Essays in Macroeconometrics: Heterogeneity, Spillovers, and Dynamic Causal Effects

Eric Yibin Qian

A Dissertation
Presented to the Faculty
of Princeton University
in Candidacy for the Degree
of Doctor of Philosophy

Recommended for Acceptance by
the Department of
Economics
Advisers: Mikkel Plagborg-Møller and Mark Watson

May 2026

Abstract

This dissertation consists of three chapters on macroeconometrics.

In the first chapter, I develop an optimal test for omitted unit-level heterogeneity applicable to moment condition models. The test is particularly useful for structural estimation, which inevitably involves a choice of which parameters to treat as homogeneous across units. Unlike existing semiparametric specification tests for heterogeneity, the test asymptotically maximizes a weighted average power criterion. I study two applications. First, applied to income dynamics, I use the heterogeneity test as a diagnostic for determining the presence of within education group-level heterogeneity. Second, I use the test as a diagnostic for the presence of within subindustry-level heterogeneity in the estimation of production functions.

In the second chapter, I propose a new estimator, called robust granular instrumental variables (RGIV), that enables the study of unit-level heterogeneity in spillovers. The estimator builds on granular instrumental variables (GIV), an estimator that has experienced sharp growth in empirical macro-finance—spanning economic environments like the estimation of spillovers and demand systems. Unlike existing methods that assume heterogeneity is a function of observables, RGIV leaves heterogeneity unrestricted. Applied to the Euro area, I find strong evidence of country-level heterogeneity in sovereign yield spillovers.

In the third chapter (coauthored with José Luis Montiel Olea, Mikkel Plagborg-Møller, and Christian Wolf), we consider impulse response inference in a locally misspecified vector autoregression (VAR) model. The conventional local projection (LP) confidence interval has correct coverage even when the misspecification is so large that it can be detected with probability approaching 1. By contrast, the conventional VAR confidence interval with short-to-moderate lag length can severely undercover for misspecification that is small, difficult to detect statistically, and cannot be ruled out based on economic theory. The VAR confidence interval has robust coverage if, and only if, the lag length is so large that the interval is as wide as the LP interval.

Acknowledgements

I am beyond thankful to the members of my dissertation committee for their generous support over the years. To Mikkel Plagborg-Møller, for challenging me to exceed even my own expectations. To Mark Watson, for helping me see the forest through the trees. To John Grigsby, for guiding me in becoming a clearer thinker. I am eager to pay their generosity forward. I am grateful to many others who have given me invaluable advice during graduate school: Ulrich Müller, Bo Honoré, Michal Kolesár, Mark Aguiar, Richard Rogerson, Karthik Sastry, Gianluca Violante, as well as the participants of the econometrics and macroeconomics student workshops. I'm especially grateful to Domenico Giannone for sparking my love for economics during my time as a Research Analyst at the New York Fed.

I have been immensely fortunate to collaborate with a talented group of coauthors over the years: Christian Wolf, José Luis Montiel Olea, Conor Walsh, Felipe Del Canto, Richard Crump, Stefano Eusepi, and Argia Sbordone. They have all gone out of their way to support me, and I have learned so much from them all.

Perhaps the real treasure was the friends I made along the way. Thank you to my classmates for the years of joy: Sebastián Guarda, Pier Paolo Creanza, Michael Jenuwine, Richard De Thorpe, Ali Zhang, Carolina Piazza, and Christine Blandhol. I'll especially miss my post-seminar conversations with Darwin Yang, Jacob Dorn, Luther Yap, Sebastian Roelsgaard, and Dexin Li.

I couldn't have done this without my family. I am forever grateful to my parents for their steadfast support from the very beginning. Thank you to Max for helping me keep the bigger picture in mind.

And finally, to Jada, for the years behind us and the many more to come.

To my parents

Contents

List of Tables

List of Figures

Chapter 1

Testing for Omitted Heterogeneity*

1.1 Introduction

The assumption of unit-level parameter homogeneity is common and necessary in econometric models with nonlinear moment conditions. Examples include homogeneity in the persistence of income processes across households (Guvenen, 2009), the dynamic response of investment to monetary shocks across firms (Ottonello and Winberry, 2020), and output elasticities of inputs for production functions within industries (Ackerberg et al., 2015). While these moment conditions are often theoretically motivated by a representative agent, researchers face the difficult task of accounting for unit-level heterogeneity in the data; incorrectly imposing parameter homogeneity across units can lead to uninterpretable or inconsistent estimates of the parameter of interest. In light of this issue, this paper's main contribution is a statistical test of parameter heterogeneity for moment condition models.

Concretely, one of this paper's applications considers the estimation of plant-level markups, which requires the estimation of the output elasticity of inputs. Here, the researcher is unwilling to fully specify the distribution of the random variables in the plant's production process (e.g. unwilling to impose Gaussian productivity shocks) but is willing to impose

*I am grateful for helpful comments from seminar participants at: Princeton University, Stanford Institute for Economic Policy Research, University of Maryland, University of Notre Dame, Johns Hopkins University, London School of Economics, and University College London.

moment conditions for a representative firm (Ackerberg et al., 2015). However, neglecting firm-level heterogeneity in production functions gives rise to estimates of output elasticities that, in the limit, fail to converge to an average value and otherwise lack a clear interpretation. While it is common to address parameter heterogeneity by estimating production functions at the subindustry level, the test allows the researcher to assess the presence of leftover parameter heterogeneity *within* these subsamples. Hence, the test can help determine the subindustries that give reliable markup estimates and those that are worthy of further analyses.

The proposed test takes the form of a test of over-identifying restrictions. Starting with moment conditions that are valid under parameter homogeneity, the test exploits Jensen's inequality; parameter heterogeneity is detectable in the moment conditions that are nonlinear in the parameter under test. The test is then constructed using a second-derivative weighted sum of the normalized fitted moment conditions and admits the interpretation of an estimate of the variance of the heterogeneous coefficient. Thus, the second derivative matrix is the only required object beyond those typically computed for the estimation of moment condition models. Unlike other moment-based specification tests, the proposed test directs power toward the specific alternative hypotheses that arise from parameter heterogeneity increasing power. Relative to the omnibus J test specifically, power is directed by placing higher weight on the moment conditions expected to be misspecified under parameter heterogeneity and using critical values that exploit the structure of the alternative hypothesis parameter space. Unlike likelihood-based specification tests, the procedure doesn't require for there to be a correct and fully-specified likelihood function for size to be controlled under the null hypothesis of no parameter heterogeneity.

The test can be viewed as an asymmetric score test on the limiting distribution of the fitted moment conditions. The test is based on a moment condition model with parameter heterogeneity that shrinks with the sample size. Under this framework, the amount of misspecification is large enough to affect the coverage of conventional GMM confidence intervals for the mean parameter yet small enough for the power of specification tests to be nontrivial. The test is based on studying the limiting distribution of the fitted moment conditions under this local parameter heterogeneity. A nonstandard feature of this problem is that the alternative hypothesis parameter space is asymmetric, containing the set of positive semidefinite matrices as opposed to non-zero matrices in general. The test achieves power gains by exploiting this asymmetry through a likelihood ratio test for the score vector of the limiting fitted moment conditions (Silvapulle and Silvapulle, 1995), giving a mixture chi-squared null distribution. In the special case of testing for scalar heterogeneity, the null distribution is Gaussian giving a one-sided test that rejects for negative values.

The proposed test is asymptotically optimal for detecting moderate amounts of heterogeneity. Adapting the semiparametric framework of van der Vaart (1998), I derive the efficient score, restricting study to the class of densities that satisfy a set of moment conditions absent parameter heterogeneity. The composite test for multivariate heterogeneity inevitably requires trading off power in detecting heterogeneity among the parameters under test (just as in the multivariate normal means problem of, for instance, Chapter 15.2 of van der Vaart (1998)). I show that the proposed test is optimal for detecting moderate amounts of parameter heterogeneity—moderate in the sense that the test maximizes a weighted average power criterion for distant alternatives in the limit experiment (Andrews, 1996) yet parameter heterogeneity is local. For the special case of testing for scalar heterogeneity, I show that the test is asymptotically uniformly most powerful. These results complement existing specification tests, which either require a fully-specified parametric model (necessarily imposing strong assumptions on the distribution of shocks), have no known optimality properties for detecting parameter heterogeneity (omnibus misspecification tests like the J test), or are optimal within narrow classes of tests (e.g. Hahn et al. (2014) is optimal among tests with limiting chi-squared distributions under the null hypothesis).

The test is most useful when economic theory gives restrictions on a generating model but is uninformative on the particular distribution of the random variables. As a result, the test's power comes from measuring disagreement in a set of moment conditions. In contrast, textbook parametric tests of the presence of mixtures, like the Neyman and Scott $C(\alpha)$ test, require moment conditions that are distinct from those used in estimation to also be satisfied under the null. These moments are derived from second-order derivatives of the log likelihood function (Lindsay, 1995). For example, the maximum likelihood estimator of the mean and variance of a normal random variable are their respective sample analogues. The corresponding $C(\alpha)$ test of parameter heterogeneity in both parameters however rejects when the third and fourth moments of the data are incompatible with normality. Therefore, a test with correct size would require the stronger assumption that the third and fourth moments are compatible with a Normal distribution under the null hypothesis of no parameter heterogeneity.

In a Monte Carlo exercise, I show that my proposed test has good finite sample properties in a short panel AR(2) model. Serving as a benchmark, the AR(2) is a relatively simple model where it is not known how to estimate the amount of heterogeneity of the autoregressive parameters absent strong distributional restrictions. By only assuming white noise errors and without restricting fixed effects, we can derive moment conditions that are valid under parameter homogeneity. When testing for heterogeneity in both autoregressive parameters, the empirical size is near its nominal size for Gaussian, Skew-t, and ARCH(1) shocks. Varying the degree of parameter heterogeneity, the test has higher power than the J test and the test of Hahn et al. (2014). The simulations demonstrate that a likelihood-based $C(\alpha)$ test for parameter heterogeneity exhibits a power-robustness tradeoff. While the power of the $C(\alpha)$ test is higher than the preceding moment-based tests when the shocks are indeed Gaussian, the $C(\alpha)$ test severely over-rejects otherwise (at worst, with an empirical rejection rate of 100% for a 5% test). Consistent with this paper's theoretical results, ignoring parameter heterogeneity causes the conventional confidence intervals to severely under-cover the mean of the autoregressive parameters.

The paper includes two empirical applications. In the first empirical application, I show that household-level heterogeneity in the persistence of income shocks is detectable even in public survey data. I estimate a model of income dynamics using the Panel Survey of Income Dynamics based on Guvenen (2007) that includes a transitory shock, persistent shock, and income profile heterogeneity. On the full sample, the J test fails to reject at 5% while the test for heterogeneity on the persistence of the persistent shock rejects. This result suggests that a researcher concerned about parameter heterogeneity may fail to detect if reliant solely on the J test. I then use the heterogeneity test as a data-driven guide for determining the presence of within education group-level parameter heterogeneity, testing the assumption of approximate homogeneity within each group. I find that the assumption of parameter homogeneity approximately holds for the "HS or less," "Some college," and "Bachelor's or more" education groups.

In the second empirical application, I use Chilean manufacturing data to show that plant-level heterogeneity in production functions estimated at the finest available level of industry aggregation threatens the estimates of markups over marginal costs. Markups are important to study because their level and dispersion distort allocations and ultimately affect welfare (Hsieh and Klenow, 2009; Edmond et al., 2023). Under the "production function approach" (De Loecker and Warzynski, 2012), plant-level production functions are typically estimated at the industry level. Relative to existing work, I entertain the possibility of parameter heterogeneity *within* 4-digit industry categories threatening estimates of markups. Combined with the J test, I then use the heterogeneity test as a diagnostic for determining the subindustries whose moments are compatible with the data. I find that 15 of the 50 total subindustries fail to reject the J test but reject the heterogeneity test at 5%. Notably, none of these industries reject the test of Hahn et al. (2014) at 5%. Next, I find 13 "well-specified" subindustries that fail to reject both the heterogeneity and J tests at 5%. Consistent with the pattern of development during the historical period under study, these subindustries mostly include those related to food and beverages. Focusing on plants belonging to these well-specified subindustries, I estimate a median sales-weighted plant-level markup of 1.24 with a 90 minus 50-percentile markup dispersion of 0.42. Investigating the implications of improper aggregation, I compute markups for plants belonging to these same well-specified subindustries, but instead plug-in production function estimates estimated using coarser 3-digit industry categories. Here, the median markup rises to 1.37 and markup dispersion falls to 0.32. Therefore ignoring parameter heterogeneity overstates the median markup and understates its dispersion.

LITERATURE. The proposed test is most useful in cases where modeling a random coefficient is difficult, like short dynamic panel models (Arellano and Bond, 1991; Arellano and Bover, 1995; Ahn and Schmidt, 1995; Blundell and Bond, 1998) and the estimation of production functions (Olley and Pakes, 1996; Blundell and Bond, 2000; Levinsohn and Petrin, 2003; Ackerberg et al., 2015; Gandhi et al., 2020). These short dynamic panel methods are often used in characterizing income dynamics (MaCurdy, 1982; Guvenen, 2007). The proposed test is less useful for cases where there are existing approaches for modeling and estimating a random coefficient, like short panel models with strictly exogenous regressors (Chamberlain, 1992; Wooldridge, 2005; Arellano and Bonhomme, 2012; Graham and Powell, 2012).

The test can be viewed as a diagnostic before proceeding to technically sophisticated methods for estimating parameter heterogeneity, especially in the context of dynamic panels. These techniques require additional assumptions and can be computationally intensive. Examples include restricting heterogeneity to a finite number of types or limiting heterogeneity to particular distributional families (Browning et al., 2010; Gu and Koenker, 2017; Alan et al., 2018; Bonhomme and Manresa, 2015; Bonhomme et al., 2019). Alternatively, Lee (2022) gives a methodology applied to dynamic panels for computing identified sets. Under further restrictions to the random coefficient, Pesaran and Yang (2024) gives conditions for estimating moments of the autoregressive coefficient of a panel AR(1). For richer models however, like the panel AR(2) considered in the simulation section, it is not known how to estimate the heterogeneity of the autoregressive parameters without restricting the distribution of shocks.

My proposed test builds on the work of statistical tests for the presence of mixture distributions in a likelihood-based framework (Neyman, 1959; Chesher, 1984; Lindsay, 1995;

Gu, 2016). Like these tests, my test is designed to detect local parameter heterogeneity, is based on a framework where parameter heterogeneity shrinks with the sample size, and requires the study of second-order expansions that lie outside of classical testing frameworks (like those described in Chapter 9 of Newey and McFadden (1994) in my case). Unlike these likelihood-based statistical tests, I focus on moment condition models. In particular, my framework for semiparametric efficiency is closely related to Gu (2016), which analyzes the $C(\alpha)$ test for parameter heterogeneity in a LeCam framework with a fully-specified parametric likelihood function. The moment condition model featured in this paper can be viewed as a semiparametric version where the distribution of the data is an infinite-dimensional nuisance parameter. To test for multivariate heterogeneity, Gu (2016) also proposes a likelihood ratio test on a score vector (following Silvapulle and Silvapulle (1995)). The arguments developed in this paper suggest that their test also maximizes a weighted average power criterion.

Just like this paper, Hahn et al. (2014) proposes a test for parameter heterogeneity in a moment-based framework. Their test also relies on a second-derivative weighted sum of normalized moment conditions. Their test however maximizes power relative to the narrow class of tests with limiting chi-square distributions. This paper's test in contrast achieves gains in power by directing power to alternative hypotheses that would arise from parameter heterogeneity, yielding semiparametric efficiency in a weighted average power sense.

This paper's test is a special case of testing problems where the parameter under test is on the boundary of the parameter space under the null hypothesis (Chernoff, 1954; Bartholomew, 1959; Perlman, 1969; Chant, 1974; Robertson et al., 1988; Andrews, 2001). The limit experiment of my test reduces to a test of an unknown mean in a multivariate Gaussian shift experiment with a known variance. Specifically, the null hypothesis is simple while the alternative hypothesis is a convex cone—included in the framework of Example 1 of Andrews (1996). Like these references, the limiting distribution of the test statistic is non-Gaussian.

