k F_r^{l'}]\mathbb{E}[S_t^k F_t^l] = 0$, since $(S_s^k F_r^{l'})$ is independent from $(S_t^k F_t^l)$ $\mathbb{E}[S_t^k F_t^l] = 0$. Next, for $r \neq t$ and for $w \neq r$, we observe that $\mathbb{E}\left[S_t^k S_r^{k'} F_w^l F_w^{l'}\right] = \mathbb{E}[S_t^k F_w^l F_w^{l'}]\mathbb{E}[S_r^{k'}] = 0$, since (S_r^k) is independent from $(S_t^k, F_w^l, F_w^{l'})$ and $\mathbb{E}[S_r^{k'}] = 0$. Next, for $r \neq t$, we observe that $\mathbb{E}\left[S_t^k S_r^{k'} F_r^l F_r^{l'}\right] = \mathbb{E}[S_r^{k'} F_r^l F_r^{l'}]\mathbb{E}[S_t^{k'}] = 0$, since (S_t^k) is independent from $(S_t^k, F_t^l, F_t^{l'})$ and $\mathbb{E}[S_t^k] = 0$. Finally, analogous arguments apply to show the same when F and S are switched.

We now apply this rule, as well as the iid nature of draws across time, to simplify further:

$$
\begin{aligned}
\mathbb{E}\left[\left(S_t^k S_s^{k'} - S_t^k \bar{S}^{k'} - \bar{S}^k S_s^{k'} + \bar{S}^k \bar{S}^{k'}\right)\left(F_t^l F_s^{l'} - F_t^l \bar{F}^{l'} - \bar{F}^l F_s^{l'} + \bar{F}^l \bar{F}^{l'}\right)\right] \\
= \mathbb{E}\left[\left(-S_t^k \bar{S}^{k'} - \bar{S}^k S_s^{k'} + \bar{S}^k \bar{S}^{k'}\right)\left(-F_t^l \bar{F}^{l'} - \bar{F}^l F_s^{l'} + \bar{F}^l \bar{F}^{l'}\right)\right] \\
= \mathbb{E}\left[\frac{2}{T^2} S_t^k S_t^{k'} F_t^l F_t^{l'} + \frac{2}{T^2} S_t^k S_t^{k'} F_s^l F_s^{l'} - \frac{2}{T^3} S_t^k S_t^{k'} \sum_{w=1}^{T} F_w^l F_w^{l'} - \right. \\
\left. \frac{2}{T^3} F_t^l F_t^{l'} \sum_{r=1}^{T} S_r^k S_r^{k'} + \frac{1}{T^4}\left(\sum_{r=1}^{T} S_r^k S_r^{k'}\right)\left(\sum_{w=1}^{T} F_w^l F_w^{l'}\right)\right] \\
=: \frac{1}{T^2} M_0(T, k, k', l, l')
\end{aligned} \tag{41}
$$

where we define $M_0(T, k, k', l, l')$ in the last line. We next observe that $|M_0(T, k, k', l, l')| < \bar{M}_0 < \infty$, for all T and (k, k', l, l'), for some constant $\bar{M}_0$ that does not depend on T or the indices, because of our assumption of bounded moments. We also define $M_1(N, k, k', l, l') = \mathbb{E}\left[\tilde{\eta}_i^k \tilde{\lambda}_i^l \tilde{\eta}_j^{k'} \tilde{\lambda}_j^{l'}\right]$ and similarly observe that $|M_1(N, k, k', l, l')| < \bar{M}_1$, for all N and (k, k', l, l'), because of bounded moments.

This allows us to write, for $t \neq s$

$$
\begin{aligned}
\tilde{\omega}(i, j, t, s) &= \sum_{k,k',l,l'} \left[\mathbb{E}\left[\left(S_t^k S_s^{k'} - S_t^k \bar{S}^{k'} - \bar{S}^k S_s^{k'} + \bar{S}^k \bar{S}^{k'}\right)\left(F_t^l F_s^{l'} - F_t^l \bar{F}^{l'} - \bar{F}^l F_s^{l'} + \bar{F}^l \bar{F}^{l'}\right)\right]\right. \\
&\qquad \left. \cdot \mathbb{E}\left[\tilde{\eta}_i^k \tilde{\lambda}_i^l \tilde{\eta}_j^{k'} \tilde{\lambda}_j^{l'}\right]\right] \\
&= \frac{1}{T^2} \sum_{k,k',l,l'} M_0(T, k, k', l, l') M_1(N, k, k', l, l')
\end{aligned} \tag{42}
$$

and to moreover observe that $|\tilde{\omega}(i, j, t, s)| < \frac{1}{T^2} J^2 K^2 \bar{M}_0 \bar{M}_1$. Therefore, $\tilde{\omega}(i, j, t, s)$ is $O(1/T^2)$ for $t \neq s$.

The proof of case one is analogous, and we omit it for brevity. □

A.3 Proof of Proposition 2

Proof. We first show that $\hat{V}^{CR} \to^p V^{CR}$. The clustered estimator for the asymptotic variance is $\hat{V}^{CR} = \hat{Q}^{-1} \hat{\Omega}^{CR} \left(\hat{Q}'\right)^{-1}$, where

$$
\hat{Q} = \frac{1}{NT} \sum_{i,t} \tilde{Z}_{it} \tilde{X}_{it}' \qquad \text{and} \qquad \hat{\Omega}^{CR} = \frac{1}{N} \sum_{i} \left[\left(\frac{1}{T} \sum_{t} \tilde{Z}_{it} \hat{u}_{it}\right)\left(\frac{1}{T} \sum_{t} \tilde{Z}_{it} \hat{u}_{it}\right)'\right] \tag{43}
$$

and where $\hat{u}_{it} := \tilde{Y}_{it} - \tilde{X}'_{it}\hat{\beta}$. We have already shown that $\hat{Q} \xrightarrow{p} \mathbb{E}\left[\tilde{Z}_{it}\tilde{X}'_{it}\right]$ in the proof of Proposition 1.

We thus need to prove that $\hat{\Omega}^{CR} \xrightarrow{p} \Omega$. To do this, we will break $\hat{\Omega}^{CR}$ out into its components, and then use a law of large numbers argument to show that it converges. We have:

$$
\begin{aligned}
\hat{\Omega}^{CR} &= \frac{1}{N}\sum_i \left[\left(\frac{1}{T}\sum_t \tilde{Z}_{it}\hat{u}_{it}\right)\left(\frac{1}{T}\sum_t \tilde{Z}_{it}\hat{u}_{it}\right)'\right] \\
&= \frac{1}{N}\sum_i \left[\left(\frac{1}{T}\sum_t \tilde{Z}_{it}\left(\tilde{Y}_{it} - \tilde{X}'_{it}\hat{\beta}\right)\right)\left(\frac{1}{T}\sum_t \tilde{Z}_{it}\left(\tilde{Y}_{it} - \tilde{X}'_{it}\hat{\beta}\right)\right)'\right] \\
&= \frac{1}{N}\frac{1}{T^2}\sum_i \Bigg[\left(\sum_t \tilde{Z}_{it}\tilde{Y}_{it}\right)\left(\sum_t \tilde{Z}_{it}\tilde{Y}_{it}\right)' - \left(\sum_t \tilde{Z}_{it}\tilde{Y}_{it}\right)\left(\sum_t \tilde{Z}_{it}\tilde{X}'_{it}\hat{\beta}\right) \\
&\qquad - \left(\sum_t \tilde{Z}_{it}\tilde{X}'_{it}\hat{\beta}\right)\left(\sum_t \tilde{Z}_{it}\tilde{Y}_{it}\right)' + \left(\sum_t \tilde{Z}_{it}\tilde{X}'_{it}\hat{\beta}\right)\left(\sum_t \tilde{Z}_{it}\tilde{X}'_{it}\hat{\beta}\right)'\Bigg]
\end{aligned}
\tag{44}
$$