Like prior work on local misspecification in moment condition models (Kitamura et al., 2013; Andrews et al., 2017; Armstrong and Kolesár, 2021; Bonhomme and Weidner, 2022)), parameter heterogeneity in this framework induces asymptotic bias of the same order as the standard errors. Unlike these papers, however, this paper focuses on constructing an optimal hypothesis test for the specific alternatives that arise from parameter heterogeneity (giving rise to gains in power) rather than studying model misspecification in general. The Taylor expansions used in this paper are similar to those featured in Evdokimov and Zeleneev (2023), who correct for scalar errors-in-variables bias for nonlinear moment condition models. This paper instead uses a multivariate expansion to motivate the construction of an optimal test of parameter heterogeneity.

Outline. Section 1.2 illustrates the main ideas of the proposed heterogeneity test in two motivating examples. Section 1.3 introduces the heterogeneity test. Section 1.4 presents a framework for semiparametric efficiency of the test. Through simulations, Section 1.5 examines the finite sample properties of the proposed test and compares with competing approaches. Section 1.6 applies the test to applications on income dynamics and markup dispersion. Section 1.7 concludes.

1.2 Motivating examples

Section 1.2.1 introduces a panel AR(2) to aid in discussing the considerations researchers face in estimating wide dynamic panel models. Section 1.2.2 illustrates the ingredients required for a test of parameter heterogeneity in a stylized measurement error model that will be used as a running example throughout this paper.

1.2.1 Example: Wide panel AR(2)

This subsection considers a relatively simple stationary wide panel AR(2). We will revisit the model in Section 1.5 in a Monte Carlo exercise.

Setting up, the researcher observes an outcome variable y_{it} for units $i = 1, \ldots, n$ and time $t = 1, \ldots, T$. The number of units is large while the number of time periods is small. The outcome variable is a function of its previous two lags

$$y_{it} = \delta_i + \phi_1 y_{it-1} + \phi_2 y_{it-2} + \varepsilon_{it} \tag{1.2.1}$$

for innovations $\varepsilon_{it} \sim (0, \sigma^2)$ (i.i.d. over units and white noise over time) and unit fixed effect δ_i. Notably, the autoregressive parameters ϕ_1 and ϕ_2 are taken to be homogeneous over units. This assumption will be the focus of study for this paper. In addition, the researcher is unwilling to impose further restrictions on the distribution of the innovations (like Gaussianity) or restrictions of higher-order moments. Consistent with the estimation of models of income dynamics (MaCurdy, 1982; Guvenen, 2007), the researcher estimates by matching the empirical autocovariances in first differences (enabling an unrestricted fixed effect) with those implied by Equation 1.2.1. The procedure is a form of nonlinear GMM since the moment conditions are nonlinear in the autoregressive parameters.

While simple, the model is useful to study because it shares attributes with richer models used in practice. Examples include those of income dynamics (Guvenen, 2009; MaCurdy, 1982), production function estimation (Blundell and Bond, 2000; Ackerberg et al., 2015) and panel local projections (Ottonello and Winberry, 2020). These models are typically motivated by moment conditions that are valid for a representative agent where the assumption of unit-level parameter homogeneity is required for estimation. In these settings, it is typically unknown how to rewrite the moment conditions to be valid if the homogeneous parameter were instead modeled as a random coefficient.

Analogously, the assumption of unit-level homogeneity in the autoregressive coefficients ϕ_1 and ϕ_2 is potentially problematic when taken to the data. In the limit, Section 1.3 will show that ignoring parameter heterogeneity gives rise to autoregressive coefficients that fail to converge to their average values and otherwise lack a clear interpretation. What's more, even in this relatively simple AR(2), it is unknown how to write valid moment conditions for unit-specific (random coefficients) autoregressive parameters.[2] The short panel also rules out long panel methodologies like estimating and pooling unit-specific AR(2) processes (Pesaran and Smith, 1995).

Challenges remain in two common solutions for addressing parameter heterogeneity. First, it is often unclear whether there is leftover heterogeneity after splitting the sample by observables and estimating the model on these subgroups. Second, additional statistical restrictions—like restricting parameter heterogeneity to discrete types or likelihood-based approaches—potentially require specialized knowledge (e.g. the choice of functional forms and tuning parameters) with intensive computing resources (Bonhomme et al., 2019; Browning et al., 2010).

A test with power directed toward detecting parameter heterogeneity is a step towards addressing these challenges. Such a test can help determine if the assumption of parameter homogeneity approximately holds when estimating a model for a particular subgroup and can help determine if there is sufficient heterogeneity to justify switching to more sophisticated statistical modeling approaches.

1.2.2 Example: Illustrative measurement error model

Having outlined the contexts in which a test with power directed toward parameter heterogeneity would be useful, this subsection gives intuition for such a test through an illustrative measurement error model.

Suppose the researcher observes an i.i.d. sample $\mathbf{y}_i = (y_{1i}, y_{2i})'$ where

$$\mathbf{y}_i = \mathbf{h}(\theta_i) + \boldsymbol{\varepsilon}_i = \begin{bmatrix} \theta_i \\ \theta_i^2 \end{bmatrix} + \begin{bmatrix} \varepsilon_{1i} \\ \varepsilon_{2i} \end{bmatrix}$$

[2]Pesaran and Yang (2024) gives moment conditions for a random coefficients AR(1) but doesn't apply to models with more complex dynamics.

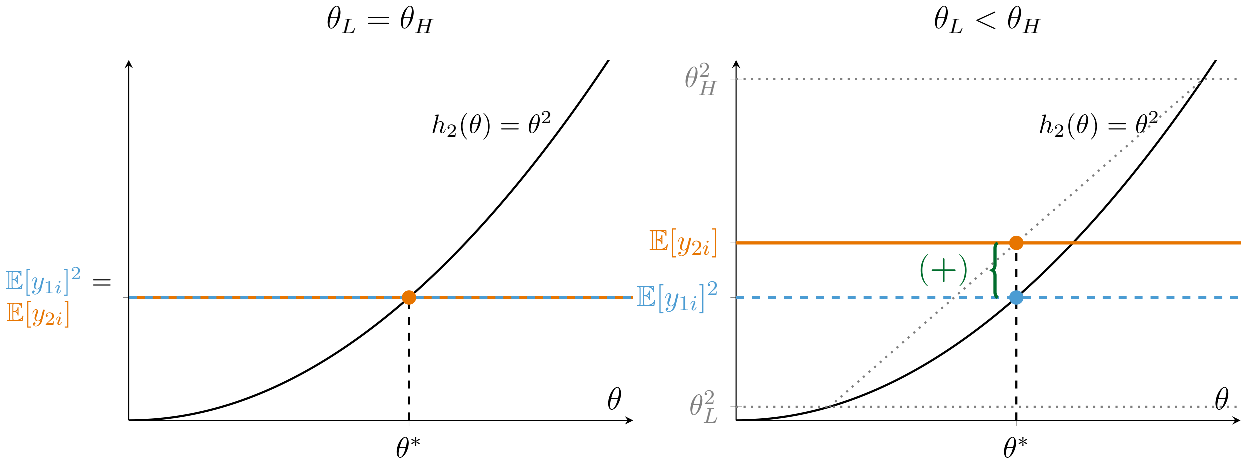

Figure 1.2.1: Plots of $h_2(\theta) = \theta^2$ and $\mathbb{E}(y_{2i})$ under parameter homogeneity (left panel) and under parameter heterogeneity (right panel).

for $\mathbf{h}(\theta) = (h_1(\theta), h_2(\theta))'$. The mean-zero measurement error $\boldsymbol{\varepsilon}_i$ is jointly uncorrelated, i.i.d. over i, and independent of θ_i. Random coefficient θ_i has mean θ^* and equally mixes between θ_L and θ_H. Data y_{1i} and y_{2i} can be interpreted as noisy observations of θ_i and θ_i^2 respectively.

Analogous to the more complicated moment condition models used in practice, the specific distribution of the measurement errors is unknown ruling out likelihood-based approaches. Restricted to using only the mean of $\mathbf{y}_i$, the researcher seeks to answer two questions. First, what is θ^*? Second, can parameter heterogeneity be detected? Comparing the data $\mathbb{E}[\mathbf{y}_i]$ to the model $\mathbf{h}(\theta)$, there are two moments ($\mathbb{E}[y_{1i}]$ and $\mathbb{E}[y_{2i}]$) and one unknown (θ^*).

Beginning with the first question, the first moment identifies θ^* . Since the measurement error is mean-zero, $\mathbb{E}[y_{1i}] = \mathbb{E}[\theta_i] + \mathbb{E}[\varepsilon_{1i}] = h_1(\theta^*) = \theta^*$. From linearity of $h_1(\theta)$, θ^* is identified under both the null and alternative hypotheses.

Moving to the second question, the second moment can be used to “test” for parameter heterogeneity where θ^* is taken from the first moment. Figure 1.2.1 plots the data ($\mathbb{E}[y_{2i}]$) against the function $h_2(\theta) = \theta^2$. Absent parameter heterogeneity (left panel), the data is compatible with θ^* and intersects $h_2(\theta)$ at θ^*. In equations, $\mathbb{E}[y_{2i}] = \mathbb{E}[y_{1i}]^2$. With parameter heterogeneity (right panel), however, y_{2i} mixes equally between processes with parameters θ_L and θ_H giving the orange horizontal line. Since $h_2(\theta)$ is convex about θ^*, Jensen's inequality implies the mean of y_{2i} under parameter heterogeneity is strictly *greater* than $\mathbb{E}[y_{1i}]^2$

$$\mathbb{E}[y_{2i}] = \frac{\theta_L^2 + \theta_H^2}{2} > \mathbb{E}[y_{1i}]^2.$$

Contrasting $\mathbb{E}[y_{1i}]$ with $\mathbb{E}[y_{2i}]$ gives a testable restriction.

In fact, the convexity/concavity of $h_2(\theta)$ about θ^* and the sign/magnitude of the gap between $\mathbb{E}[y_{2i}]$ and $\mathbb{E}[y_{1i}]^2$ are informative of the presence of parameter heterogeneity. The positive gap between $\mathbb{E}[y_{2i}]$ and $\mathbb{E}[y_{1i}]^2$ would widen if $h_2(\theta)$ were more convex about θ^* and would shrink if $h_2(\theta)$ were more linear about θ^*. If $h_2(\theta)$ were instead concave about θ^* (like if $h_2(\theta) = \sqrt{\theta}$), then $\mathbb{E}[y_{2i}] < \sqrt{\mathbb{E}[y_{1i}]}$.

Taking stock. This simple example presents two lessons for more general analysis. First, applicable to many economic applications, parameter heterogeneity is detectable without fully specifying the density of $\boldsymbol{\varepsilon}_i$. Economic theory often restricts the generating model ($\mathbf{h}(\theta)$ in this simple example), but is uninformative of the specific distribution. Second, overidentifying restrictions and nonlinearity of the moment condition in the parameter of interest are key ingredients for a test of parameter heterogeneity. Jensen's inequality implies that the sign and magnitude of misspecification in the nonlinear moment about θ^* are informative of the presence of parameter heterogeneity. Shown in Section 1.3, accounting for the sign of the gap gives rise to power gains from considering a one-sided versus a two-sided test.

The following section will outline a test of overidentifying restrictions with power directed toward parameter heterogeneity for a general nonlinear moment condition model. Unlike the special case of the simple example, the proposed test will consider environments without a clear separation between "estimation" and "test" moments. I will show that the GMM estimator lacks a clear interpretation when unit-level parameter heterogeneity is present.

1.3 Testing for parameter heterogeneity

This section describes the assumptions required for the test of heterogeneity and the construction of the test.

1.3.1 Generating model

Suppose the researcher observes an i.i.d. sample $(\mathbf{x}_i', \mathbf{y}_i')'$ for covariates $\mathbf{x}_i$ and T-dimensional outcome variable $\mathbf{y}_i$. Outcome variable $\mathbf{y}_i$ is determined by a smooth generating model that depends on a p-dimensional unit-specific parameter vector $\boldsymbol{\theta}_i$ and vector of unobserved shocks $\boldsymbol{\varepsilon}_i$:

$$\mathbf{y}_i = \mathbf{f}(\mathbf{x}_i, \boldsymbol{\varepsilon}_i; \boldsymbol{\theta}_i) \tag{1.3.1}$$

$$\theta_{ki} = \theta_k^* + u_{ki}\sigma_k \text{ for } k = 1, \ldots p. \tag{1.3.2}$$

Random vector $\mathbf{u}_i = (u_{1i}, \ldots, u_{pi})'$ has mean zero with covariance matrix $\mathbf{C}$ with 1's along its diagonal and is independent of both the covariates $\mathbf{x}_i$ and unobserved shocks $\boldsymbol{\varepsilon}_i$. p-dimensional vector $\boldsymbol{\sigma} = (\sigma_1, \ldots, \sigma_p)'$ has non-negative elements. All random variables are i.i.d. across units. The generating model is potentially unknown and is subject to restrictions governed by a set of moment conditions to be described later. Vector $\boldsymbol{\theta}^* = (\theta_1^*, \ldots, \theta_p^*)'$ is the mean of $\boldsymbol{\theta}_i$.

The correlation of the elements of the parameter vector $\boldsymbol{\theta}_i$ capture features potentially present in applications. Consider for example the estimation of production functions. For sub-industries with advanced manufacturing technologies, capital-intensive firms may be efficient users of capital or materials. This efficiency may induce correlation among the output elasticity of capital and the output elasticity of materials.

It is typically difficult *a priori* to establish a relationship between parameter heterogeneity and unobserved shocks $\boldsymbol{\varepsilon}_i$. The assumption of independence can be viewed as a starting point for a statistical test.

The researcher has access to an m-dimensional moment vector $\mathbf{g}(\mathbf{x}, \mathbf{y}; \boldsymbol{\theta})$ that has a mean of zero absent parameter heterogeneity when evaluated at $\boldsymbol{\theta}^*$

$$\mathbb{E}[\mathbf{g}(\mathbf{x}, \mathbf{y}^0; \boldsymbol{\theta}^*)] = \mathbf{0}_{m\times 1} \text{ where } \mathbf{y}^0 = \mathbf{f}(\mathbf{x}, \boldsymbol{\varepsilon}; \boldsymbol{\theta}^*). \tag{1.3.3}$$

The setup captures the scenario that a researcher has derived a set of moment conditions that would be correct for a representative agent, but is concerned about the assumption of parameter heterogeneity. These moment conditions are all that is willing to be assumed for two leading reasons:

1. The researcher is unwilling to fully specify the distributions of $\mathbf{x}_i$ and $\boldsymbol{\varepsilon}_i$; the underlying economic argument yields implications for the generating model but is uninformative on the specific form of the underlying random variables.

2. Deriving moment conditions that would be valid under parameter heterogeneity either isn't possible or would require additional implausible restrictions.

Assumption 1.3.1. $\boldsymbol{\sigma} = \mathbf{s}n^{-1/4}$ *where p-dimensional vector* $\mathbf{s}$ *is non-negative.*

Assumption 1.3.1 states that parameter heterogeneity shrinks with the sample size. The assumption is used to establish formal properties of the testing procedure: the amount of heterogeneity is large enough to threaten the coverage of the conventional confidence intervals of a GMM estimator yet small enough that no test can detect it with certainty. As will be discussed in Section 1.3.3, the specific rate of $n^{-1/4}$ is consistent with the literature on testing for the presence of mixtures in parametric models, as the asymptotic bias will appear in the moment condition's second order expansion. Intuitively, the rate can be viewed as a root-n rate on the variance of the random coefficient since $\text{Var}(\boldsymbol{\theta}_i) = \boldsymbol{\Lambda}/\sqrt{n}$, where $\boldsymbol{\Lambda} := \mathbf{s}\mathbf{s}' \odot \mathbf{C}$ and $\odot$ is the Hadamard product (the unit-wise product of matrix elements). These different rates are also a feature in the study of weak instruments, unit roots, and misspecification in moment condition models (Staiger and Stock, 1997; Elliott et al., 1996; Andrews et al., 2017; Armstrong and Kolesár, 2021; Bonhomme and Weidner, 2022).