Given our assumption of finite fourth moments, we can guarantee that the expectation of each of these components will be finite. This, combined with out previous arguments about the data being i.i.d. across regions, will let us use the law of large numbers so that each component converges to its expectation. Finally, adding in the previously proven fact that $\hat{\beta} \xrightarrow{p} \beta$, this tells us that $\hat{\Omega}^{CR} \xrightarrow{p} \mathbb{E}\left[\left(\frac{1}{T}\sum_t \tilde{Z}_{it}\tilde{u}_{it}\right)\left(\frac{1}{T}\sum_t \tilde{Z}_{it}\tilde{u}_{it}\right)'\right]$..

We next show that $AVAR\left(\sqrt{N}\cdot\hat{\beta}\right) = V^{CR}$. It is sufficient to show that $\Omega^{CR} = \Omega$. We note that

$$
\begin{aligned}
\Omega &= \text{AVAR}\left(\frac{1}{\sqrt{N}}\cdot\frac{1}{T}\sum_{i,t}\tilde{Z}_{it}\tilde{u}_{it}\right) \\
&= \lim_{N\to\infty}\frac{1}{NT^2}\mathbb{E}\left[\left(\sum_{i,t}\tilde{Z}_{it}\tilde{u}_{it}\right)\left(\sum_{i,t}\tilde{Z}_{it}\tilde{u}_{it}\right)'\right] \\
&= \lim_{N\to\infty}\frac{1}{NT^2}\mathbb{E}\left[\sum_{i,t}\sum_{j,s}\tilde{Z}_{it}\tilde{u}_{it}\tilde{Z}_{js}\tilde{u}_{js}\right]
\end{aligned}
\tag{45}
$$

To simplify this, we first consider terms where $i \neq j$. We have:

$$
\mathbb{E}\left[\tilde{Z}_{it}\tilde{u}_{it}\tilde{Z}_{js}\tilde{u}_{js}\right] = \mathbb{E}\left[\left(\tilde{\lambda}'_i\tilde{F}_t + \tilde{\varepsilon}_{it}\right)\left(\tilde{\lambda}'_j\tilde{F}_s + \tilde{\varepsilon}_{js}\right)\left(\tilde{\eta}'_i\tilde{S}_t\right)\left(\tilde{\eta}'_j\tilde{S}_s\right)\right] \tag{46}
$$

We first observe that $\mathbb{E}[\tilde{\varepsilon}_{it}\tilde{\lambda}'_j\tilde{F}_s\tilde{Z}_{it}\tilde{Z}_{js}] = \mathbb{E}[\tilde{\varepsilon}_{it}]\mathbb{E}[\tilde{\lambda}'_j\tilde{F}_s\tilde{Z}_{it}\tilde{Z}_{js}] = 0$ because ε_{it} is i.i.d.

across regions, mean zero, and independent from the factor draws and shocks. Similarly, $\mathbb{E}[\tilde{\lambda}_i' \tilde{F}_t \tilde{\varepsilon}_{js} \tilde{Z}_{it} \tilde{Z}_{js}] = 0$. We thus have

$$\begin{aligned}\mathbb{E}\left[\tilde{Z}_{it}\tilde{u}_{it}\tilde{Z}_{js}\tilde{u}_{js}\right] &= \mathbb{E}\left[\tilde{\lambda}_i'\tilde{F}_t\tilde{\lambda}_j'\tilde{F}_s\left(\tilde{\eta}_i'\tilde{S}_t\right)\left(\tilde{\eta}_j'\tilde{S}_s\right) + \tilde{\varepsilon}_{it}\tilde{\varepsilon}_{js}\left(\tilde{\eta}_i'\tilde{S}_t\right)\left(\tilde{\eta}_j'\tilde{S}_s\right)\right] \\ &= \tilde{\omega}\left(i,j,t,s\right) + \mathbb{E}\left[\tilde{\varepsilon}_{it}\tilde{\varepsilon}_{js}\left(\tilde{\eta}_i'\tilde{S}_t\right)\left(\tilde{\eta}_j'\tilde{S}_s\right)\right]\end{aligned}$$

We know, from Lemma 2, that $\tilde{\omega}\left(i,j,t,s\right) = O\left(\frac{1}{N^2}\right)$. We also know $\mathbb{E}\left[\tilde{\varepsilon}_{it}\tilde{\varepsilon}_{js}\left(\tilde{\eta}_i'\tilde{S}_t\right)\left(\tilde{\eta}_j'\tilde{S}_s\right)\right]$ is $O\left(\frac{1}{N^2}\right)$, from the proof of Proposition 4. Using this, we further simplify:

$$\begin{aligned}\Omega &= \lim_{N\to\infty}\frac{1}{NT^2}\mathbb{E}\left[\sum_{i,t}\sum_{j,s}\tilde{Z}_{it}\tilde{u}_{it}\tilde{Z}_{js}\tilde{u}_{js}\right] \\ &= \lim_{N\to\infty}\frac{1}{NT^2}\mathbb{E}\left[\sum_{i,t}\sum_{j,s}\mathbf{1}\left(i=j\right)\tilde{Z}_{it}\tilde{u}_{it}\tilde{Z}_{js}\tilde{u}_{js} + \sum_{i,t}\sum_{j,s}\mathbf{1}\left(i\neq j\right)O(1/N^2)\right] \\ &= \lim_{N\to\infty}\left[\frac{1}{NT^2}\mathbb{E}\left[\sum_{i,t}\sum_{j,s}\mathbf{1}\left(i=j\right)\tilde{Z}_{it}\tilde{u}_{it}\tilde{Z}_{js}\tilde{u}_{js}\right] + \frac{1}{NT^2}T^2N(N-1)O(1/N^2)\right] \\ &= \lim_{N\to\infty}\frac{1}{NT^2}\mathbb{E}\left[\sum_{i,t}\sum_{j,s}\mathbf{1}\left(i=j\right)\tilde{Z}_{it}\tilde{u}_{it}\tilde{Z}_{js}\tilde{u}_{js}\right] \\ &= \mathbb{E}\left[\left(\frac{1}{T}\sum_{t}\tilde{Z}_{it}\tilde{u}_{it}\right)\left(\frac{1}{T}\sum_{t}\tilde{Z}_{it}\tilde{u}_{it}\right)'\right] \\ &= \Omega^{CR}\end{aligned} \tag{47}$$

where the second step uses our result that $\mathbb{E}\left[\tilde{Z}_{it}\tilde{u}_{it}\tilde{Z}_{js}\tilde{u}_{js}\right]$ is $O\left(\frac{1}{N^2}\right)$ for $i \neq j$, the second to last step uses the Law of Large Numbers, and the last step is the definition of Ω^{CR}. $\square$

A.4 Proof of Proposition 3

Proof. Recall that $\Omega := \text{AVAR}\left(\frac{1}{\sqrt{N}}\cdot\frac{1}{T}\sum_{i,t}\tilde{Z}_{it}\tilde{u}_{it}\right)$, in the limit where $N\to\infty$. This implies that

$$\Omega = \lim_{N\to\infty}\frac{1}{NT^2}\mathbb{E}\left[\left(\sum_{i,t}\tilde{Z}_{it}\tilde{u}_{it}\right)\left(\sum_{i,t}\tilde{Z}_{it}\tilde{u}_{it}\right)'\right] = \lim_{N\to\infty}\frac{1}{NT^2}\mathbb{E}\left[\sum_{i,t}\sum_{j,s}\tilde{Z}_{it}\tilde{u}_{it}\tilde{Z}_{js}\tilde{u}_{js}\right] \tag{48}$$

In contrast,

$$\Omega^{CR} := \mathbb{E}\left[\left(\frac{1}{T}\sum_{t}\tilde{Z}_{it}\tilde{u}_{it}\right)\left(\frac{1}{T}\sum_{t}\tilde{Z}_{it}u_{it}\right)'\right] = \frac{1}{NT^2}\mathbb{E}\left[\sum_{i,t}\sum_{j,s}\mathbf{1}\left(i=j\right)\tilde{Z}_{it}\tilde{u}_{it}\tilde{Z}_{js}\tilde{u}_{js}\right] \tag{49}$$

Taking the difference between the two, pre-multiplying by $\frac{1}{N}$, and taking the limit as $N \to \infty$, we can write

$$\begin{aligned}\lim_{N\to\infty}\frac{1}{N}\left(\Omega^{CR}-\Omega\right) &= -\lim_{N\to\infty}\frac{1}{N^2T^2}\sum_{i,t}\sum_{j,s}\mathbf{1}\left(i\neq j\right)\mathbb{E}\left[\tilde{Z}_{it}\tilde{u}_{it}\tilde{Z}_{js}\tilde{u}_{js}\right]\\ &= -\lim_{N\to\infty}\frac{1}{N^2T^2}\sum_{i,t}\sum_{j,s}\mathbf{1}\left(i\neq j\right)\mathbb{E}\left[\tilde{Z}_{it}(\tilde{\lambda}_i'\tilde{F}_t+\tilde{\varepsilon}_{it})\tilde{Z}_{js}(\tilde{\lambda}_j'\tilde{F}_s+\tilde{\varepsilon}_{js})\right]\\ &= -\lim_{N\to\infty}\frac{1}{N^2T^2}\sum_{i,t}\sum_{j,s}\mathbf{1}\left(i\neq j\right)\mathbb{E}\Big[\tilde{Z}_{it}\tilde{\lambda}_i'\tilde{F}_t\tilde{Z}_{js}\tilde{\lambda}_j'\tilde{F}_s+\tilde{Z}_{it}\tilde{\lambda}_i'\tilde{F}_t\tilde{Z}_{js}\tilde{\varepsilon}_{js}+\\ &\qquad \tilde{Z}_{it}\tilde{\varepsilon}_{it}\tilde{Z}_{js}\tilde{\lambda}_j'\tilde{F}_s+\tilde{Z}_{it}\tilde{\varepsilon}_{it}\tilde{Z}_{js}\tilde{\varepsilon}_{js}\Big]\end{aligned} \tag{50}$$

For all (i,t,j,s), $\mathbb{E}[\tilde{Z}_{it}\tilde{\lambda}_i'\tilde{F}_t\tilde{Z}_{js}\tilde{\varepsilon}_{js}] = \mathbb{E}[\tilde{Z}_{it}\tilde{\lambda}_i'\tilde{F}_t\tilde{Z}_{js}]\mathbb{E}[\tilde{\varepsilon}_{js}] = 0$, where the first equality uses the fact that ε is independent of (Z,λ,F) and the second equality uses $\mathbb{E}[\tilde{\varepsilon}_{js}] = 0$. Similarly, for all (i,t,j,s), $\mathbb{E}[\tilde{Z}_{it}\tilde{\varepsilon}_{it}\tilde{Z}_{js}\tilde{\lambda}_j'\tilde{F}_s] = 0$. Thus,

$$\lim_{N\to\infty}\frac{1}{N}\left(\Omega^{CR}-\Omega\right) = \lim_{N\to\infty}-\frac{1}{N^2T^2}\sum_{i,t}\sum_{j,s}\mathbf{1}\left(i\neq j\right)\mathbb{E}\left[\tilde{Z}_{it}\tilde{\lambda}_i'\tilde{F}_t\tilde{Z}_{js}\tilde{\lambda}_j'\tilde{F}_s+\tilde{Z}_{it}\tilde{\varepsilon}_{it}\tilde{Z}_{js}\tilde{\varepsilon}_{js}\right] \tag{51}$$

We start by considering the second term. We show that this term is zero. First, we observe that $\mathbb{E}\left[\tilde{Z}_{it}\tilde{\varepsilon}_{it}\tilde{Z}_{js}\tilde{\varepsilon}_{js}\right] = \mathbb{E}\left[\tilde{Z}_{it}\tilde{Z}_{js}\right]\mathbb{E}\left[\tilde{\varepsilon}_{it}\tilde{\varepsilon}_{js}\right]$ because ε is independent of (Z,λ,F). Next,

$\left|\mathbb{E}\left[\tilde{Z}_{it}\tilde{Z}_{js}\right]\right| < C < \infty$ because of the bounded moments of Z. Therefore

$$\begin{aligned}
\left|\lim_{N\to\infty}\frac{1}{N^2T^2}\sum_{i,t,j,s}\mathbf{1}\left(i\neq j\right)\mathbb{E}\left[\tilde{Z}_{it}\tilde{\varepsilon}_{it}\tilde{Z}_{js}\tilde{\varepsilon}_{js}\right]\right| &< \lim_{N\to\infty} C\frac{1}{N^2T^2}\sum_{i,t,j,s}\mathbf{1}\left(i\neq j\right)\left|\mathbb{E}\left[\tilde{\varepsilon}_{it}\tilde{\varepsilon}_{js}\right]\right| \\
&= \lim_{N\to\infty} C\frac{1}{T^2}\sum_{s,t}\left|\mathbb{E}\left[\tilde{\varepsilon}_{it}\tilde{\varepsilon}_{js}\right]\right| \\
&= C\frac{1}{T^2}\sum_{s,t}\lim_{N\to\infty}\left|\mathbb{E}\left[\tilde{\varepsilon}_{it}\tilde{\varepsilon}_{js}\right]\right| \\
&= C\frac{1}{T^2}\sum_{s,t}\lim_{N\to\infty}\Big|\mathbb{E}\Big[(\varepsilon_{it}-\bar{\varepsilon}_i-\bar{\varepsilon}_t+\bar{\varepsilon}) \\
&\qquad\qquad (\varepsilon_{jt}-\bar{\varepsilon}_j-\bar{\varepsilon}_s+\bar{\varepsilon})\Big]\Big| \\
&= C\frac{1}{T^2}\sum_{s,t}\left|\mathbb{E}\left[(\varepsilon_{it}-\bar{\varepsilon}_i)(\varepsilon_{jt}-\bar{\varepsilon}_j)\right]\right| = 0
\end{aligned} \tag{52}$$

where the second line uses the exchangeability of units (i,j), the third line exchanges the limit and the (finite) sum, the fourth line writes out the definition of $\tilde{\varepsilon}_{it}$, the fifth line uses the fact that means across cross-sectional units go to zero, and the last equality uses the fact that ε are drawn independently across cross-sectional units.