For what follows, the parameter under test consist of elements of the heterogeneity covariance matrix. Specifically, the parameter under test is $\boldsymbol{\lambda} := \mathbf{E}\text{vech}(\boldsymbol{\Lambda})$ for a selection matrix $\mathbf{E} = (\mathbf{e}_{i_1}, \ldots, \mathbf{e}_{i_{d_\lambda}})'$ that is chosen by the researcher (for unit vector $\mathbf{e}_i$ and $i_1 < \cdots < d_\lambda$) and where the half-vectorization operator vech() stacks the elements of the lower triangular portion of a square matrix into a column vector. If the researcher is interested in testing for heterogeneity in all elements of the heterogeneity covariance matrix $\boldsymbol{\Lambda}$, then $\mathbf{E}$ equals the identity matrix $\mathbf{I}$. Otherwise, $\mathbf{E}$ selects the subset of vech$(\boldsymbol{\Lambda})$ under test. Such a situation would arise if a parameter were already a random coefficient or if a researcher were interested in directing power toward detecting heterogeneity in a subset of the parameter vector.

Assumption 1.3.2. $\mathbf{E}'\boldsymbol{\lambda} = \text{vech}(\boldsymbol{\Lambda})$.

Assumption 1.3.2 restricts $\boldsymbol{\Lambda}$ to be compatible with the selection matrix $\mathbf{E}$. For example, if $\mathbf{E}$ selected the variances of the heterogeneity covariance matrix $\boldsymbol{\Lambda}$, then the off-diagonal elements of $\boldsymbol{\Lambda}$ would equal zero. The condition is trivial when all elements of the heterogeneity covariance matrix are under test (when $\mathbf{E} = \mathbf{I}$) and will be used for deriving an optimal test for subvector heterogeneity.

1.3.2 Setup

Since Equation 1.3.3 is all that the researcher is willing to assume in the absence of unit-level parameter heterogeneity, the researcher would proceed with estimating $\boldsymbol{\theta}^*$ by the generalized method of moments. The GMM estimator is defined as

$$\widehat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \frac{1}{n}\sum_{i=1}^{n} \mathbf{g}(\mathbf{x}, \mathbf{y}; \boldsymbol{\theta})'\widehat{\mathbf{W}}\mathbf{g}(\mathbf{x}, \mathbf{y}; \boldsymbol{\theta})$$

for positive semidefinite weight matrix estimator $\widehat{\mathbf{W}}$.

To establish the upcoming results on the asymptotic behavior of the GMM estimator and heterogeneity test, Assumption A.1.1 found in Appendix A.1.2 adapts the conditions of Newey and McFadden (1994) to my context. These assumptions restate textbook conditions required for GMM estimation to counterfactual data $\mathbf{y}^0$ generated absent parameter heterogeneity. For example, Condition (i) of Assumption A.1.1 states that absent parameter heterogeneity, parameter $\boldsymbol{\theta}^*$ is identified by the moment condition. Based on these conditions, I define the expected Jacobian matrix $\mathbf{G} = \mathbb{E}[\frac{\partial}{\partial \boldsymbol{\theta}'}\mathbf{g}(\mathbf{x}, \mathbf{y}^0; \boldsymbol{\theta}^*)]$ and the (positive definite) moment covariance matrix $\boldsymbol{\Sigma} = \mathbb{E}[\mathbf{g}(\mathbf{x}, \mathbf{y}^0; \boldsymbol{\theta}^*)\mathbf{g}(\mathbf{x}, \mathbf{y}^0; \boldsymbol{\theta}^*)']$ using the counterfactual data $\mathbf{y}^0$.

The Hessian matrix $\mathbf{H}$ (with dimensions m by $p(p+1)/2$) is the only new population object relative to those required for constructing confidence intervals in moment condition models and measures the local curvature of the moment vector about $\boldsymbol{\theta}^*$. For the simpler case of scalar θ, $\mathbf{H} = \mathbb{E}[\frac{\partial^2}{\partial \theta^2}\mathbf{g}(\mathbf{x}_i, \mathbf{y}_i^0; \theta^*)]$ is an m-dimensional vector and stores the second derivatives of the moment function with respect to θ evaluated at θ^*. For multivariate $\boldsymbol{\theta}$, $\mathbf{H}$ stores its own and cross second partial derivatives for each moment in the moment function:

$$\mathbf{H} = \begin{bmatrix} \mathrm{vech}(\mathbf{H}_1)' \\ \vdots \\ \mathrm{vech}(\mathbf{H}_m)' \end{bmatrix} \text{ where } \mathbf{H}_r = \mathbb{E}\Big[\frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} g_r(\mathbf{x}_i, \mathbf{y}_i^0; \boldsymbol{\theta}^*)\Big] \text{ for } r = 1, \ldots, m \tag{1.3.4}$$

For what follows, the duplication matrix $\mathbf{D}$ maps the half-vectorization operator to the vectorization operator $\mathrm{vec}(\mathbf{A}) = \mathbf{D}\mathrm{vech}(\mathbf{A})$ for symmetric matrix $\mathbf{A}$.

Assumption A.1.2 of Appendix A.1.2 includes regularity conditions that are required for the consistency under the local alternative.

Assumption 1.3.3 (Smoothness)**.** *For the model in Equation 1.3.1, the following conditions also hold:*

i) For all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, $\mathbf{f}(\mathbf{x}, \boldsymbol{\varepsilon}, \boldsymbol{\theta})$ is twice continuously differentiable in $\boldsymbol{\theta}$ with probability 1.

ii) For all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, $\mathbf{g}(\mathbf{x}, \mathbf{y}; \boldsymbol{\theta})$ is three times continuously differentiable in $\mathbf{y}$ and $\boldsymbol{\theta}$ with probability 1.

iii) For moment r, $\mathbf{0}_{T\times p} = \frac{\partial^2 g_r(\mathbf{x}, \mathbf{y}; \boldsymbol{\theta})}{\partial \mathbf{y} \partial \boldsymbol{\theta}'}$.

Assumption 1.3.3 imposes smoothness on the moment function and generating model. These conditions are consistent with the applications to dynamic panel models. The conditions are also required for the Taylor expansions used in proving consistency and asymptotic normality of the GMM estimator. In particular, the stronger requirement of three times continuous differentiability of the moment function imposed by the second condition is used in Lemma A.1.4 of Appendix A.1.2 for proving consistency of the sample expected Hessian matrix estimator $\widehat{\mathbf{H}}$ under local parameter heterogeneity. The requirement $\mathbf{0}_{T\times p} = \frac{\partial^2 g_r(\mathbf{x},\mathbf{y};\boldsymbol{\theta})}{\partial \mathbf{y}\partial \boldsymbol{\theta}'}$ in part iii) is a common feature of dynamic panel applications. Simplifying expressions, the assumption allows the bias induced by parameter heterogeneity to be summarized by $\mathbf{H}$ alone (see the forthcoming results).[3]

Example. (Illustrative measurement error model, continued from Section 1.2.2.) Written in a moment condition framework, the moment function, expected Jacobian matrix, and expected Hessian matrix are

$$\mathbf{g}(\mathbf{y};\theta) = \begin{bmatrix} y_1 - \theta \\ y_2 - \theta^2 \end{bmatrix}, \mathbf{G} = \begin{bmatrix} -1 \\ -2\theta^* \end{bmatrix}, \mathbf{H} = \begin{bmatrix} 0 \\ -2 \end{bmatrix}.$$

The moment condition contrasts the data with the model under parameter homogeneity. The first element of the expected Hessian matrix is zero since the first moment condition is linear in the parameter θ. To ease exposition, I will take the moment covariance matrix for this example to equal the identity matrix $\boldsymbol{\Sigma} = \mathbf{I}$.

[3]Condition iii) of Assumption 1.3.3 can be dropped. Applying Lemma A.1.3, it suffices to modify $\mathbf{H}$ in Definition 1.3.4 to include additional terms that contain second cross partial derivatives with respect to $\boldsymbol{\theta}$ and $\mathbf{y}$. Specifically, replace $\mathbf{H}_r$ with $\mathbb{E}\Big[\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} g_r(\mathbf{x}_i, \mathbf{y}_i^0; \boldsymbol{\theta}^*)\Big] + \mathbb{E}\Big[\Big(\frac{\partial \mathbf{f}(\mathbf{x}_i,\boldsymbol{\varepsilon}_i;\boldsymbol{\theta}^*)}{\partial\boldsymbol{\theta}'}\Big)' \frac{\partial^2 g_r(\mathbf{x}_i,\mathbf{y}_i^0;\boldsymbol{\theta}^*)}{\partial\mathbf{y}\partial\boldsymbol{\theta}'}\Big] + \mathbb{E}\Big[\Big(\frac{\partial \mathbf{f}(\mathbf{x}_i,\boldsymbol{\varepsilon}_i;\boldsymbol{\theta}^*)}{\partial\boldsymbol{\theta}'}\Big)' \frac{\partial^2 g_r(\mathbf{x}_i,\mathbf{y}_i^0;\boldsymbol{\theta}^*)}{\partial\mathbf{y}\partial\boldsymbol{\theta}'}\Big]'$.

1.3.3 Asymptotic bias and fitted moments

Lemma 1.3.1. *Impose Assumptions 1.3.1, 1.3.2, 1.3.3, A.1.1, and A.1.2. Then,*

$$\frac{\sqrt{n}}{n}\sum_{i=1}^{n}\mathbf{g}(\mathbf{x}_i,\mathbf{y}_i;\boldsymbol{\theta}^*) = \frac{\sqrt{n}}{n}\sum_{i=1}^{n}\Big[\mathbf{g}(\mathbf{x}_i,\mathbf{f}(\mathbf{x}_i,\boldsymbol{\varepsilon}_i;\boldsymbol{\theta}^*);\boldsymbol{\theta}^*) - \frac{1}{2}\mathbf{H}(\mathbf{D}'\mathbf{D})\mathbf{E}'\boldsymbol{\lambda}\Big] + o_p(1).$$

Lemma 1.3.1 shows the moment conditions evaluated at the mean parameter $\boldsymbol{\theta}^*$ are asymptotically biased. The rate of $n^{-1/4}$ in Assumption 1.3.1 gives rise to a second order term coming from the second order Taylor expansion of $\mathbf{y}_i = \mathbf{f}(\mathbf{x}_i,\boldsymbol{\varepsilon}_i;\boldsymbol{\theta}_i)$ about $\boldsymbol{\theta}^*$. Notably, the first order term from the Taylor expansion of $\mathbf{y}_i$ disappears because $\mathbf{u}_i$ is mean zero and is independent of $\mathbf{x}_i$ and $\boldsymbol{\varepsilon}_i$.[4] The rate $n^{1/4}$, disappearance of a first-order term in a Taylor expansion, and reliance on the second-order term appear when testing for the presence of heterogeneous parameters for fully-specified parametric models (Lindsay, 1995; Gu, 2016). Mirroring the Jensen's inequality intuition found in the measurement error example of Section 1.2.2, the asymptotic bias term is nonzero when the moment function is nonlinear in $\boldsymbol{\theta}$ about $\boldsymbol{\theta}^*$ (measured through Hessian matrix $\mathbf{H}$) and under parameter heterogeneity (when the heterogeneity covariance matrix $\boldsymbol{\Lambda}$ is non-zero). The duplication matrix $\mathbf{D}$ appears because the expected Hessian matrix $\mathbf{H}$ is written using the half-vectorization operator.

Proposition 1.3.1 (Asymptotic normality)**.** *Impose Assumptions 1.3.1, 1.3.2, 1.3.3, A.1.1, and A.1.2. Then,*

$$\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*) \xrightarrow{d} \mathcal{N}(\frac{1}{2}(\mathbf{G}'\mathbf{W}\mathbf{G})^{-1}\mathbf{G}'\mathbf{W}\mathbf{H}(\mathbf{D}'\mathbf{D})\mathbf{E}'\boldsymbol{\lambda}, \operatorname{avar}(\widehat{\boldsymbol{\theta}}))$$

where $\operatorname{avar}(\widehat{\boldsymbol{\theta}}) = (\mathbf{G}'\mathbf{W}\mathbf{G})^{-1}\mathbf{G}'\mathbf{W}\boldsymbol{\Sigma}\mathbf{W}'\mathbf{G}(\mathbf{G}'\mathbf{W}\mathbf{G})^{-1}$.

Local parameter heterogeneity induces asymptotic bias in the GMM estimator, which ultimately affects inference. Proposition 1.3.1 shows that the asymptotic bias (which is at

[4] An alternative approach considers a rate of $n^{-1/2}$ on the reparametrized $\eta_k = \sigma_k^2$ and uses L'Hôpital's rule (Chesher, 1984; Hahn et al., 2014). Such an approach would give rise to a similar asymptotic expansion but would require the additional assumption of symmetry on the heterogeneity distribution as discussed in Lindsay (1995).

the same magnitude as the standard errors) is zero under parameter homogeneity ($\boldsymbol{\Lambda} = \mathbf{0}$) and increases with nonlinearity in the moment condition about $\boldsymbol{\theta}^*$. The asymptotic bias ultimately affects coverage of the average parameter. Consider a $(1-\alpha) \times 100$ percent conventional confidence interval for $\boldsymbol{\theta}^*$, which is computed as $\widehat{\boldsymbol{\theta}} \pm z^*_{1-\alpha/2}\sqrt{\mathrm{avar}(\widehat{\boldsymbol{\theta}})/n}$ for the $1-\alpha/2$ quantile of a standard normal distribution $z^*_{1-\alpha/2}$. Since the asymptotic variance is the same as it would be under parameter homogeneity, the conventional confidence interval necessarily under-covers.

Proposition 1.3.2 (Fitted moments)**.** *Impose Assumptions 1.3.1, 1.3.2, 1.3.3, A.1.1, and A.1.2. Then,*

$$\frac{\sqrt{n}}{n}\sum_{i=1}^{n}\mathbf{g}(\mathbf{x}_i, \mathbf{y}_i; \widehat{\boldsymbol{\theta}}) \xrightarrow{d} \mathcal{N}(-\frac{1}{2}\mathbf{M_W}\mathbf{H}(\mathbf{D}'\mathbf{D})\mathbf{E}'\boldsymbol{\lambda}, \mathbf{M_W}\boldsymbol{\Sigma}\mathbf{M_W}')$$

for $\mathbf{M_W} = (\mathbf{I}_m - \mathbf{G}(\mathbf{G}'\mathbf{W}\mathbf{G})^{-1}\mathbf{G}'\mathbf{W})$.

Parameter heterogeneity however is detectable in the fitted GMM moments when there are more moments than parameters. Proposition 1.3.2 shows that the fitted GMM moments are generally non-zero. Like the expression for the asymptotic bias of the GMM estimator, the asymptotic mean of the fitted GMM moments shrinks with the sample size, depends on the magnitude of parameter heterogeneity (through heterogeneity covariance matrix $\boldsymbol{\Lambda}$), and increases with nonlinearity in the moment condition. Through $\mathbf{M_W}$ (compare to P_W in Newey (1985)), the term can be interpreted as the residual (scaled by $-1/2$) of the generalized least squares regression of $\mathbf{H}(\mathbf{D}'\mathbf{D})\mathbf{E}'\boldsymbol{\lambda}$ on $\mathbf{G}$ with precision matrix $\mathbf{W}$. Then, the resulting residual is non-zero when there are more moments than parameters just as in standard results on GMM over-identification tests (Hansen, 1982). Note that the asymptotic covariance matrix given in Proposition 1.3.2 is singular.