We therefore have

$$\lim_{N\to\infty}\frac{1}{N}\left(\Omega^{CR}-\Omega\right) = \lim_{N\to\infty}-\frac{1}{N^2T^2}\sum_{i,t}\sum_{j,s}\mathbf{1}\left(i\neq j\right)\mathbb{E}\left[\tilde{Z}_{it}\tilde{\lambda}_i'\tilde{F}_t\tilde{Z}_{js}\tilde{\lambda}_j'\tilde{F}_s\right] \tag{53}$$

We now apply Lemma 2 to observe that, if $t\neq s$, then $\mathbb{E}\left[\tilde{Z}_{it}\tilde{\lambda}_i'\tilde{F}_t\tilde{Z}_{js}\tilde{\lambda}_j'\tilde{F}_s\right] = O(1/T^2) <$

$\frac{M}{T^2} < \infty$, if $t \neq s$. We therefore write

$$
\begin{aligned}
\lim_{N\to\infty} \frac{1}{N}\left(\Omega^{CR} - \Omega\right) &= \lim_{N\to\infty} -\frac{1}{N^2T^2}\sum_{i,t}\sum_{j,s}\Bigg(\mathbf{1}\left(i \neq j, t = s\right)\mathbb{E}\left[\tilde{Z}_{it}\tilde{\lambda}_i'\tilde{F}_t\tilde{Z}_{js}\tilde{\lambda}_j'\tilde{F}_s\right] \\
&\qquad + \mathbf{1}\left(i \neq j, t \neq s\right)O(1/T^2)\Bigg) \\
&= \lim_{N\to\infty} -\frac{1}{N^2T^2}\sum_{i,t}\sum_{j,s}\left(\mathbf{1}\left(i \neq j, t = s\right)\mathbb{E}\left[\tilde{Z}_{it}\tilde{\lambda}_i'\tilde{F}_t\tilde{Z}_{js}\tilde{\lambda}_j'\tilde{F}_s\right]\right) \\
&\quad - \lim_{N\to\infty}\frac{N\left(N-1\right)T\left(T-1\right)}{N^2T^2}\cdot O(1/T^2) \\
&= \lim_{N\to\infty} -\frac{1}{N^2T^2}\sum_{i,t}\sum_{j,s}\left(\mathbf{1}\left(i \neq j, t = s\right)\mathbb{E}\left[\tilde{Z}_{it}\tilde{\lambda}_i'\tilde{F}_t\tilde{Z}_{js}\tilde{\lambda}_j'\tilde{F}_s\right]\right) - O(1/T^2)
\end{aligned} \tag{54}
$$

We then substitute in $\tilde{Z}_{it} = \tilde{\eta}_i{}'\tilde{S}_t$ and simplify

$$
\begin{aligned}
\lim_{N\to\infty} \frac{1}{N}\left(\Omega^{CR} - \Omega\right) + O(1/T^2) &= \lim_{N\to\infty} -\frac{1}{N^2T^2}\sum_{i,t}\sum_{j}\mathbf{1}\left(i \neq j\right)\mathbb{E}\left[\tilde{\eta}_i'\tilde{S}_t\tilde{\lambda}_i'\tilde{F}_t\tilde{\eta}_j'\tilde{S}_t\tilde{\lambda}_j'\tilde{F}_t\right] \\
&= \lim_{N\to\infty} -\frac{1}{N^2T^2}\sum_{i,t}\sum_{j}\mathbf{1}\left(i \neq j\right)\mathbb{E}\left[\mathbb{E}\left[\tilde{\eta}_i'\tilde{S}_t\tilde{\lambda}_i'\tilde{F}_t\eta_j'\tilde{S}_t\tilde{\lambda}_j'\tilde{F}_t \mid \tilde{F}_t, \tilde{S}_t\right]\right] \\
&= \lim_{N\to\infty} -\frac{1}{N^2T^2}\sum_{i,t}\sum_{j}\mathbf{1}\left(i \neq j\right)\mathbb{E}\left[\tilde{S}_t'\mathbb{E}\left[\tilde{\eta}_i\tilde{\lambda}_i'\right]\tilde{F}_t\tilde{S}_t'\mathbb{E}\left[\tilde{\eta}_j\tilde{\lambda}_j'\right]\tilde{F}_t\right]
\end{aligned} \tag{55}
$$

where, in the second line, we condition on $\tilde{F}_t, \tilde{S}_t$ and, in the third line, we exploit the independence across regions. We finally simplify this expression further using the fact that regions are exchangeable, that $\tilde{F}_t, \tilde{S}_t$ are i.i.d. across time periods, and that $\tilde{\eta}$ and $\tilde{\lambda}$ converge to η and λ as $N \to \infty$:

$$
\begin{aligned}
\lim_{N\to\infty} \frac{1}{N}\left(\Omega^{CR} - \Omega\right) &= \lim_{N\to\infty} -\frac{N(N-1)}{N^2T^2}\sum_{t}\mathbb{E}\left[\tilde{S}_t'\mathbb{E}\left[\tilde{\eta}_i\tilde{\lambda}_i'\right]\tilde{F}_t\tilde{S}_t'\mathbb{E}\left[\tilde{\eta}_j\tilde{\lambda}_j'\right]\tilde{F}_t\right] - O(1/T^2) \\
&= -\frac{1}{T^2}\sum_{t}\mathbb{E}\left[\tilde{S}_t'\mathbb{E}\left[\eta_i\lambda_i'\right]\tilde{F}_t\tilde{S}_t'\mathbb{E}\left[\eta_j\lambda_j'\right]\tilde{F}_t\right] - O(1/T^2) \\
&= -\frac{1}{T}\mathbb{E}\left[(\tilde{S}_t'\mathbb{E}\left[\eta_i\lambda_i'\right]\tilde{F}_t)^2\right] - O(1/T^2)
\end{aligned} \tag{56}
$$

as desired.

Under the scalar case, this simplifies to:

$$
\lim_{N\to\infty} \frac{1}{N}\left(\Omega^{CR} - \Omega\right) = -\frac{1}{T}\left(\mathbb{E}[\tilde{\eta}_i\tilde{\lambda}_i]\right)^2\ \mathbb{E}\left[\left(\tilde{S}_t\tilde{F}_t\right)^2\right] - O(1/T^2) \tag{57}
$$

$\square$

A.5 Proof of Proposition 4

Proof. First, note that

$$\Omega := AVAR\left(\sqrt{N} \cdot \hat{\beta}\right) = \lim_{N\to\infty, T\to\infty, N/T\to C} \frac{1}{NT^2} \sum_{i,t} \sum_{j,s} \mathbb{E}\left[\tilde{u}_{it}\tilde{u}_{js}\tilde{Z}_{it}\tilde{Z}_{js}\right] \tag{58}$$

Let B denote the above infinite sum, subsetting to the terms for which $i \neq j$ and $t \neq s$. It suffices to show that $B = 0$. We have:

$$B = \lim_{N\to\infty, T\to\infty, N/T\to C} \frac{1}{NT^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{t=1}^{T} \sum_{s=1}^{T} \mathbf{1}\left(i \neq j \text{ AND } t \neq s\right) \mathbb{E}[\tilde{u}_{it}\tilde{u}_{js}\tilde{Z}_{it}\tilde{Z}_{js}] \tag{59}$$

Using our definitions of u and Z, we re-write the terms in the expectation as

$$\tilde{u}_{it}\tilde{u}_{js}\tilde{Z}_{it}\tilde{Z}_{js} = (\tilde{\lambda}_i'\tilde{F}_t + \tilde{\varepsilon}_{it})(\tilde{\lambda}_j'\tilde{F}_s + \tilde{\varepsilon}_{js})(\tilde{\eta}_i'\tilde{S}_t)(\tilde{\eta}_j'\tilde{S}_s) \tag{60}$$

We first observe that $\mathbb{E}[\tilde{\varepsilon}_{it}\tilde{\lambda}_j'\tilde{F}_s\tilde{Z}_{it}\tilde{Z}_{js}] = \mathbb{E}[\tilde{\varepsilon}_{it}]\mathbb{E}[\tilde{\lambda}_j'\tilde{F}_s\tilde{Z}_{it}\tilde{Z}_{js}] = 0$ because ε_{it} is i.i.d. across regions, mean zero, and independent from the factor draws and shocks. Similarly, $\mathbb{E}[\tilde{\lambda}_i'\tilde{F}_t\tilde{\varepsilon}_{js}\tilde{Z}_{it}\tilde{Z}_{js}] = 0$.

We now study the sum of the terms $\mathbb{E}[\tilde{\varepsilon}_{it}\tilde{\varepsilon}_{js}\tilde{Z}_{it}\tilde{Z}_{js}]$, or

$$B_1 := \lim_{N\to\infty, T\to\infty, N/T\to C} \frac{1}{NT^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{t=1}^{T} \sum_{s=1}^{T} \mathbf{1}\left(i \neq j \text{ AND } t \neq s\right) \mathbb{E}[\tilde{\varepsilon}_{it}\tilde{\varepsilon}_{js}\tilde{Z}_{it}\tilde{Z}_{js}] \tag{61}$$

We now show that $B_1 = 0$. Our calculation is similar to the equivalent calculation in the proof of Proposition 3. We first observe that $\mathbb{E}\left[\tilde{Z}_{it}\tilde{\varepsilon}_{it}\tilde{Z}_{js}\tilde{\varepsilon}_{js}\right] = \mathbb{E}\left[\tilde{Z}_{it}\tilde{Z}_{js}\right]\mathbb{E}\left[\tilde{\varepsilon}_{it}\tilde{\varepsilon}_{js}\right]$ because ε is independent of (Z, λ, F). We focus first on $\mathbb{E}\left[\tilde{\varepsilon}_{it}\tilde{\varepsilon}_{js}\right]$

$$
\begin{aligned}
\mathbb{E}\left[\tilde{\varepsilon}_{it}\tilde{\varepsilon}_{js}\right] &= \mathbb{E}\left[(\varepsilon_{it}-\bar{\varepsilon}_i-\bar{\varepsilon}_t+\bar{\varepsilon})(\varepsilon_{js}-\bar{\varepsilon}_j-\bar{\varepsilon}_s+\bar{\varepsilon})\right] \\
&= \mathbb{E}\Bigg[\frac{4}{NT}\varepsilon_{is}^2-\frac{2}{N}\varepsilon_{it}\varepsilon_{is}+\frac{1}{NT}\sum_{s=1}^{T}\varepsilon_{it}\varepsilon_{is}-\frac{2}{NT^2}\sum_{s=1}^{T}\sum_{r=1}^{T}\varepsilon_{ir}\varepsilon_{is}+ \\
&\qquad \frac{1}{N^2}\sum_{k=1}^{N}\varepsilon_{ks}\varepsilon_{kt}-\frac{2}{N^2T}\sum_{k=1}^{N}\sum_{r=1}^{T}\varepsilon_{kt}\varepsilon_{kr}+\frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{s=1}^{T}\sum_{r=1}^{T}\varepsilon_{ir}\varepsilon_{is}\Bigg] \\
&= \frac{4}{NT}\mathbb{E}\left[\varepsilon_{is}^2\right]-\frac{2}{N}\mathbb{E}\left[\varepsilon_{it}\varepsilon_{is}\right]+\frac{1}{NT}\sum_{s=1}^{T}\mathbb{E}\left[\varepsilon_{it}\varepsilon_{is}\right]-\frac{2}{NT^2}\sum_{s=1}^{T}\sum_{r=1}^{T}\mathbb{E}\left[\varepsilon_{ir}\varepsilon_{is}\right]+ \\
&\qquad \frac{1}{N^2}\sum_{k=1}^{N}\mathbb{E}\left[\varepsilon_{ks}\varepsilon_{kt}\right]-\frac{2}{N^2T}\sum_{k=1}^{N}\sum_{r=1}^{T}\mathbb{E}\left[\varepsilon_{kt}\varepsilon_{kr}\right]+\frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{s=1}^{T}\sum_{r=1}^{T}\mathbb{E}\left[\varepsilon_{ir}\varepsilon_{is}\right] \\
&= O\left(\frac{1}{N}\right)
\end{aligned} \tag{62}
$$

where we expand terms in the second line, simplify in the third, and use the boundedness of moments in the fourth. We next consider $\mathbb{E}\left[\tilde{Z}_{it}\tilde{Z}_{js}\right]$. We have:

$$
\begin{aligned}
\mathbb{E}\left[\tilde{Z}_{it}\tilde{Z}_{js}\right] &= \mathbb{E}\left[\tilde{\eta}_i'\tilde{S}_t\tilde{\eta}_j'\tilde{S}_s\right] \\
&= \mathbb{E}\left[\tilde{S}_t'\tilde{\eta}_i\tilde{\eta}_j'\tilde{S}_s\right] \\
&= \operatorname{tr}\left(\mathbb{E}\left[\tilde{\eta}_i\tilde{\eta}_j'\tilde{S}_s\tilde{S}_t'\right]\right) \\
&= \operatorname{tr}\left(\mathbb{E}\left[\tilde{\eta}_i\tilde{\eta}_j'\right]\mathbb{E}\left[\tilde{S}_s\tilde{S}_t'\right]\right)
\end{aligned} \tag{63}
$$