Example. (Illustrative measurement error model, continued) Deviating from the illustration, now suppose that the researcher attempts to estimate θ^* using *both* moments rather than solely the linear one. Under local heterogeneity, the GMM estimator with an identity weight matrix is asymptotically biased:

$$\sqrt{n}(\widehat{\theta} - \theta^*) \xrightarrow{d} \mathcal{N}\Big(\frac{2\theta^*}{4(\theta^*)^2 + 1}\Lambda, \frac{1}{4(\theta^*)^2 + 1}\Big).$$

The bias depends on the value of θ^* and on the degree of parameter heterogeneity. The fitted GMM moments are also off-center

$$\frac{\sqrt{n}}{n}\sum_{i=1}^{n}\mathbf{g}(\mathbf{y}_i; \widehat{\theta}) \xrightarrow{d} \mathcal{N}\Big(\frac{1}{4(\theta^*)^2 + 1}\begin{bmatrix} -2\theta^* \\ 1 \end{bmatrix}\Lambda, \frac{1}{4(\theta^*)^2 + 1}\begin{bmatrix} 4(\theta^*)^2 & -2\theta^* \\ -2\theta^* & 1 \end{bmatrix}\Big).$$

The first linear fitted GMM moment condition is off-center for $\theta^* \neq 0$ since the GMM estimator is asymptotically biased.

1.3.4 Constructing a test for parameter heterogeneity

Section 1.3.4.1 describes a limit experiment that forms the basis of the test for parameter heterogeneity. Section 1.3.4.2 introduces the heterogeneity test in the special case of testing for scalar heterogeneity. Section 1.3.4.3 generalizes the heterogeneity test to the multivariate case. See Section 1.4 for a discussion on semiparametric efficiency.

1.3.4.1 Limit experiment

The limiting distribution of the fitted GMM moments in Proposition 1.3.2 summarizes the over-identifying restrictions of a moment condition model and can be viewed as a limit experiment. Proposition 1.3.1 and Lemma A.1.4 of Appendix A.1.2 imply that the sample estimators of the expected Jacobian $\widehat{\mathbf{G}} = \frac{1}{n}\sum_{i=1}^{n}\frac{\partial}{\partial\boldsymbol{\theta}'}\mathbf{g}(\mathbf{x}_i, \mathbf{y}_i; \widehat{\boldsymbol{\theta}})$, expected Hessian $\widehat{\mathbf{H}}_r = \frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}g_r(\mathbf{x}_i, \mathbf{y}_i; \widehat{\boldsymbol{\theta}})$, and moment covariance $\widehat{\boldsymbol{\Sigma}} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{g}(\mathbf{x}_i, \mathbf{y}_i; \widehat{\boldsymbol{\theta}})\mathbf{g}(\mathbf{x}_i, \mathbf{y}_i; \widehat{\boldsymbol{\theta}})'$ matrices converge in probability to $\mathbf{G}, \mathbf{H}_r$, and $\boldsymbol{\Sigma}$ respectively under the alternative hypothesis. These matrices can therefore be treated as known for hypothesis testing using the limiting distribution of Proposition 1.3.2 noted as $\mathbf{X}$:

$$\mathbf{X} \sim \mathcal{N}(-\frac{1}{2}\mathbf{M_W H}(\mathbf{D'D})\mathbf{E'}\boldsymbol{\lambda}, \mathbf{M_W \Sigma M_W}'). \tag{1.3.5}$$

The above display can be viewed as a limit experiment consisting of a single draw from a multivariate normal distribution with a known covariance matrix. I formalize this discussion in Section 1.4.

The score vector for testing $\boldsymbol{\lambda} = \mathbf{0}$ in Equation 1.3.5 summarizes the deviations of $\mathbf{X}$ from the null hypothesis of no parameter heterogeneity. The score vector in this case is the gradient of the log-likelihood function of the multivariate normal random variable described in Equation 1.3.5 with respect to $\boldsymbol{\lambda}$ evaluated at $\boldsymbol{\lambda} = \mathbf{0}$. Proposition A.1.2 of Appendix A.1.2 shows that the score vector is equal in distribution (up to a constant) to the Hessian-weighted sum of the normalized fitted GMM moments

$$\mathbf{S} = \mathbf{E}(\mathbf{D'D})\mathbf{H}'\boldsymbol{\Sigma}^{-1}\mathbf{X} \sim \mathcal{N}(-\frac{1}{2}\boldsymbol{\Omega\lambda}, \boldsymbol{\Omega}) \tag{1.3.6}$$

where $\boldsymbol{\Omega} = \mathbf{E}(\mathbf{D'D})\mathbf{H}'\boldsymbol{\Sigma}^{-1}\mathbf{MH}(\mathbf{D'D})\mathbf{E}'$ and $\mathbf{M} = \mathbf{M}_{\boldsymbol{\Sigma}^{-1}}$. Proposition A.1.2 also implies that the score vector is invariant to the choice of weight matrix $\mathbf{W}$ so long as $\mathbf{G'WG}$ is full rank—the test statistic remains the same regardless of the choice of weight matrix. If testing for heterogeneity for a subset of the parameter vector, $\mathbf{E}$ select the columns of the expected Hessian matrix $\mathbf{H}$ that correspond to the parameters under test since $\mathbf{D'D}$ is diagonal. Thus, in practice, the full expected Hessian matrix needn't be computed if testing for heterogeneity for a subset of the parameter vector.

Assumption 1.3.4. $\boldsymbol{\Omega}$ *is full rank.*

Assumption 1.3.4 can be viewed as a joint requirement of the existence of over-identifying restrictions (through $\mathbf{M}$) and local nonlinearity of the moment function corresponding to the parameters under test (through $\mathbf{E}(\mathbf{D}'\mathbf{D})\mathbf{H}$). Since $\boldsymbol{\Omega}$ is positive semidefinite by construction, the condition ensures $\boldsymbol{\Omega}$ is invertible and positive definite.

1.3.4.2 Test for scalar heterogeneity

When Λ is a scalar, a test for heterogeneity based on the limiting score vector of Equation 1.3.6 is one-sided. Under Assumption 1.3.4, the scalar score S defined in Equation 1.3.6 is mean-zero under the null hypothesis and necessarily has a negative mean under the alternative hypothesis since Ω is positive. Thus, an *infeasible* level α test rejects when $\frac{\mathbf{H}'\boldsymbol{\Sigma}^{-1}\mathbf{X}}{\sqrt{\Omega}} < z_\alpha$ where z_α gives the α quantile of a standard normal distribution. Such a test is infeasible because the population objects are required to be known. A feasible test requires replacing the population matrices and vectors with their sample analogues (with full conditions specified in the multivariate test of Definition 1.3.1 to follow).

Example. (Illustrative measurement error model, continued) For $\Omega = \frac{4}{1+4(\theta^*)^2}$, the sample analogue of the scalar score is

$$\widehat{S} = \mathbf{H}'\boldsymbol{\Sigma}^{-1}\frac{\sqrt{n}}{n}\sum_{i=1}^{n}\mathbf{g}(\mathbf{y}_i;\widehat{\theta}) = -2\frac{\sqrt{n}}{n}\sum_{i=1}^{n}[y_{2i} - \widehat{\theta}^2] \xrightarrow{d} \mathcal{N}(-\frac{1}{2}\Omega\Lambda, \Omega).$$

Recall that the expected Hessian matrix $\mathbf{H}$ doesn't depend on $\widehat{\theta}$ since the polynomial order of the two moment conditions with respect to θ is less than or equal to two (and $\boldsymbol{\Sigma}$ is taken to be the identity matrix for ease of exposition). Observe that the first element of the expected Hessian matrix is zero since the first moment condition is linear in θ. Consistent with the Jensen's inequality intuition from Section 1.2, the score vector examines deviations between the second and first moment conditions. A feasible asymptotically level α test rejects when $\widehat{S}/\sqrt{\widehat{\Omega}} < z_\alpha$. Consistent with the visual illustration, rejection occurs when the sample mean of y_{2i} is sufficiently greater than $\widehat{\theta}^2$.

1.3.4.3 Test for multivariate heterogeneity

When $\mathbf{\Lambda}$ is instead a matrix (and again imposing Assumption 1.3.4), the one-sided scalar test can be generalized by computing the generalized likelihood ratio test statistic on the score vector of Equation 1.3.6. This procedure is based on Silvapulle and Silvapulle (1995) and is analogous to the proposal of the parametric test of Gu (2016). I study the power properties of the test in Section 1.4.2. For now, begin with defining an infeasible test statistic that requires knowledge of the population objects

$$\mathcal{T} = \mathbf{S}'\mathbf{\Omega}^{-1}\mathbf{S} - \inf_{\boldsymbol{\lambda}:\, \mathrm{vech}^{-1}(\mathbf{E}'\boldsymbol{\lambda})\in\mathcal{C}} \left[\mathbf{S} + \frac{1}{2}\mathbf{\Omega}\boldsymbol{\lambda}\right]'\mathbf{\Omega}^{-1}\left[\mathbf{S} + \frac{1}{2}\mathbf{\Omega}\boldsymbol{\lambda}\right] \tag{1.3.7}$$

where $\mathcal{C}$ gives the set of positive semidefinite matrices and vech^{-1} reconstructs a symmetric matrix from its half-vectorized form. A level α test rejects when $\mathcal{T} > cv_\alpha$ for critical value cv_α and works by contrasting the score vector under the null (first term) and alternative hypothesis (second term).

Under the null hypothesis, the score vector is mean zero $\mathbf{S} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Omega})$ and $\mathcal{T}$ follows the chi-bar-square distribution. This distribution is formed as a mixture of chi-squared random variables with further properties described in textbook references on ordered regressions like Robertson et al. (1988) and Silvapulle and Sen (2001). More generally, this distribution appears in tests on the boundary of the parameter space (see for example Chernoff (1954); Andrews (1996)). Computation of the test statistic involves a convex objective function with convex constraints and is thus easy to implement using standard optimization software.[5] The critical value for a level α test can be computed through simulation: (1) Take draws of $\mathbf{S} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Omega})$, (2) Using the draws from Step 1, compute the test statistic of Equation 1.3.7, (3) The critical value cv_α is the $1-\alpha$ quantile of the distribution from Step 2. The infeasible scalar test outlined in Section 1.3.4.2 is a special case of the multivariate test outlined here when $\alpha < 0.5$.

[5]Throughout this paper, I compute the test statistic using the `fmincon` function in MATLAB constraining the eigenvalues of $\mathbf{\Lambda}$ to be non-negative.

Accounting for the positive semidefinite constraint of $\mathbf{\Lambda}$ gives rise to power gains relative to tests with a limiting chi-squared distribution under the null hypothesis. Let's proceed term-by-term. The first term of $\mathcal{T}$ is $\mathbf{S}'\mathbf{\Omega}^{-1}\mathbf{S}$ and can be understood as a measure of general misspecification of the score. The term examines general deviations of $\mathbf{S}$ from zero rather than solely those arising from positive semidefinite matrices $\mathcal{C}$. By itself, this first term is the traditional Rao score test statistic (see e.g. Chapter 12.4.3 of Lehmann and Romano (2022)) and follows a chi-squared distribution under the null hypothesis. The second term $\inf_{\boldsymbol{\lambda}:\,\mathrm{vech}^{-1}(\mathbf{E}'\boldsymbol{\lambda})\in\mathcal{C}}\left[\mathbf{S}+\frac{1}{2}\mathbf{\Omega}\boldsymbol{\lambda}\right]'\mathbf{\Omega}^{-1}\left[\mathbf{S}+\frac{1}{2}\mathbf{\Omega}\boldsymbol{\lambda}\right]$ directs power toward the specific alternative hypothesis of parameter heterogeneity. To see this, suppose for simplicity that $\mathbf{E}=\mathbf{I}$ and that the constraint of positive semidefiniteness of $\mathbf{\Lambda}$ were dropped—that $\mathcal{C}$ instead equaled $\mathbb{R}^{p(p+1)/2\times p(p+1)/2}$. Then the second term would be identically zero, giving rise to a test statistic $\mathcal{T}$ that equals the first term.

Definition 1.3.1 is the feasible analogue to the infeasible test statistic of Equation 1.3.7, replacing the population objects with their corresponding estimators:

Definition 1.3.1 (Heterogeneity test statistic)**.** *Impose Assumptions 1.3.1, 1.3.2, 1.3.3, A.1.1, A.1.2, and 1.3.4. Let* $\mathbf{S}_n=\mathbf{E}(\mathbf{D}'\mathbf{D})\widehat{\mathbf{H}}'\widehat{\mathbf{\Sigma}}^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbf{g}(\mathbf{x}_i,\mathbf{y}_i;\widehat{\boldsymbol{\theta}}^{\mathrm{eff}})$*. Then, the **heterogeneity test statistic** is*

$$\mathcal{T}_n=\mathbf{S}_n'\widehat{\mathbf{\Omega}}^{-1}\mathbf{S}_n-\inf_{\boldsymbol{\lambda}:\,\mathrm{vech}^{-1}(\mathbf{E}'\boldsymbol{\lambda})\in\mathcal{C}}\left[\mathbf{S}_n+\frac{1}{2}\widehat{\mathbf{\Omega}}\boldsymbol{\lambda}\right]'\widehat{\mathbf{\Omega}}^{-1}\left[\mathbf{S}_n+\frac{1}{2}\widehat{\mathbf{\Omega}}\boldsymbol{\lambda}\right]$$

where $\widehat{\boldsymbol{\theta}}^{\mathrm{eff}}$ *is an efficient GMM estimator,* $\widehat{\mathbf{\Omega}}=\mathbf{E}(\mathbf{D}'\mathbf{D})\widehat{\mathbf{H}}'\widehat{\mathbf{\Sigma}}^{-1}\widehat{\mathbf{M}}\widehat{\mathbf{H}}(\mathbf{D}'\mathbf{D})\mathbf{E}'$ *and* $\mathcal{C}$ *is the set of positive semidefinite matrices.*

Critical values for the multi-dimensional heterogeneity test statistic can be computed by following the same simulation procedure for the infeasible test after substituting $\mathbf{\Omega}$ with $\widehat{\mathbf{\Omega}}$.

Remarks

1. A normalized version of the score vector can be interpreted as a single draw from a multivariate normal distribution centered at $\boldsymbol{\lambda}$

$$-2\boldsymbol{\Omega}^{-1}\mathbf{S} \sim \mathcal{N}(\boldsymbol{\lambda}, \boldsymbol{\Omega}^{-1}).$$

   Rescaling, the observation can be interpreted as an estimate of parameter heterogeneity since $\text{Cov}(\boldsymbol{\theta}_i) = \boldsymbol{\Lambda}/\sqrt{n}$. Analogous to the "Moderate Measurement Error" condition of Evdokimov and Zeleneev (2023), this interpretation is most appropriate when the true data generating process is well approximated by the generating model of Section 1.3.1—in particular, that parameter heterogeneity is not too large and that parameter heterogeneity is independent of the underlying shock.

2. The proposed heterogeneity test achieves power gains over the J test in two main ways. First, the heterogeneity test is based on the score vector, which weighs the sample moments evaluated at the GMM estimator by the expected Hessian matrix $\mathbf{H}$. In contrast, the J test examines deviations of the fitted sample moments weighted by the inverse moment covariance matrix. The two tests therefore place different weights on the moment conditions. Second, the heterogeneity test directs power by exploiting the specific geometry of the alternative hypothesis parameter space through the likelihood ratio test. The J test looks to general deviations of the moment conditions using a quadratic form, where the degrees of freedom of the null chi-squared distribution increases with the number of moment conditions. Thus the heterogeneity test achieves power gains by looking to a narrower set of alternative hypotheses.

1.4 Semiparametrically efficient testing

Section 1.4.1 sets up the framework for evaluating semiparametric efficiency. Section 1.4.2 discusses the main result, showing that the proposed test maximizes a weighted average power criterion for distant alternatives. Section 1.4.3 contains additional technical details.

1.4.1 Setup

The infinite-dimensional nuisance parameters corresponding to the densities of $(\mathbf{x}', \mathbf{y}')'$ and $\mathbf{u}$ present a technical challenge in deriving results on asymptotic efficiency. The analysis is based on a *least favorable* submodel—one that makes the testing problem the "hardest." In this section, I adapt the discussion of Chapter 25 of van der Vaart (1998), allowing for the second order expansion required by the random coefficient framework and for multivariate testing.