We proceed by analyzing cases. In case one, we have independent draws of η_i across regions, which yields.

$$
\begin{aligned}
\mathbb{E}\left[\tilde{\eta}_i\tilde{\eta}_j'\right] &= \mathbb{E}\left[\left(\eta_i-\bar{\eta}\right)\left(\eta_j-\bar{\eta}\right)'\right] \\
&= \mathbb{E}\left[-\eta_i\bar{\eta}'-\bar{\eta}\eta_j'+\bar{\eta}\bar{\eta}'\right] \\
&= \mathbb{E}\left[-\frac{1}{N}\eta_i\eta_i'-\frac{1}{N}\eta_j\eta_j'+\frac{1}{N^2}\sum_{\iota}\eta_\iota\eta_\iota'\right] \\
&= \mathbb{E}\left[-\frac{1}{N}\eta_i\eta_i'\right] \\
&= O\left(\frac{1}{N}\right) \\
\implies \mathbb{E}\left[\tilde{Z}_{it}\tilde{Z}_{js}\right] &= O\left(\frac{1}{N}\right)
\end{aligned} \tag{64}
$$

where the fourth line uses the fact that η_i and η_j are drawn from the same distribution. Case two is analogous, and yields $\mathbb{E}\left[\tilde{Z}_{it}\tilde{Z}_{js}\right] = O\left(\frac{1}{T}\right)$. We thus have that $\mathbb{E}\left[\tilde{Z}_{it}\tilde{\varepsilon}_{it}\tilde{Z}_{js}\tilde{\varepsilon}_{js}\right]$ is $O\left(\frac{1}{N^2}\right)$ in case one, and $O\left(\frac{1}{NT}\right)$ in case two. In either case, we can write that $\mathbb{E}\left[\tilde{Z}_{it}\tilde{\varepsilon}_{it}\tilde{Z}_{js}\tilde{\varepsilon}_{js}\right] = O\left(\frac{1}{N^2} + \frac{1}{NT}\right)$.

We now have:

$$\begin{aligned}
B_1 &= \lim_{N\to\infty, T\to\infty, N/T\to C}\left(\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbf{1}\left(i \neq j \text{ AND } t \neq s\right) O\left(\frac{1}{N^2} + \frac{1}{NT}\right)\right) \\
&= \lim_{N\to\infty, T\to\infty, N/T\to C}\left(\frac{1}{NT^2}N\left(N-1\right)T\left(T-1\right)O\left(\frac{1}{N^2} + \frac{1}{NT}\right)\right) \\
&= \lim_{N\to\infty, T\to\infty, N/T\to C} O\left(\frac{1}{N} + \frac{1}{T}\right) = 0
\end{aligned} \tag{65}$$

Therefore, we can drop the ε_{it} terms, and re-write Equation 59 as

$$B = \lim_{N\to\infty, T\to\infty, N/T\to C}\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbf{1}\left(i \neq j \text{ AND } t \neq s\right)\cdot\tilde{\omega}(i,j,t,s) \tag{66}$$

We first observe, using Lemma 2, that $\tilde{\omega}(i,j,t,s) = O(1/N^2)$ in case one and $\tilde{\omega}(i,j,t,s) = O(1/T^2)$ in case two. We observe that B can therefore be written under either case as

$$\begin{aligned}
B &= \lim_{N\to\infty, T\to\infty, N/T\to C}\frac{1}{NT^2}N\left(N-1\right)T\left(T-1\right)O\left(\frac{1}{N^2} + \frac{1}{T^2}\right) \\
&= \lim_{N\to\infty, T\to\infty, N/T\to C} N\cdot O\left(\frac{1}{N^2} + \frac{1}{T^2}\right) \\
&= 0
\end{aligned} \tag{67}$$

where the last line uses the fact that $\frac{N}{T} \to C$. Thus, since $B = 0$, it follows that $\Omega = \Omega^{TWC}$. □

A.6 Formulae for Weak-Instrument Robust Tests

In this Appendix, we provide expressions for the two weak-instrument robust test statistics that we study in Section 4. These are the "Anderson-Rubin Minimum Distance" (AR-MD) and "Anderson-Rubin Lagrange Multiplier" (AR-LM) tests introduced in Finlay and Magnusson (2009) and Magnusson (2010) and implemented in the weakiv command of Stata.

To introduce these statistics, we first remind the reader of our two model equations (in double-demeaned form):

$$\tilde{Y}_{it} = \tilde{X}_{it}'\beta + \tilde{u}_{it} \tag{68}$$
$$\tilde{X}_{it} = \tilde{Z}_{it}'\pi + \tilde{e}_{it} \tag{69}$$

In contrast to our baseline model, we let $\tilde{Z}_{it}$ by an $n_I \times 1$ vector and π be a $n_I \times 1$ vector. We also introduce the "reduced form" model that directly relates $\tilde{Y}$ and $\tilde{Z}$:

$$\tilde{Y}_{it} = \delta \cdot \tilde{Z}_{it} + \tilde{v}_{it} \tag{70}$$

We observe that the OLS estimate of this equation is $\hat{\delta} = (\tilde{Z}'\tilde{Z})^{-1}\tilde{Z}'Y$. In the scalar instrument case, this is related to the OLS estimates of the original model equations by $\hat{\delta} = \hat{\pi}\hat{\beta}$.

We next define some objects from Finlay and Magnusson (2009). The first is the asymptotic covariance matrix of $\sqrt{N}\left[(\hat{\delta}-\delta)', (\hat{\pi}-\pi)'\right]$, where N is the appropriate asymptotic normalization. We write an estimator of this object as

$$\hat{\Lambda}(\delta,\pi) = \begin{bmatrix} \hat{\Lambda}_{\delta,\delta}(\delta,\pi) & \hat{\Lambda}_{\delta,\pi}(\delta,\pi) \\ \hat{\Lambda}_{\pi,\delta}(\delta,\pi) & \hat{\Lambda}_{\pi,\pi}(\delta,\pi) \end{bmatrix} \tag{71}$$

which, in a way that we will soon clarify, depends on a parameters δ, π. We next define the object $\Psi(\beta_0, \beta_1)$ which is the asymptotic coviarance matrix of $\sqrt{N}(\hat{\delta} - \hat{\pi}\beta)$:

$$\hat{\Psi}(\beta,\delta,\pi) = \hat{\Lambda}_{\delta\delta}(\delta,\pi) - \beta\hat{\Lambda}_{\delta,\pi}(\delta,\pi) - \beta\hat{\Lambda}_{\pi,\delta}(\delta,\pi) + \beta^2\hat{\Lambda}_{\pi,\pi}(\delta,\pi) \tag{72}$$