The generating model of Section 1.3.1 can be expressed as the semiparametric model $P_{\boldsymbol{\theta}^*,\boldsymbol{\sigma}^*,\text{vech}(\mathbf{C}),p_{xy},p_u}$ where the densities of data $(\mathbf{x}', \mathbf{y}')'$ and $\mathbf{u}$ are p_{xy} and p_u respectively. These densities can be viewed as infinite-dimensional nuisance parameters. I consider a one-dimensional submodel indexed by t, which can be viewed as a smooth curve embedded within a larger semiparametric model that intersects the true model at $t = 0$. Later, I'll study local asymptotic power, examining paths that scale with the size of the sample $t = n^{-1/2}$.

Specifically, the model parameters drift from their true values $(\boldsymbol{\theta}^{*\prime}, \boldsymbol{\sigma}^{*\prime})'$ as t increases

$$\boldsymbol{\theta}_t = \boldsymbol{\theta}^* + \mathbf{c}t \text{ and } \boldsymbol{\sigma}_t = \boldsymbol{\sigma}^* + \mathbf{s}\sqrt{t} \tag{1.4.1}$$

where $\boldsymbol{\sigma}^* = \mathbf{0}$ (representing the null hypothesis of no parameter heterogeneity), $\mathbf{c} \in \mathbb{R}^p$, and $\mathbf{s}$ is a p-dimensional vector on the non-negative orthant. As will be discussed in Section 1.4.3, the curved path for $\boldsymbol{\sigma}_t$ is required for the second-order expansion of the semiparametric score function. The densities $p_{xy,t}(\mathbf{x}, \mathbf{y}; \boldsymbol{\theta})$ and $p_{u,t}(\mathbf{u})$ are unknown, differentiable with respect to t, and equal their true values $p_{xy}(\mathbf{x}, \mathbf{y}; \boldsymbol{\theta})$ and $p_u(\mathbf{u})$ when $t = 0$. Along the path t, the densities are subject to the moment condition of Equation 1.3.3.

Testing for the presence of parameter heterogeneity can then be stated as testing for the presence of a mixture distribution. Absent parameter heterogeneity (when $t = 0$), the density for $(\mathbf{x}', \mathbf{y}')'$ is $q_0(\mathbf{x}, \mathbf{y}) = p_{xy}(\mathbf{x}, \mathbf{y}; \boldsymbol{\theta}^*)$. With parameter heterogeneity (when $t > 0$), the density of $(\mathbf{x}', \mathbf{y}')'$ mixes over $\mathbf{u}$

$$q_t(\mathbf{x}, \mathbf{y}) = \int p_{xy,t}(\mathbf{x}, \mathbf{y}; \boldsymbol{\theta}_t + \text{diag}(\mathbf{u})\boldsymbol{\sigma}_t) p_{u,t}(\mathbf{u}) d\mathbf{u}.$$

Unlike parametric tests for the presence of mixture distributions, the specific distribution for $p_{xy,t}$ is unknown. These challenges are discussed in Section 1.4.3.

Score function. For computation of the semiparametric score function, Assumption A.1.3 of Appendix A.1.3 imposes two times continuous differentiability of the density of $(\mathbf{x}', \mathbf{y}')'$ with respect to $\boldsymbol{\theta}$. Lemma 1.4.1 shows that the semiparametric score function can be decomposed into parametric and non-parametric components:

Lemma 1.4.1. *Let operator $\frac{\partial_+}{\partial t}$ give the right-sided partial derivative with respect to t. Impose Assumptions 1.3.2, A.1.3 and A.1.4. Let $(\mathbf{x}', \mathbf{y}')'$ be on the support of the probability model $P_{\boldsymbol{\theta}^*, \boldsymbol{\sigma}^*, \text{vech}(\mathbf{C}), p_{xy}, p_u}$. Then,*

$$\frac{\partial_+}{\partial t} \log q_t(\mathbf{x}, \mathbf{y}) \mid_{t=0} = \mathbf{a}' \dot{\boldsymbol{\ell}}(\mathbf{x}, \mathbf{y}) + \eta(\mathbf{x}, \mathbf{y})$$

$$\dot{\boldsymbol{\ell}}(\mathbf{x}, \mathbf{y}) = \frac{1}{p_{xy}(\mathbf{x}, \mathbf{y}; \boldsymbol{\theta}^*)} \begin{bmatrix} \frac{\partial}{\partial \boldsymbol{\theta}} p_{xy}(\mathbf{x}, \mathbf{y}; \boldsymbol{\theta}^*) \\ \frac{1}{2}\mathbf{E}(\mathbf{D}'\mathbf{D})\text{vech}(\frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} p_{xy}(\mathbf{x}, \mathbf{y}; \boldsymbol{\theta}^*)) \end{bmatrix}$$

for $\mathbf{a} = (\mathbf{c}', \boldsymbol{\lambda}')'$, measurable function $\eta(\mathbf{x}, \mathbf{y}) = \partial \log p_t(\mathbf{x}, \mathbf{y}; \boldsymbol{\theta}^*)/\partial t$, and $\boldsymbol{\lambda} = \mathbf{E}\text{vech}(\boldsymbol{\Lambda})$.

$\dot{\boldsymbol{\ell}}(\mathbf{x}, \mathbf{y})$ can be interpreted as the score function for the parametric part of the model holding fixed the non-parametric part. The first term of $\dot{\boldsymbol{\ell}}(\mathbf{x}, \mathbf{y})$ is the score for $\boldsymbol{\theta}^*$ and the second term is the score for $\boldsymbol{\sigma}^*$. The second term arises from the curved path for $\boldsymbol{\sigma}_t$ and differs from standard results on semiparametric models where the parameter of interest is partitioned from the nuisance parameter (like Chapter 25.4 of van der Vaart (1998)). The selection matrix $\mathbf{E}$ selects the non-zero elements of the heterogeneity covariance matrix $\boldsymbol{\lambda} = \mathbf{E}\text{vech}(\boldsymbol{\Lambda})$ and appears to account for the case where the researcher is only interested in testing for parameter heterogeneity for a subset of the estimated parameters (when certain elements of $\boldsymbol{\Lambda}$ are known to be zero). $\eta(\mathbf{x}, \mathbf{y})$ can be interpreted as the score function for the density of $(\mathbf{x}', \mathbf{y}')'$ of the model holding fixed the parametric part. Notably, the score function for the density of $\mathbf{u}$ doesn't appear in the semiparametric score. For the results to follow, the technical condition of differentiability in quadratic mean described in Definition A.1.1 of Appendix A.1.3 is required for local asymptotic normality to hold.

1.4.2 Semiparametric efficiency

Having defined the semiparametric submodel, this subsection discusses semiparametric efficiency of the heterogeneity test in the scalar and multivariate cases. Theorem 1.4.1, the main result of this subsection, will show that the heterogeneity test is the most powerful test for moderate amounts of parameter heterogeneity. For what follows, let $t \mapsto P_{t,\mathbf{a}'\dot{\boldsymbol{\ell}}+\eta}$ be the submodel described in Section 1.4.1 (indexed by t) with arbitrary score function $\mathbf{a}'\dot{\boldsymbol{\ell}} + \eta$ matching the convention of Chapter 25.6 of van der Vaart (1998). Selection matrix $\mathbf{E}$ selects the elements of the heterogeneity covariance matrix under test. The "tangent set" describes the set of score functions and is $(\mathbf{c}', \boldsymbol{\lambda}')'\dot{\boldsymbol{\ell}}+\eta$ for $\mathbf{c} \in \mathbb{R}^p$, $\boldsymbol{\lambda} = \mathbf{E}\text{vech}(\boldsymbol{\Lambda})$ where $\boldsymbol{\Lambda} \in \mathcal{C}$, and $\eta \in \dot{\mathcal{P}}$ for the nuisance tangent space $\dot{\mathcal{P}}$. Since $\mathcal{C}$ is a convex cone, the tangent set is a convex cone. Lemma A.1.5 of Appendix A.1.3 shows that $\dot{\mathcal{P}}$ consists of all measurable mean-zero random functions $\eta(\mathbf{x}, \mathbf{y})$ with finite variance such that $\mathbf{0}_{m\times 1} = \iint g(\mathbf{x}, \mathbf{y}; \boldsymbol{\theta}^*)\eta(\mathbf{x}, \mathbf{y})p_{xy}(\mathbf{x}, \mathbf{y}; \boldsymbol{\theta}^*)d\mathbf{x}d\mathbf{y}$.

The tool underpinning Theorem 1.4.1 (to follow) is the asymptotic representation theorem (see e.g. Theorem 9.3 of van der Vaart (1998)).[6] A statistic that converges weakly in the model $P_{n^{-1/2},\mathbf{a}'\dot{\boldsymbol{\ell}}+\eta}$ is "matched" by some statistic in the limit experiment—in this case, the

[6]Care is required for applying the result. As discussed in the proof of Theorem 25.44 of van der Vaart (1998), the argument requires adapting results on parametric testing by specifying an orthonormal base of score functions. The covariance matrix, which is the inverse of the efficient information matrix, arises from a sufficiently large base.

limit experiment is a single observation $\mathbf{Z}$ of a multivariate normal distribution with mean $\text{vech}(\mathbf{\Lambda})$ and variance $4\mathbf{\Omega}^{-1}$:

$$\mathbf{Z} \sim \mathcal{N}(\text{vech}(\mathbf{\Lambda}), 4\mathbf{\Omega}^{-1}). \tag{1.4.2}$$

Consequently, no sequence of tests can asymptotically be "better" (as will be defined below) than the "best" test in the limit experiment. Hence, it suffices to study Equation 1.4.2 for establishing the asymptotic properties of tests. Equation 1.4.2 also implies that the discussion of Section 1.3.4—particularly on score vector $\mathbf{S}$— focuses on the correct limit experiment. To see this, a rescaled version of $\mathbf{Z}$ given by $-\frac{1}{2}\mathbf{\Omega}\mathbf{Z}$ is equal in distribution to $\mathbf{S}$.

Since the matrix $\mathbf{\Lambda}$ is unknown, Equation 1.4.2 shows that tests for multivariate parameter heterogeneity inevitably require trading off power in detecting heterogeneity among the parameters under test. Illustrating, one approach for evaluating the tests in the limit experiment of Equation 1.4.2 is to compare the power of a test for *one* particular pointwise alternative. Call such an alternative $\text{vech}(\overline{\mathbf{\Lambda}})$. Then, the Neyman-Pearson lemma implies that any other test is less powerful than the likelihood ratio test comparing the pointwise alternative $\text{vech}(\overline{\mathbf{\Lambda}})$ against the pointwise null $\mathbf{0}$. However, this criterion would fail to reward tests with power against other potential alternative hypotheses.

I will instead consider a weighted average power criterion (Andrews and Ploberger, 1994). The criterion is a weighted sum of the power of a test across potentially many alternatives. Specifically, I consider a criterion that maximizes power against ellipses centered about the null hypothesis. Mathematically, let's define the weight function $w_r(\boldsymbol{\lambda})$. For radius $r > 0$, let $p_\lambda(\boldsymbol{\lambda}; r)$ be the probability density function of $r\mathbf{\Omega}^{-1/2}\boldsymbol{\zeta}$ where $\boldsymbol{\zeta}$ is uniformly distributed on the $p(p+1)/2$-dimensional unit sphere. Then, the weight function $w_r(\boldsymbol{\lambda})$ places positive weight on values of the ellipse whose corresponding matrices $\text{vech}^{-1}(\boldsymbol{\lambda})$ are positive semidefinite

$$w_r(\boldsymbol{\lambda}) = 1[\text{vech}^{-1}(\boldsymbol{\lambda}) \in \mathcal{C}]p_\lambda(\boldsymbol{\lambda}; r)/K$$

where $K = \int 1[\text{vech}^{-1}(\boldsymbol{\lambda}) \in \mathcal{C}]p_\lambda(\boldsymbol{\lambda}; r)d\boldsymbol{\lambda}$.

To formally establish results, Assumption A.1.4 of Appendix A.1.3 gives technical conditions that allow the interchange of integrals and derivatives. Theorem 1.4.1 shows that the heterogeneity test asymptotically maximizes a weighted average power criterion with weight function $w_r(\boldsymbol{\lambda})$ as r diverges:

Theorem 1.4.1. *Impose Assumptions 1.3.2, 1.3.3, 1.3.4, A.1.1, A.1.2, and A.1.4. Suppose $P_{t,(\mathbf{c}',\boldsymbol{\lambda}')\dot{\boldsymbol{\ell}}+\eta}$ is differentiable in quadratic mean at $t = 0$ and $(\mathbf{c}', \boldsymbol{\lambda}')\dot{\boldsymbol{\ell}} + \eta$ is a member of the tangent set. Suppose $\mathbf{Z} \sim \mathcal{N}(\boldsymbol{\lambda}, 4\boldsymbol{\Omega}^{-1})$ and $\pi^*(\boldsymbol{\lambda}; r)$ is the power function of a level-α test of $\mathbf{Z}$ for $H_0\colon \boldsymbol{\lambda} = 0$ that maximizes the weighted average power criterion $\int \pi^*(\boldsymbol{\lambda}; r)w_r(\boldsymbol{\lambda})d\lambda$ for $r > 0$.*

1. *Let $P \mapsto \pi_n(P)$ be a sequence of power functions in the local experiment that is level-α for each n. Then,*

$$\lim_{r\to\infty} \limsup_{n\to\infty} \int \pi_n(P_{n^{-1/2},(\mathbf{c}',\boldsymbol{\lambda}')\dot{\boldsymbol{\ell}}+\eta})w_r(\boldsymbol{\lambda})d\boldsymbol{\lambda} \leq \lim_{r\to\infty} \int \pi^*(\boldsymbol{\lambda}; r)w_r(\boldsymbol{\lambda})d\boldsymbol{\lambda}.$$

2. *Let $\varphi_n^{\text{het}} = 1$ when the asymptotically level-α heterogeneity test based on $\mathbf{S}_n$ rejects and $\varphi_n^{\text{het}} = 0$ otherwise. Then,*

$$\lim_{r\to\infty} \lim_{n\to\infty} \int \mathbb{E}_{n^{-1/2},(\mathbf{c}',\boldsymbol{\lambda}')\dot{\boldsymbol{\ell}}+\eta}[\varphi_n^{\text{het}}]w_r(\boldsymbol{\lambda})d\boldsymbol{\lambda} = \lim_{r\to\infty} \int \pi^*(\boldsymbol{\lambda}; r)w_r(\boldsymbol{\lambda})d\boldsymbol{\lambda}.$$

The first part of the theorem establishes a power bound according to a weighted average power with weight function $\lim_{r\to\infty} w_r(\boldsymbol{\lambda})$ with diverging radius. The result can be viewed as an extension of the scalar semiparametric optimality result of Theorem 25.44 of van der Vaart (1998) to a multivariate weighted average power criterion applied to the context of this paper's setting. The second part of the theorem establishes that the heterogeneity test achieves this power bound for distant (through diverging radius r) though local (since heterogeneity shrinks with the sample size) alternatives. It is in this sense that the heterogeneity test is optimal against "moderate" amounts of parameter heterogeneity. In addition, the result shows that the heterogeneity test is asymptotically admissible, that any other test with higher power for a particular alternative hypothesis necessarily has lower power for another. The second part of the theorem can be viewed as a multivariate generalization of Theorem 25.45 of van der Vaart (1998) combined with the result of Andrews (1996).

For the special case of scalar heterogeneity, Corollary A.1.1 found in Appendix A.1.3 shows that the scalar test for heterogeneity described in Section 1.3.4.2 is the locally uniformly most powerful test. The result can be seen in Equation 1.4.2 for a scalar Λ. Here, the best test is one-sided and doesn't depend on the particular value of $\Lambda > 0$, rejecting for large positive values of Z. Corollary A.1.1 shows that the heterogeneity test "matches" this optimal test.

Remarks.

1. In the limit experiment, the test of Hahn et al. (2014) can be matched with a test that rejects for large values of $\frac{1}{4}\mathbf{Z}'\mathbf{\Omega}\mathbf{Z}$. As a result, the test rejects for a broader set of alternatives (for general deviations from $\mathbf{0}$ and not just those from the constraint of positive semi-definiteness) and experiences a power loss relative to one that accounts for the positive semidefiniteness constraint.

2. The analyses in this section extend the fully parametric analyses of Gu (2016), which also study heterogeneity tests in a LeCam framework, to moment condition models. Fully specifying a likelihood as is done in the framework of Gu (2016) gives rise to a $C(\alpha)$ test based on the second-order derivatives of the log likelihood function. These moments based on second-order scores are generally distinct from the moments used for estimation (those from the score function) and often can be interpreted as measuring the compatibility of higher order moments of the data as later illustrated in Section 1.5. The expansion of Lemma 1.4.1 reduces to that of Gu (2016) if the likelihood were instead fully specified, as the moment restrictions would be redundant.

3. For a general set of problems, it is difficult to make arguments in favor of one weight function over another (Andrews, 1998). The particular weight function in Theorem 1.4.1 can be viewed as the one that rationalizes the use of the likelihood ratio test—a test that is computationally tractable and well-studied.

4. The weighted average power result follows from the geometry of the null and alternative hypothesis parameter spaces. In particular, the null hypothesis is simple while the alternative hypothesis space is positively homogeneous since the space of positive semidefinite matrices is closed under multiplication by positive constants.

   Rewriting, the multivariate Gaussian limit experiment can be written as a Gaussian linear regression model with as many observations as unknowns:

$$\overline{\mathbf{Z}} = \mathbf{\Omega}^{1/2\prime}\boldsymbol{\lambda} + \boldsymbol{\eta}$$

   for independent standard normal vector $\boldsymbol{\eta}$ and $\overline{\mathbf{Z}} = -2\mathbf{\Omega}^{-1/2}\mathbf{S}$. $\overline{\mathbf{Z}}$ and $\mathbf{\Omega}^{1/2\prime}$ can be interpreted as data, while $\boldsymbol{\lambda}$ can be interpreted as regression coefficients. Thus, the limit experiment is a special case of Theorem 1 of Andrews (1996).

5. The multivariate test of Gu (2016) also considers a likelihood ratio test based on the score vector. Consequently, the application of the result of Andrews (1996) in Theorem 1.4.1 can also be applied, so the test of Gu (2016) also asymptotically maximizes weighted average power against distant alternatives in a fully parametric baseline model.

6. The scalar optimality result is a direct application of the results on semiparametric testing in Chapter 25.6 of van der Vaart (1998) while the general multivariate result of Theorem 1.4.1 extends these same arguments to the particular multivariate hypothesis testing problem considered in this paper.

1.4.3 Proof sketch

This subsection gives theoretical context for Theorem 1.4.1 by describing the computation of the efficient influence function and efficient information matrix, which can be viewed as an information bound for the heterogeneity covariance matrix.

The functional of interest is $\boldsymbol{\psi}(P_{t,(\mathbf{c}',\boldsymbol{\lambda}')\dot{\boldsymbol{\ell}}+\eta}) = \boldsymbol{\lambda}_0 + \boldsymbol{\lambda}t$ for $\boldsymbol{\lambda}_0 = \mathbf{0}$. The test consists of a simple null hypothesis on the boundary versus a composite alternative

$$H_0\colon \boldsymbol{\psi}(P) = \mathbf{0} \text{ vs. } H_1\colon \boldsymbol{\psi}(P) \in \{\boldsymbol{\lambda} : \text{vech}^{-1}(\mathbf{E}'\boldsymbol{\lambda}) \in \mathcal{C}\} \setminus \mathbf{0}. \tag{1.4.3}$$

Looking to the one-dimensional submodel, $\boldsymbol{\psi}(P_{t,(\mathbf{c}',\boldsymbol{\lambda}')\dot{\boldsymbol{\ell}}+\eta}) = \boldsymbol{\lambda}t$ belongs to the alternative when $t > 0$ while $\boldsymbol{\psi}(P_{0,(\mathbf{c}',\boldsymbol{\lambda}')\dot{\boldsymbol{\ell}}+\eta}) = \mathbf{0}$ belongs to the null hypothesis. I consider a neighborhood of the true model that shrinks with the sample size $t = n^{-1/2}$, so $\boldsymbol{\psi}(P_{n^{-1/2},(\mathbf{c}',\boldsymbol{\lambda}')\dot{\boldsymbol{\ell}}+\eta}) = \boldsymbol{\lambda}n^{-1/2}$.

Analogous to the information matrix in likelihood theory, the efficient information matrix can be interpreted as the summary of the information available for $\boldsymbol{\lambda}_0$. To compute, note that $\boldsymbol{\psi}(P_{t,(\mathbf{c}',\boldsymbol{\lambda}')\dot{\boldsymbol{\ell}}+\eta})$ is differentiable in the ordinary sense at $t = 0$ as $\partial_+\boldsymbol{\psi}(P_{t,(\mathbf{c}',\boldsymbol{\lambda}')\dot{\boldsymbol{\ell}}+\eta})/\partial t\,|_{t=0} = \boldsymbol{\lambda}$. Then, the Riesz representation theorem implies $\boldsymbol{\lambda}_0$ is differentiable *as a parameter on the model*[7] if and only if there exists some function $\widetilde{\boldsymbol{\psi}}(\mathbf{x},\mathbf{y})$, the *efficient influence function*, such that

$$\boldsymbol{\lambda} = \mathbb{E}[\widetilde{\boldsymbol{\psi}}(\mathbf{x},\mathbf{y})(\mathbf{a}'\dot{\boldsymbol{\ell}}(\mathbf{x},\mathbf{y}) + \eta(\mathbf{x},\mathbf{y}))].$$

Setting $\boldsymbol{\lambda}$ to zero, it suffices to find $\widetilde{\boldsymbol{\psi}}(\mathbf{x},\mathbf{y})$ such that $\mathbb{E}[\widetilde{\boldsymbol{\psi}}(\mathbf{x},\mathbf{y})\eta(\mathbf{x},\mathbf{y})] = \mathbf{0}$ and $\mathbb{E}[\widetilde{\boldsymbol{\psi}}(\mathbf{x},\mathbf{y})\frac{\partial}{\partial\boldsymbol{\theta}'}p_y(\mathbf{x},\mathbf{y};\boldsymbol{\theta}_0)] = \mathbf{0}$. Lemma A.1.7 in Appendix A.1.3 finds such a function $\widetilde{\boldsymbol{\psi}}(\mathbf{x},\mathbf{y})$, showing that the efficient information matrix and efficient influence function are

$$\widetilde{\mathbf{I}} = \boldsymbol{\Omega}/4 \text{ and } \widetilde{\boldsymbol{\psi}}(\mathbf{x},\mathbf{y}) = -\frac{1}{2}\widetilde{\mathbf{I}}^{-1}\mathbf{E}(\mathbf{D}'\mathbf{D})\mathbf{H}'\boldsymbol{\Sigma}^{-1}\mathbf{M}\mathbf{g}(\mathbf{x},\mathbf{y},\boldsymbol{\theta}^*).$$

[7] See page 363 of van der Vaart (1998) for a more detailed discussion.

The efficient information matrix reflects the information loss from ignorance of the density $p_{xy}(\mathbf{x}, \mathbf{y})$ and parameter $\boldsymbol{\theta}^*$. The inverse of the efficient information matrix is an informational bound on $\boldsymbol{\lambda}_0$ which is exactly the covariance of the limit experiment $\mathbf{Z}$ described in Equation 1.4.2.

1.5 Simulations

In this section, I show that the moment-based testing procedure has good finite sample properties in a panel AR(2) like that considered in Section 1.2.1. While likelihood-based tests have higher power when the distribution of shocks is correctly-specified, I show that the likelihood-based $C(\alpha)$ test over-rejects when the likelihood function is incorrectly specified. Furthermore, the simulations illustrate that neglecting parameter heterogeneity can lead to severe under-coverage of the mean autoregressive coefficient under conventional confidence intervals.

Setup. I consider a stationary random coefficients panel AR(2) with $T = 10$ periods and 6000 observations

$$y_{it} = \phi_{1i} y_{it-1} + \phi_{2i} y_{it-2} + \varepsilon_{it}, \quad \phi_{1i} \sim \mathcal{N}(.5, \delta^2), \quad \phi_{2i} \sim \mathcal{N}(.3, \delta^2). \tag{1.5.1}$$

Parameter heterogeneity is determined by the parameter δ, which ranges from 0 to 0.16. Mean zero and unit variance shocks ε_{it} are independent over units i and are white noise over time t.

Motivated by features of the data in macroeconomic applications, I consider shocks ε_{it} that are (1) i.i.d. Gaussian, (2) i.i.d. Skew-t with 8 degrees of freedom and shape parameter of -0.5 (Hansen, 1994), and (3) ARCH(1) (with ARCH parameter 0.4) shocks. Shocks of the first type serve as a benchmark. Shocks of the second type are skewed and have tails that are heavier than those of a normal distribution. Shocks of the third type are conditionally heteroskedastic, mimicking features found in time series datasets. I consider a two-step

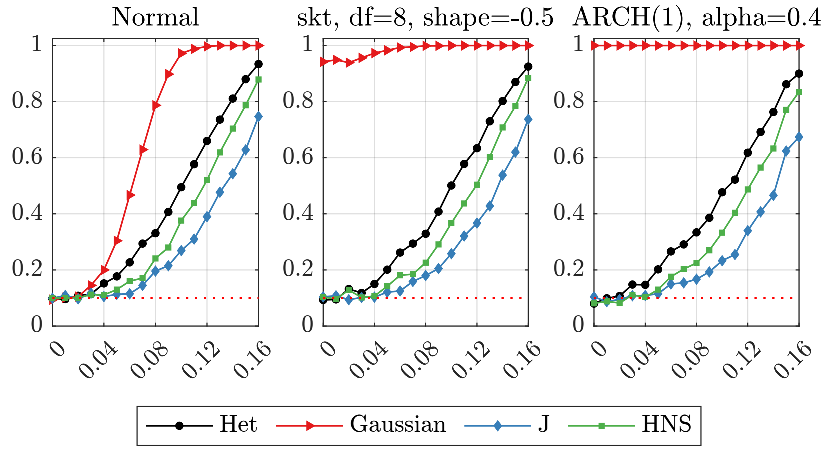

Figure 1.5.1: Empirical rejection probabilities by test and shock distribution. The horizontal axis gives the standard deviation of the random coefficient δ and the vertical axis gives the empirical rejection probability. "Het" gives this paper's moment-based heterogeneity test, "Gaussian" gives a Gaussian likelihood $C(\alpha)$ test, "J" gives a GMM Sargan-Hansen J-test, and HNS gives the Hahn et al. (2014) test.

efficient GMM estimator (that neglects parameter heterogeneity) whose moments are formed by matching autocovariances of $\mathbb{E}[\Delta y_{it} \Delta y_{it-\ell}]$. Just as in the illustration, I consider the model in first differences to mimic dynamic panel applications that use first differences to eliminate unit-specific fixed effects.

I compare this paper's test to three alternative testing procedures. The first test is the $C(\alpha)$ test of Gu (2016). For comparability, the likelihood-based test is based on the stationary panel AR(2) of Equation 1.5.1 in first differences under the assumption of homogeneous autoregressive coefficients and i.i.d. Gaussian shocks. The test represents a researcher who is interested in testing for parameter heterogeneity where shocks are potentially non-Gaussian, but proceeds anyway with a Gaussian likelihood-based test. The second test is a standard J-test of overidentifying restrictions, and represents a textbook procedure for model misspecification. The third test is the heterogeneity test of Hahn et al. (2014) (HNS), formed as a GMM Lagrange Multiplier test with a chi-squared distribution under the null. HNS serves as a comparison for the value of considering an unconventional test statistic over a two-sided test that doesn't exploit the structure of the alternative hypothesis parameter space.

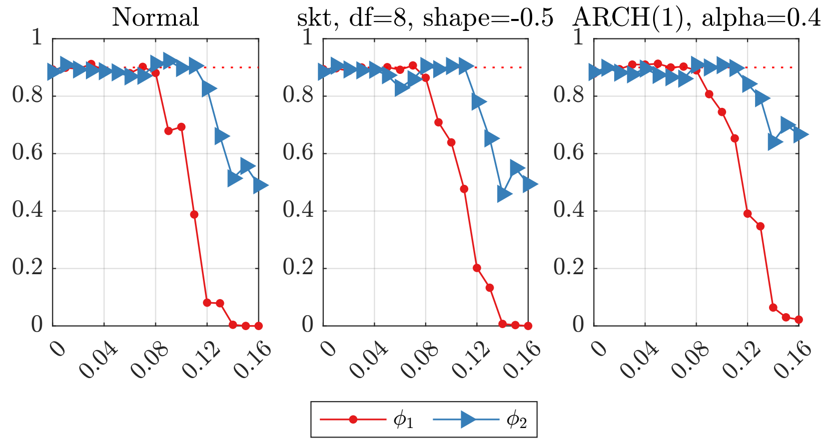

Figure 1.5.2: Empirical coverage by parameter and shock type. The horizontal axis gives the standard deviation of the random coefficient δ and the vertical axis gives the empirical coverage for 95% confidence intervals.

From Figure 1.5.1, all four tests control size under Gaussian shocks. Exploiting deviations in higher-order moments (like kurtosis), the Gaussian $C(\alpha)$ test is also more powerful across δ. Unlike the GMM tests, size control however is lost for the $C(\alpha)$ test under skew-t and ARCH(1) shocks—the same sensitivity to higher order moments that give rise to power gains in the Gaussian case make the test vulnerable to misspecification of the shock distribution. The three GMM tests control size under non-Gaussian shocks, with the J giving the lowest power across values of δ. The J test's power is spread across many possible alternatives, not just parameter heterogeneity. Finally, the power gains of my proposed test over HNS suggest the value of taking the positive semidefinite constraint into account.

Consistent with the theoretical results on the asymptotic bias of GMM estimators described in Section 1.3, Figure 1.5.2 shows that parameter heterogeneity affects the estimator's finite sample properties. Under the three choices of shocks, the estimator's empirical coverage for pointwise conventional confidence intervals is near their nominal coverage for small amounts of parameter heterogeneity. As parameter heterogeneity increases (particularly above $\delta = 0.08$), the empirical confidence intervals potentially severely under-cover. These results illustrate that neglecting parameter heterogeneity can potentially lead to severely distorted inference when the parameter of interest is the mean autoregressive parameter.

1.6 Empirical applications

I demonstrate the applied utility of my test with two applications. First, I show how the proposed test can be used as a diagnostic for determining the appropriate level of aggregation in characterizing income dynamics. Second, I show that neglecting plant-level heterogeneity in the output elasticity of variable inputs affects estimates of the distribution of markups.

1.6.1 Income dynamics

Under incomplete markets, earnings risk is a critical ingredient in a large range of economic questions—including understanding consumption/wealth inequality (Heathcote et al., 2010), designing counterfactuals for tax policy (Conesa et al., 2009), and the determination of asset prices (Constantinides and Duffie, 1996). Researchers have recognized the need to allow for heterogeneity across some dimension, like education.

In this section, I show that unit-level heterogeneity in the persistence of income shocks is detectable even in a standard survey dataset. Across different aggregations of educational subgroups, I show how the proposed test can be used as a guide for choosing appropriate sample splits for characterizing income dynamics by educational groups.

Data. Following Guvenen (2009), my sample is the Panel Survey of Income Dynamics, including household heads that worked between 520-5110 hours per year, with average hourly earnings between \$2 and \$400 (in 1992 terms), and excluding the poverty sample. y_{it} gives labor income excluding wages, bonuses, commissions, and the labor portion of farm/business income.

Estimation. I consider a model for income dynamics with income profile heterogeneity

$$y_{it} = \delta_i + \delta_t + \beta_i h_{it} + \eta_{it} + z_{it}, \quad z_{it} = \rho z_{it-1} + \varepsilon_{it}$$

for transitory shock η_{it}, autoregressive income shock z_{it}, unit fixed effect δ_i, and time fixed effect δ_t. h_{it} is tenure (proxied by age) where β_i is a random coefficient that varies by individual with variance $\text{Var}[\beta_i^2] = \sigma_\beta^2$. Mean zero shocks η_{it} and ε_{it} are independent across units, are jointly independent, and white noise over time. The white noise condition allows for conditional heteroskedasticity (like ARCH dynamics). Unlike likelihood-based tests, I do not need to commit to a specific model for modeling heteroskedasticity or to the shock's specific distribution. Averaging over units, let $\mathbb{E}[\eta_{it}^2] = \sigma_\eta^2$ and $\mathbb{E}[\varepsilon_{it}^2] = \sigma_\varepsilon^2$.