We finally define the estimators $\hat{\Lambda}$. $\hat{\Lambda}_{\delta\delta}$ and $\hat{\Lambda}_{\pi\pi}$ are the corresponding cluster-robust variance estimators for the regressions of $\tilde{Z}_{it}$ on $\tilde{Y}_{it}$ and $\tilde{Z}_{it}$ on $\tilde{X}_{it}$. In the one-way clustering case, these are

$$\begin{aligned} \hat{\Lambda}_{\delta,\delta}(\delta,\pi) &= (\tilde{Z}'\tilde{Z})^{-1}\left(\sum_c \tilde{Z}_c'(\tilde{Y}_c - \delta\tilde{Z}_c)(\tilde{Y} - \delta\tilde{Z}_c)'\tilde{Z}_c\right)(\tilde{Z}'\tilde{Z})^{-1} \\ \hat{\Lambda}_{\pi,\pi}(\delta,\pi) &= (\tilde{Z}'\tilde{Z})^{-1}\left(\sum_c \tilde{Z}_c'(\tilde{X}_c - \pi\tilde{Z}_c)(\tilde{X} - \pi\tilde{Z}_c)'\tilde{Z}_c\right)(\tilde{Z}'\tilde{Z})^{-1} \end{aligned} \tag{73}$$

where each object in the summation is the vector of the corresponding variable in cluster c. These objects are defined analogously for multi-way clustering, The covariance terms are defined as

$$\begin{aligned}\hat{\Lambda}_{\delta,\pi}(\delta,\pi) &= (\tilde{Z}'\tilde{Z})^{-1}\left(\sum_c \tilde{Z}_c'(\tilde{Y}_c - \delta\tilde{Z}_c)(\tilde{X} - \pi\tilde{Z}_c)'\tilde{Z}_c\right)(\tilde{Z}'\tilde{Z})^{-1} \\ \hat{\Lambda}_{\pi,\delta}(\delta,\pi) &= (\tilde{Z}'\tilde{Z})^{-1}\left(\sum_c \tilde{Z}_c'(\tilde{X}_c - \pi\tilde{Z}_c)(\tilde{Y} - \delta\tilde{Z}_c)'\tilde{Z}_c\right)(\tilde{Z}'\tilde{Z})^{-1}\end{aligned} \tag{74}$$

We now define the estimators. The AR-MD statistic is

$$\mathrm{AR}_{\mathrm{MD}} = N(\hat{\delta} - \hat{\pi}\beta_0)'(\hat{\Psi}(\beta_0, \hat{\delta}, \hat{\pi}))^{-1}(\hat{\delta} - \hat{\pi}\beta_0) \tag{75}$$

and observe from Finlay and Magnusson (2009) and Magnusson (2010) that this follows a $\chi^2(n_I)$ distribution. The AR-LM statistic, by contrast, replaces the middle covariance with the same evaluated at $\delta = \hat{\pi}\beta_0$:

$$\mathrm{AR}_{\mathrm{LM}} = N(\hat{\delta} - \hat{\pi}\beta_0)'(\hat{\Psi}(\beta_0, \hat{\pi}\beta_0, \hat{\pi}))^{-1}(\hat{\delta} - \hat{\pi}\beta_0) \tag{76}$$

Equivalence with se_{β_0}in One-Dimensional Case. We now show that the AR-LM test coincides with the conventional test with the t-statistic

$$t_{\beta_0} = \frac{\hat{\beta} - \beta_0}{\mathrm{se}_{\beta_0}} \tag{77}$$

in one dimension. To do this, we will show that

$$\hat{\Psi}(\beta_0, \hat{\pi}\beta_0, \hat{\pi}) = N\hat{\pi}^2(\mathrm{se}_{\beta_0})^2 \tag{78}$$

The equivalence of the tests follows from observing that if $t \sim N(0,1)$, then $t^2 \sim \chi^2(1)$. We show (78) in the case of single-variable clustering; the calculation is analogous for multi-way clustering (or multi-way HAC), by expressing those covariance estimators as the weighted sum of one-dimensional clustering estimators.

We first simplify

$$\hat{\Psi}(\beta,\delta,\pi) = (\tilde{Z}'\tilde{Z})^{-1}\left(\sum_c \tilde{Z}_c' W_c \tilde{Z}_c\right)(\tilde{Z}'\tilde{Z})^{-1} \tag{79}$$

where

$$\begin{aligned}W_c = &(\tilde{Y}_c - \delta\tilde{Z}_c)(\tilde{Y} - \delta\tilde{Z}_c)' \\ &- \beta(\tilde{Y}_c - \delta\tilde{Z}_c)(\tilde{X} - \pi\tilde{Z}_c)' - \beta(\tilde{X}_c - \pi\tilde{Z}_c)(\tilde{Y} - \delta\tilde{Z}_c)' \\ &+ \beta^2(\tilde{X}_c - \pi\tilde{Z}_c)(\tilde{X} - \pi\tilde{Z}_c)'\end{aligned} \tag{80}$$

Next, we complete the square to write $W_c = V_c' V_c$ where

$$V_c = \tilde{Y}_c - \delta \tilde{Z}_c - \beta(\tilde{X}_c - \pi \tilde{Z}_c) = \tilde{Y}_c - \beta \tilde{X}_c + \beta\pi \tilde{Z}_c - \delta \tilde{Z}_c \tag{81}$$

Evaluating this at $\beta = \beta_0$, $\pi = \hat{\pi}$, and $\delta = \hat{\pi}\beta_0$ we obtain $V_c = \tilde{Y}_c - \beta_0 \tilde{X}_c$. Therefore, we have

$$\hat{\Psi}(\beta, \delta, \pi) = (\tilde{Z}'\tilde{Z})^{-1} \left(\sum_c \tilde{Z}_c'(\tilde{Y}_c - \beta_0 \tilde{X}_c)(\tilde{Y}_c - \beta_0 \tilde{X}_c)' \tilde{Z}_c \right) (\tilde{Z}'\tilde{Z})^{-1} \tag{82}$$

We next use the fact that $\hat{\pi} = (\tilde{Z}'\tilde{Z})^{-1}\tilde{Z}'\tilde{X}$ and the definition

$$\mathrm{se}_{\beta_0} = \sqrt{\frac{1}{N}(\tilde{Z}'\tilde{X})^{-2} \left(\sum_c \tilde{Z}_c'(\tilde{Y}_c - \beta_0 \tilde{X}_c)(\tilde{Y}_c - \beta_0 \tilde{X}_c)' \tilde{Z}_c \right)} \tag{83}$$

to write

$$\begin{aligned} \hat{\Psi}(\beta, \delta, \pi) &= \hat{\pi}^2 (\tilde{Z}'\tilde{X})^{-2} \left(\sum_c \tilde{Z}_c'(\tilde{Y}_c - \beta_0 \tilde{X}_c)(\tilde{Y}_c - \beta_0 \tilde{X}_c)' \tilde{Z}_c \right) \\ &= N\hat{\pi}^2 (\mathrm{se}_{\beta_0})^2 \end{aligned} \tag{84}$$

as desired.