Similar to Guvenen (2009), I estimate parameters $(\rho, \sigma_\beta^2, \sigma_\eta^2, \sigma_\varepsilon^2)'$ by matching empirical and model-implied autocovariances. For $\widetilde{y}_{it} = y_{it} - \frac{1}{n}\sum_{i=1}^n y_{it}$, I consider the lag 0 through lag 11 autocovariances of $\Delta\widetilde{y}_{it} = \widetilde{y}_{it} - \widetilde{y}_{it-1}$:

$$\mathbb{E}[(\Delta\widetilde{y}_{it})^2] = \sigma_\beta^2 + 2\sigma_\eta^2 + 2\frac{\sigma_\varepsilon^2}{1+\rho}$$
$$\mathbb{E}[(\Delta\widetilde{y}_{it})(\Delta\widetilde{y}_{it-\ell})] = \sigma_\beta^2 - 1[\ell = 1]\sigma_\eta^2 + \rho^\ell \frac{\sigma_\varepsilon^2(\rho - 1)}{\rho + 1}, \quad \ell \geq 1.$$

For the $m = 12$ moments, the $p = 4$ model parameters are estimated using two-step efficient GMM. To coarsely capture time variation in parameters, I estimate separately on the 1968-1980 and 1980-1992 subsamples.[8]

RESULTS. I test whether parameter ρ is homogeneous across households. The parameter is important for macroeconomic models with households, as it is a crucial determinant for a household's marginal propensity to consume. Motivated by the education subsamples often used in the literature on income dynamics (Hubbard et al., 1994; Carroll and Samwick, 1997; Guvenen, 2007), I illustrate how the test for parameter heterogeneity can be used as a diagnostic for determining the appropriate level of aggregation.

Including all households, Table 1.6.1 shows that the estimated persistence in both subperiods is around 0.7. Moreover, the J test fails to reject the null hypothesis of correct specification at 5%. However, the null hypothesis of homogeneity across households is rejected at 5% for both this paper's test and the test of Hahn et al. (2014). These results indicate that a researcher interested in testing for parameter heterogeneity using only the J test would

[8] Extending the analyses to richer administrative data, like that of Guvenen et al. (2021), with more granular sample splits and alternative estimators would be a valuable area of future work.

	Spec. tests						
	p	p_{HNS}	p_J	ρ	SE	n	$\widehat{\text{Var}}(\rho_i)$
<u>1968–1980</u>							
All	0.01	0.02	0.08	0.68	0.06	809	0.18
Some college or less	0.01	0.02	0.06	0.55	0.06	599	0.25
HS or less	0.22	0.44	0.36	0.54	0.08	309	–
Some college	0.72	0.54	0.42	0.87	0.13	173	–
Bachelor's or more	0.09	0.17	0.71	0.95	0.25	167	–
<u>1980–1992</u>							
All	0.02	0.04	0.40	0.74	0.05	1022	0.09
Some college or less	0.03	0.07	0.31	0.73	0.06	652	0.10
HS or less	0.18	0.36	0.68	0.73	0.09	262	–
Some college	0.75	0.49	0.31	0.79	0.11	207	–
Bachelor's or more	0.20	0.41	0.34	0.76	0.08	302	–

Table 1.6.1: Estimation results for the model of income dynamics by education group. The first three columns give p-values for this paper's heterogeneity test, the Hahn et al. (2014) procedure, and the J test respectively. The last three columns give point estimates for persistence ρ (under the assumption of within-group parameter homogeneity), the associated standard error, and sample size n. The final column gives an estimate for the variance of the random coefficient.

potentially miss it in this context. Rescaling this paper's heterogeneity test statistic (see Remark 1 of Section 1.3.4) gives a noisy estimate of parameter heterogeneity—the variance of the persistence parameter across units is estimated to be 0.18 (with a 90% confidence interval of $[0.05, 0.31]$) for the 1968-1980 sample versus 0.09 for the 1980-1992 sample (with a 90% confidence interval of $[0.02, 0.16]$). Thus, more granular subsamples are needed.

Next, I split the sample into the "Bachelor's or more" and "Some college or less" categories. Here, my test rejects the null hypothesis of parameter homogeneity for the "some college or less" category, while fails to reject for the "Bachelor's or more" category at 5%. These results suggest that further splits for the "Some college or less" category are needed. In contrast, the Hahn et al. (2014) test fails to reject parameter homogeneity for the "Some college or less" category in the 1980-1992 subperiod, consistent with the gain in power from considering a one-sided versus a two-sided test. Finally, my test fails to reject parameter homogeneity after splitting the "Some college or less category" into "HS or less" and "Some college." The assumption of parameter homogeneity appears to approximately hold for these more granular education groups.

Summarizing, estimating a single model of income dynamics for all households masks education group-level heterogeneity in the persistence of income shocks—particularly in the earlier 1968–1980 sample. I show how this paper's test for parameter heterogeneity can be used as a data-driven diagnostic for determining sample splits.[9]

1.6.2 Markup heterogeneity

Competition between firms combined with the possibility of new entrants ensures that firms set prices that reflect marginal costs. Without competition, firms gain market power and are able to set higher prices—ultimately affecting consumer welfare, discouraging innovation, and decreasing the investment in capital (De Loecker et al., 2020). However, despite the importance of understanding market power, there are substantial challenges in estimating markups over marginal cost.

In this subsection, I give practical guidance for estimating output elasticities, one of the main challenges in estimating firm-level markups. Estimation requires the assumption of homogeneity of the output elasticity of inputs at the industry level. However, even within granular industry groups, there is reason to expect these elasticities to differ across plants (e.g. plants that produce dissimilar products with distinct input mixes, plants that produce similar products using different production processes). Incorrectly imposing parameter homogeneity gives rise to unreliable estimates of these output elasticities and consequently unreliable estimates of markups. The level and dispersion of markups distort allocations ultimately affecting welfare (Hsieh and Klenow, 2009; Edmond et al., 2023).

While researchers have recognized the importance of aggregation in estimation (see Section 2.2 of Fernald et al. (2025) for a review), I present evidence of unobserved plant-level heterogeneity in production functions even at the (most granular) four-digit industry level. I

[9] Like the usage of other specification tests in applied work, multiple testing (arising from comparing p values across subsamples) is also a concern in this application (see the discussion in Chapter 9 of Lehmann and Romano (2022)).

then estimate the distribution of markups for subindustries where the moment conditions are compatible with the data.

Data. My dataset is the Chilean annual manufacturing census managed by Encuesta Nacional Industrial Anual (ENIA). The survey is a common benchmark in the production function estimation literature (Liu, 1993; Levinsohn and Petrin, 2003; Gandhi et al., 2020). The census' coverage is extensive, covering all Chilean manufacturing plants with at least 10 employees. I focus on the ten year sample from 1987-1996, giving approximately 5000 plants per year. Each plant is associated with a four-digit International Standard Industrial Classification Revision 2 code (ISIC4).

Estimation. To estimate markups, I consider the production approach (Hall et al., 1986; Hall, 1988, 1990; De Loecker and Warzynski, 2012). The main assumption is that firms are cost-minimizing and are price takers in the input market. This approach to markup estimation is attractive since the assumption of constant returns to scale needn't be imposed and the user cost of capital needn't be observed. Firms set marginal products equal to the factor prices. Their first order condition implies that the markup of firm i equals the output elasticity of materials β^m_{it} divided by the revenue share of materials

$$\mu_{it} = \frac{\beta^m_{it}}{s^m_{it}}. \tag{1.6.1}$$

While s^m_{it} is taken from the data, β^m_{it} must be estimated.

I consider a Cobb-Douglas production function. For plant i and time t, I observe output y_{it}, materials m_{it}, capital k_{it}, and labor ℓ_{it}. Plants are subject to unobserved productivity shocks ω_{it} that follow an AR(1) process with productivity innovation ζ_{it}. In addition, the production function is subject to measurement error ε_{it}. In logs, the production function for a plant in industry j is

$$y_{it} = \beta^k_j k_{it} + \beta^\ell_j \ell_{it} + \beta^m_j m_{it} + \omega_{it} + \varepsilon_{it} \text{ and } \omega_{it} = \rho_{j0} + \rho_{j1}\omega_{it-1} + \zeta_{it}. \tag{1.6.2}$$

The elasticity β_j^m is used to estimate markups according to Equation 1.6.1. With the short estimation window of ten years, the production function coefficients are taken to be constant over time. While the production function coefficients will be taken to vary across 4-digit subindustries, we are interested in testing if there is remaining heterogeneity within subindustries.

Matching Raval (2023) and De Loecker et al. (2020), I estimate the production function parameters using the control function approach of Ackerberg et al. (2015). These more sophisticated approaches are required because the productivity process is partially observed by firms that choose their own inputs. As a result, ordinary least squares estimation of the production function elasticities is biased. For brevity, I suppress the industry subscripts j. There are two sets of moment conditions that are valid absent plant-level production function heterogeneity. For the first set, the assumptions of Ackerberg et al. (2015) imply that output can be decomposed as

$$y_{it} = \phi_t(\ell_{it}, m_{it}, k_{it}) + \varepsilon_{it} \tag{1.6.3}$$

where the measurement error ε_{it} is uncorrelated with some flexible function $\phi_t(\ell_{it}, m_{it}, k_{it})$. Consistent with applied practice, I take $\phi_t(\ell_{it}, m_{it}, k_{it})$ to be a second-order polynomial in the inputs.[10] For the second set, Equations 1.6.2 and 1.6.3 imply that the productivity shock is $\omega_{it} = \phi_t(\ell_{it}, m_{it}, k_{it}) - \beta^\ell \ell_{it} - \beta^m m_{it} - \beta^k k_{it}$. After substituting, the innovation to the productivity process is $\zeta_{it} = \omega_{it} - \rho_0 - \rho_1 \omega_{it-1} = [\phi_t(\ell_{it}, m_{it}, k_{it}) - \beta^\ell \ell_{it} - \beta^m m_{it} - \beta^k k_{it}] - \rho_0 - \rho_1[\phi_{t-1}(\ell_{it-1}, m_{it-1}, k_{it-1}) - \beta^\ell \ell_{it-1} - \beta^m m_{it-1} - \beta^k k_{it-1}]$. Then, the second set of moment conditions exploit the uncorrelatedness of the unobserved productivity innovation ζ_{it} with the lagged inputs and contemporaneous capital:

$$0 = \mathbb{E}\left[\zeta_{it}(1,\, k_{it},\, \ell_{it-1},\, m_{it-1},\, \phi_{t-1}(\ell_{it-1}, m_{it-1}, k_{it-1}),\, \beta^k k_{it-1},\, \beta^\ell \ell_{it-2},\, \beta^m m_{it-2})\right]. \tag{1.6.4}$$

[10] ϕ_t contains an additive time fixed effect where remaining polynomial coefficients are time-invariant.

The first five moments in Equation 1.6.4 are similar to those featured in Equation 28 of Ackerberg et al. (2015). Capital enters the moment condition contemporaneously since it is inflexible, while labor and materials enter with lags since they are flexible. The final three are lagged contributions of capital, labor, and materials to output and are included to provide over-identifying restrictions. These moments' nonlinearity in β^{ℓ}, β^{m}, and β^{k} allow for this paper's heterogeneity test to be applied. Pooling observations over time, the model is estimated using the generalized method of moments with the efficient choice of weight matrix (consistent with the discussion of Wooldridge (2009)) for each 4-digit ISIC category (ISIC4).

Results. Ignoring within–subindustry heterogeneity in production function coefficients gives rise to misspecified markup estimates that are difficult to characterize. From Equation 1.6.4, markup estimates are directly related to the output elasticity. Thus, if the output elasticities were incorrectly taken to be homogeneous (even at the mean), the effect on the estimated distribution of markups is ambiguous, ultimately depending on the joint distribution of the output elasticity and the shares s_{it}^{m}. Unfortunately, this joint distribution is difficult to characterize absent additional information.

Consequently, the results to follow focus on using specification tests to guide which subindustries fit the data well. Specifically, I demonstrate how the J-test and heterogeneity tests can be used as diagnostics for determining which granular 4-digit (ISIC) categories are compatible with the Ackerberg et al. (2015) moment conditions. The results are subject to two caveats. First, since the null hypothesis of both tests is of correct specification, failure to reject is an indication of the absence of evidence for rejection and not of proof of correct specification. Second, like Section 1.6.1 and in the use of specification tests in applied work, multiple testing arises here too.

Starting with the J-test, the first column of Table 1.6.2 lists the ISIC4 categories that reject the J test at 5% (see Tables A.2.1 and A.2.2 found in Appendix A.2.1 for the full results). There are several types. First, the column includes categories with plants that have complex production processes (like ship-building/repairing and the manufacture of motor

Misspecified	Heterogeneous	Well-specified
1. Grain mill products	1. Wine industries	1. Slaughtering, preparing and preserving meat
2. Manufacture of bakery products	2. Manufacture of made-up textile goods except wearing apparel	2. Manufacture of dairy products
3. Manufacture of food products not elsewhere classified	3. Manufacture of carpets and rugs	3. Canning and preserving of fruits and vegetables
4. Malt liquors and malt	4. Sawmills, planing and other wood mills	4. Canning, preserving and processing of fish, crustaces and similar foods
5. Soft drinks and carbonated waters industries	5. Manufacture of wooden and cane containers and small cane ware	5. Manufacture of vegetable and animal oils and fats
6. Spinning, weaving and finishing textiles	6. Manufacture of wood and cork products not elsewhere classified	6. Manufacture of cocoa, chocolate and sugar confectionery
7. Knitting mills	7. Manufacture of pulp, paper and paperboard	7. Manufacture of prepared animal feeds
8. Manufacture of wearing apparel, except footwear	8. Manufacture of containers and boxes of paper and paperboard	8. Distilling, rectifying and blending spirits
9. Manufacture of footwear, except vulcanized or moulded rubber or plastic footwear	9. Printing, publishing and allied industries	9. Manufacture of furniture and fixtures, except primarily of metal
10. Manufacture of pulp, paper and paperboard articles not elsewhere classified	10. Manufacture of soap and cleaning preparations, perfumes, cosmetics and other toilet preparations	10. Manufacture of drugs and medicines
11. Manufacture of paints, varnishes and laquers	11. Manufacture of non-metallic mineral products not elsewhere classified	11. Manufacture of chemical products not elsewhere classified
12. Manufacture of plastic products not elsewhere classified	12. Manufacture of cutlery, hand tools and general hardware	12. Organic composite solvents and thinners, not elsewhere specified or included; prepared paint or varnish removers
13. Manufacture of structural clay products	13. Reaction initiators, reaction accelerators and catalytic preparations n.e.c. or included	13. Manufacture of special industrial machinery and equipment except metal and wood working machinery
14. Manufacture of cement, lime and plaster	14. Manufacture of agricultural machinery and equipment	
15. Manufacture of sports goods	15. Manufacture of motorcycles and bicycles	
16. Manufacture of furniture and fixtures primarily of metal		
17. Manufacture of structural metal products		
18. Manufacture of fabricated metal products except machinery and equipment not elsewhere classified		
19. Manufacture of metal and wood working machinery		
20. Machinery and equipment except electrical not elsewhere classified		
21. Ship building and repairing		
22. Manufacture of motor vehicles		

Table 1.6.2: Summary of specification tests (5% significance level). "Misspecified" lists ISIC4 categories that reject the J test. "Heterogeneous" lists ISIC4 categories that fail to reject the J test but reject the heterogeneity test. "Well-specified" lists ISIC4 categories that fail to reject both the J test and heterogeneity test.

vehicles) that potentially don't satisfy the requirement of input flexibility—like through unions[11] or material inputs contracted in advance. Second, the list includes categories with plants that face market power in inputs due to high transportation costs (like limestone in cement manufacturing) (Beach et al., 2000). These plants rely on a small number of local suppliers, creating exposure to local supply constraints and inflexibility arising from market power. Third, the list includes catch-all categories (like manufacture of food products not elsewhere classified) that include dissimilar plants.