B Additional Tables and Figures

Table 5: Placebo Test with Alternative Data-Generating Processes

	Panel A: Baseline		Panel B: IID		Panel C: Persistent		Panel D: Mixture		Panel E: AR(2)	
	$se_{\hat{\beta}}$	se_{β_0}	$se_{\hat{\beta}}$	se_{β_0}	$se_{\hat{\beta}}$	se_{β_0}	$se_{\hat{\beta}}$	se_{β_0}	$se_{\hat{\beta}}$	se_{β_0}
Cluster by State	25.4%	19.8%	21.6%	16.0%	27.0%	19.3%	21.8%	12.8%	21.8%	19.0%
Cluster by Year	24.4%	20.8%	5.8%	3.6%	28.2%	29.2%	24.6%	22.0%	22.8%	19.6%
Two-way Cluster	21.1%	9.0%	11.9%	3.8%	22.2%	11.5%	16.9%	7.0%	18.2%	8.6%
Two-way HAC	20.3%	3.0%	15.6%	2.7%	19.8%	3.5%	16.2%	2.0%	16.8%	2.2%

Notes: This table shows the frequency at which the null hypothesis of $\beta_0 = 0$ is rejected at the 5% level in several variants of the placebo test based on Nakamura and Steinsson (2014). Panel A corresponds to the baseline simulation. In Panel B, we simulate national defense spending shocks as an IID Gaussian variable with the same variance as observed shocks. In Panel C, we simulate national defense spending shocks as an Gaussian AR(1) with coefficient 0.9, holding fixed the variance of the variable. In Panel D, we simulate national defense spending shocks as an AR(1) with an empirically estimated coefficient $\rho = \hat{\rho} = 0.66$ and innovations whose distribution is a mixture of two Gaussian distributions. In Panel E, we simulate national defense spending shocks as an AR(2) process with Gaussian innovations. Since the placebo defense spending shocks are drawn at random for each placebo draw, a correctly calibrated 5% test would reject 5% of the time.

Table 6: 90% Confidence Intervals for Conventional IV Estimate of Regional Fiscal Multiplier in Nakamura and Steinsson (2014)

Panel A: Initial Share (Point Estimate: 2.477)

	Conventional		Weak-IV Robust	
	$\text{se}_{\hat{\beta}}$	se_{β_0}	AR-MD	AR-LM
Cluster by State	$(0.887, 4.066)$	$(1.171, \infty)$	$(1.044, 4.433)$	$(1.171, \infty)$
Two-way Cluster	$(0.709, 4.245)$	$(1.095, \infty)$	$(0.982, 4.794)$	$(1.095, \infty)$
Two-way HAC ($L = 3$)	$(0.436, 4.518)$	$(0.473, \infty)$	$(0.767, 5.146)$	$(0.473, \infty)$
Randomization Inference	$(0.46, 4.72)$			

Panel B: State FE (Point Estimate: 1.426)

	Conventional		Weak-IV Robust	
	$\text{se}_{\hat{\beta}}$	se_{β_0}	AR-MD	AR-LM
Cluster by State	$(0.820, 2.032)$	$(0.88, 2.36)$	Empty	$(-\infty, \infty)$
Two-way Cluster	$(0.502, 2.351)$	$(0.63, 5.15)$	$(-\infty, 0.985) \cup (1.467, \infty)$	$(-\infty, \infty)^*$
Two-way HAC ($L = 3$)	$(0.256, 2.597)$	$(0.18, \infty)$	$(-\infty, 0.014) \cup (0.361, 1.562) \cup (1.731, \infty)$	$(-\infty, \infty)^*$
Randomization Inference	$(-3.3, 7.1)$			

Notes: This table shows 90% confidence intervals for the regional fiscal multiplier, estimated in the setting of Nakamura and Steinsson (2014) using the IV estimator. Panel A shows results based on the instrumental variable strategy that interacts defense spending growth with the pre-period share of military procurement spending in state output. Panel B shows results based on the instrumental variable strategy that interacts defense spending growth with state fixed effects. The first four rows of each panel show results from tests that implement clustering by state, clustering by year, using two-way clustering (state and year), and using two-way HAC standard errors with a kernel bandwidth of three years. In each of these rows, we report results from conventional t-tests with standard error estimates $\text{se}_{\hat{\beta}}$ and se_{β_0} and weak-instrument-robust tests using the Anderson-Rubin Minimum Distance and the Anderson-Rubin Lagrange Multiplier statistics of Finlay and Magnusson (2009) and Magnusson (2010) (see Section 4.2 for details). In Panel A, the se_{β_0} and AR-LM tests exactly coincide. The fifth row of each panel reports results from randomization inference. *: All of the weak-instrument-robust confidence intervals in Column 1, marked with this symbol, contain "holes" (i.e., intervals of width close to 0.01 in which one can reject the null hypothesis).

Table 7: 68% Confidence Intervals for Conventional IV Estimate of Regional Fiscal Multiplier in Nakamura and Steinsson (2014)

Panel A: Initial Share (Point Estimate: 2.477)

	Conventional		Weak-IV Robust	
	$\mathrm{se}_{\hat{\beta}}$	se_{β_0}	AR-MD	AR-LM
Cluster by State	$(1.516, 3.438)$	$(1.664, 3.860)$	$(1.587, 3.544)$	$(1.664, 3.860)$
Two-way Cluster	$(1.408, 3.546)$	$(1.624, 4.443)$	$(1.522, 3.715)$	$(1.624, 4.443)$
Two-way HAC ($L = 3$)	$(1.243, 3.711)$	$(1.462, \infty)$	$(1.384, 3.906)$	$(1.462, \infty)$
Randomization Inference	$(1.34, 3.70)$			

Panel B: State FE (Point Estimate: 1.426)

	Conventional		Weak-IV Robust	
	$\mathrm{se}_{\hat{\beta}}$	se_{β_0}	AR-MD	AR-LM
Cluster by State	$(1.060, 1.793)$	$(1.10, 1.87)$	Empty	$(-\infty, \infty)$
Two-way Cluster	$(0.867, 1.985)$	$(0.96, 2.31)$	$(-\infty, 0.989) \cup$ $(1.514, \infty)$	$(-\infty, \infty)^*$
Two-way HAC ($L = 3$)	$(0.719, 2.134)$	$(0.84.2.82)$	$(-\infty, 0.016) \cup$ $(0.380, 1.565) \cup$ $(1.741, \infty)$	$(-\infty, \infty)^*$
Randomization Inference	$(-1.3, 5.1)$			

Notes: This table shows 68% confidence intervals for the regional fiscal multiplier, estimated in the setting of Nakamura and Steinsson (2014) using the IV estimator. Panel A shows results based on the instrumental variable strategy that interacts defense spending growth with the pre-period share of military procurement spending in state output. Panel B shows results based on the instrumental variable strategy that interacts defense spending growth with state fixed effects. The first four rows of each panel show results from tests that implement clustering by state, clustering by year, using two-way clustering (state and year), and using two-way HAC standard errors with a kernel bandwidth of three years. In each of these rows, we report results from conventional t-tests with standard error estimates $\mathrm{se}_{\hat{\beta}}$ and se_{β_0} and weak-instrument-robust tests using the Anderson-Rubin Minimum Distance and the Anderson-Rubin Lagrange Multiplier statistics of Finlay and Magnusson (2009) and Magnusson (2010) (see Section 4.2 for details). In Panel A, the se_{β_0} and AR-LM tests exactly coincide. The fifth row of each panel reports results from randomization inference. *: All of the weak-instrument-robust confidence intervals in Column 1, marked with this symbol, contain "holes" (i.e., intervals of width close to 0.01 in which one can reject the null hypothesis).