Next, the second column of Table 1.6.2 lists the ISIC4 categories that fail to reject the J test at 5% but reject the heterogeneity test at 5%. This group can be interpreted as categories whose moments are consistent with plant-level heterogeneity in production functions but are otherwise well-specified. Notably, Tables A.2.1 and A.2.2 show that none of the fifteen categories reject the Hahn et al. (2014) heterogeneity test at 5%, consistent with the power gains of this paper's heterogeneity test from considering narrower alternatives. The column includes broad categories with plants that manufacture dissimilar products that require distinct input mixes (like motorcycles/bicycles and agricultural machinery/equipment). The list also contains industries that are related to lumber and paper products, consistent with the substantial structural changes in these industries in Chile during the sampling period. The late 1980s and 1990s were marked by a sharp increase in foreign investment, leading to increased mechanization and automation throughout the production process (Klubock, 2014). The heterogeneity test is potentially capturing heterogeneity arising from plants with capital-intensive processes supported by foreign investment and pre-existing labor-intensive ones.

The final column of Table 1.6.2 lists the ISIC4 categories that fail to reject both the J test and the heterogeneity test at 5%. These industries can be interpreted as those that are compatible with the Ackerberg et al. (2015) moment conditions (and consistent with the requirement

[11]Union rights were reinstated by the Pinochet regime in 1979, though in a weaker form relative to the 1960s and the early 1970s. Nonetheless, strikes still occurred (3.5 per 100,000 workers in 1996 versus 9.9 in the 1960s) suggestive of labor rigidities in industries with firm-level unionization. See Table 4 of Edwards et al. (2000) for details on strike activity and the text for a detailed discussion of the historical context.

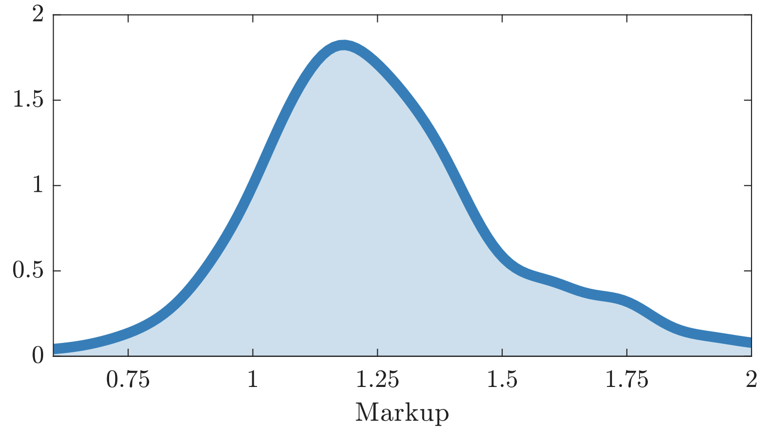

Figure 1.6.1: Estimated sales-weighted distribution of plant-level markups μ_{it} for plants belonging to categories listed in Column 3 of Table 1.6.2.

of within-ISIC4 category production function homogeneity). This list is primarily composed of industries that are related to food products (like slaughtering/preparing/preserving meat, manufacture of dairy products, and canning/preserving of fruits/vegetables). In the period under study, plants in this category have consistent manufacturing standards and competitive input markets, resulting in goods with similar input mixes. These results are consistent with the historical development of the food industry in Chile. By the early 1990s, three-quarters of fresh milk were produced by industrial dairy establishments (Llorca-Jaña et al., 2020).

Guided by Table 1.6.2, Figure 1.6.1 gives the sales-weighted distribution of plant-level markups for the subset of plants that belong to industries that fail to reject both the J test and the heterogeneity test at 5%. The median markup for plants belonging to these industries, weighted by sales, is 1.24 with a 90 minus 50-percentile markup dispersion of 0.42. These values are comparable to the full distribution of plant-level markups under the restrictive assumption of no sample selection—that the distribution of markups for plants belonging to "well-specified" industries are no different than the distribution of markups for plants in the full sample.[12]

[12]For the goal of estimating the full distribution of plant-level markups, additional steps include considering (1) alternative approaches to production function estimation for plants belonging to the "misspecified" category and (2) splitting the subindustries belonging to the "heterogeneous" category by additional observables (like size). These steps are left for future work.

As a counterfactual, now suppose that the 4-digit subindustries categorized as "well-specified" were indeed correct and suppose that the researcher estimated production functions using the coarser 3-digit subindustries instead. Using these production function coefficients to compute markups for the plants belonging to the same four-digit subindustries, the median markup is now 1.37 with a 90 minus 50 percentile markup dispersion of 0.32. Thus using the incorrect level of aggregation overstates the median markup while understating the dispersion. Viewed as welfare costs (as in the model of Edmond et al. (2023) for example), the researcher would overstate the "aggregate markup as uniform output tax" channel but understate the "misallocation of factors of production" channel.

Testing for omitted parameter heterogeneity can be applied to other specifications used in the literature on markup estimation. Alternative specifications include computing markups using labor rather than materials in Equation 1.6.1, using a translog production function instead of a Cobb-Douglas one, using other unconditional moments for the estimation procedure of Ackerberg et al. (2015) (like writing the second stage moments using $\zeta_{it} + \varepsilon_{it}$ rather than ζ_{it} alone), other data sources, and other estimators (like the dynamic panel estimator of Blundell and Bond (2000) or the estimator of Gandhi et al. (2020)). See Fernald et al. (2025) for a comprehensive review.

1.7 Conclusion

This paper proposes a test of over-identifying restrictions with power directed toward detecting moderate amounts of parameter heterogeneity. The test is particularly useful for contexts where economic theory gives moment conditions that are valid under a representative agent, but is uninformative of the particular distribution of the shocks—a common feature of dynamic panel models. Adapting these moment conditions to allow for parameter heterogeneity is often difficult. What's worse, incorrectly imposing parameter homogeneity gives rise to parameter estimates that are difficult to interpret and fail to converge to the average parameter. Here, the test can be used as an initial data-driven diagnostic before splitting a sample by observables or proceeding to sophisticated methodologies requiring additional statistical assumptions.

Exploiting Jensen's inequality, the test statistic takes the form of a second-derivative weighted average of the normalized, fitted moment conditions. The test achieves power gains by focusing on alternatives compatible with parameter heterogeneity, a form of testing on the boundary of the parameter space. After normalizing, the test statistic can also be interpreted as the covariance of the random coefficients. I prove that the test asymptotically maximizes a weighted average power criterion, and for the special case of testing for scalar heterogeneity, the test is the asymptotically uniformly most powerful test. Simulations show power gains of the proposed test relative to other semiparametric tests, and that incorrectly specifying the likelihood function for a likelihood-based heterogeneity test can lead to severe over-rejection.

Applied to a model of income dynamics, I use the test as a data-driven diagnostic for determining appropriate sample splits and find evidence of heterogeneity in the persistence of income shocks by education group. In addition, I construct estimates of plant-level markups, which requires the assumption of the homogeneity of production function elasticities across plants within an industry. Using the proposed test, I find evidence of plant-level heterogeneity in production function elasticities *within* granular 4-digit industry categories. I then use the test as a guide for constructing markup estimates for plants belonging to the subset of industries that fit the representative firm model.

Chapter 2

Robust Granular Instrumental Variables*

2.1 Introduction

In many macroeconomic settings, researchers wish to estimate the spillover of an idiosyncratic shock to one unit onto other units. Examples include estimating the contagion effects of financial market distress, aggregate demand externalities on individual consumption, strategic complementarities in price setting, and the price elasticity of demand of the stock market (Allen et al., 2009; Auclert et al., 2023; Alvarez et al., 2023; Gabaix and Koijen, 2021). Identifying plausibly exogenous variation in these settings however is notoriously difficult and contributes to the continued challenges of conducting empirical work.

Gabaix and Koijen (2024) tackles this difficulty with a novel technique, called granular instrumental variables (GIV). They consider environments where an individual unit's outcome is partially determined by the size-weighted outcome. Thus, idiosyncratic shocks to a single unit spill over to all other units in equilibrium. To overcome the resulting endogeneity bias,

*This paper has been presented at the 2024 NBER-NSF Time Series Conference. A revision has been requested at the Journal of Applied Econometrics. The latest public working paper version of the paper has been reproduced here.

their instrument for the size-weighted outcome is constructed as the difference between size- and equal-weighted outcomes. Their instrument is "granular" in that idiosyncratic shocks to large units are the primary source of identifying variation as they disproportionately contribute to movements in the size-weighted outcome variable. The broad adoption of GIV in applied macroeconomics and finance is indicative of its usefulness (Chodorow-Reich et al., 2021; Adrian et al., 2022; Camanho et al., 2022; Gabaix and Koijen, 2021).

The baseline GIV estimator of Gabaix and Koijen (2024), however, requires strong assumptions to satisfy the instrumental variables relevance condition and exclusion restriction and its extensions require further conditions. The baseline estimator requires homogeneous spillovers across units, homogeneous shock variances, skewed unit size, and idiosyncratic shocks (uncorrelated shocks between units after accounting for a *known* factor structure). In particular, there is typically no ex-ante rationale for there to be homogeneous spillovers and shock variances across units, contributing to the gap between theory and practice. Furthermore, I show that there is no general guarantee that the GIV spillover estimand admits a positive-weighted average of unit-specific spillovers. To address these concerns, Gabaix and Koijen (2024) generalize their procedure to allow for heterogeneity that is a function of observables, but the empirically important question of how to best tackle unobserved heterogeneity remains open.

This paper's main contribution is to establish global identification for a GIV model with unit-specific spillovers and unknown shock variances. I propose an estimator, called robust granular instrumental variables (RGIV), that is robust in the sense that it is applicable to a wider set of environments than the baseline Gabaix and Koijen (2024) GIV estimator. Ultimately, RGIV brings the study of unit-level heterogeneity—a feature that is of substantial importance in other areas of macroeconomics[2]—to the estimation of spillovers.

RGIV uses internally estimated idiosyncratic shocks as instruments for the size-weighted outcome variable. Informally, the procedure can be described sequentially. An initial guess of

[2]For instance, take household-level differences in the marginal propensity to consume akin to Fagereng et al. (2021); Lewis et al. (2026); Fuster et al. (2021).

the first unit's spillover gives an estimate of the first unit's idiosyncratic shock. Using this estimated shock as an instrument for the size-weighted outcome variable, the remaining units' spillover coefficients and estimated idiosyncratic shocks can be computed. If one or more pairs of estimated idiosyncratic shocks are correlated, we can guess a new spillover for the first unit and repeat the procedure until the estimated idiosyncratic shocks are uncorrelated. RGIV makes use of granularity by exploiting the contribution of individual shocks (even to relatively small units) to movements in the size-weighted outcome variable. This approach differs from the baseline GIV estimator in that skewness in the size distribution of units and homogeneity in shock variances are not required.

I prove that there is an inevitable trade-off in assumptions between allowing for general spillover heterogeneity and a general shock covariance structure. Like GIV, the RGIV estimator can easily accommodate cross-unit correlations that are due to observed covariates (both time-varying and unit-specific) or a known factor structure in the residuals. However, I prove that a model with an *unknown* factor structure and unrestricted spillover heterogeneity is not identified. Hence, absent additional structure, researchers must restrict either spillover heterogeneity or the correlation structure of the shocks.

I develop tests that evaluate the appropriateness of (1) the RGIV framework and (2) the homogeneous spillovers assumptions featured in Gabaix and Koijen (2024). These hypothesis tests are applications of standard GMM results (Newey and McFadden, 1994). The first test is the Sargan–Hansen test, where the null hypothesis of correct specification is rejected when one or more pairs of estimated idiosyncratic shocks are correlated. The second test is the distance metric test, where the null hypothesis is rejected when the constraint of spillover homogeneity binds.

Building on the analysis of a working paper version of Gabaix and Koijen (2024), I apply RGIV to sovereign yield spillovers in the Euro area and find strong evidence of spillover heterogeneity among Euro area members. I control for correlations induced by heterogeneous loadings on unobserved aggregate shocks by including a rich set of observed explanatory variables. In addition, I account for the well-documented correlation of shocks among Euro area countries by size-aggregating countries into larger "core" and "periphery" blocks (Bayoumi and Eichengreen, 1992). Under the preferred specification, I fail to reject the null hypothesis of correct specification. Meanwhile, the null hypothesis of spillover coefficient homogeneity is strongly rejected. I find that the size-weighted spillovers of countries in the western-periphery block (Portugal, Spain, and Ireland) are approximately twice that of the core country block.

In a simulation study, I also show that RGIV confidence intervals have good finite sample coverage properties under a DGP taken from the empirical application. GIV confidence intervals in contrast can undercover under spillover heterogeneity and when shock variances are taken to be unknown.

Literature My model generalizes the baseline environment of Gabaix and Koijen (2024) to allow for spillover heterogeneity, shock variance heterogeneity, and equal-sized units. For estimation, RGIV fully exploits second moment information through GMM and can also be adapted to the case in which correlations in shocks are induced by factors with known loadings. In contrast, the spillover coefficient homogeneity, skewed size, and known shock variance conditions are used to justify the validity of the GIV instrument. Gabaix and Koijen (2024) also propose extensions that separately estimate spillover coefficient heterogeneity[3] and unknown shock variances[4]. In contrast, RGIV doesn't require skewness of the size distribution, heterogeneity to depend on observables, or for shock variances to be known; uncorrelated shocks is sufficient for jointly allowing unit-level spillovers and unknown shock variances.

[3]Gabaix and Koijen (2024) Proposition 6 considers the case where unit-level spillover heterogeneity is determined by observables, Section D.9.1 proposes unit-specific "leave-one-out" granular instruments for when shock variances are known, and Section D.9.2 lists moment conditions for adapting GIV to heterogeneous spillovers though without the analysis of its econometric properties.

[4]Gabaix and Koijen (2024) Section D.3 appends shock variance moments after assuming homogeneous spillovers and global identification.

Baumeister and Hamilton (2023) applies the insights of GIV to show that a rich, structural model of the oil market (with unobserved explanatory variables and inventories) can be estimated using full information maximum likelihood. Their estimator similarly exploits shock orthogonality (in their case orthogonality of supply and demand shocks) and also allows for elasticity heterogeneity. In contrast, I extend the simple setting originally considered by Gabaix and Koijen (2024) to transparently illustrate the conditions needed for identification, for non-identification with unobserved explanatory variables, and to characterize the RGIV estimator's limit behavior. Rather than focusing on modeling assumptions that are specific to particular applications, this paper seeks to provide researchers with a parsimonious framework that could be adapted to their specific applications.

Banafti and Lee (2022) propose a refinement to the GIV estimator, allowing for unobserved explanatory variables but require the number of economic units to be large and the size distribution to be skewed, unlike RGIV. Here, the principal challenge is to prevent the law of large numbers from averaging away the granularity of units. The authors overcome this challenge by requiring the right tail of the unit size distribution to be highly skewed, with a Pareto tail index of less than 1. In contrast, RGIV is consistent even when the size distribution is uniform.

GIV procedures are closely linked to models found in the spatial panel and peer effects literature (Su et al., 2023; Aquaro et al., 2021; Chen et al., 2022; Manski, 1993). GIV's spillover network structure can be viewed as a restricted form of spatial autocorrelation. Unlike these papers, I allow the number of units to be small (finite), show that the GIV spillovers under unit-level heterogeneity are globally identified, and estimate spillovers without the use of external instruments. More broadly, my procedure can also be viewed as a special case of a simultaneous equation model with covariance restrictions, which is closely related to impact matrix identification in the structural VAR setting (Sims, 1980; Hausman, 1983).

Outline Section 2.2 illustrates the properties of GIV and RGIV in a simple three unit setting. Section 2.3 presents the main consistency and asymptotic normality results for the RGIV estimator. Section 2.4 describes the RGIV specification test and parameter homogeneity tests. Section 2.5 extends RGIV to include observable explanatory variables and discusses non-identification under unobserved explanatory variables. Section 2.6 applies the RGIV estimator to investigate sovereign yield spillovers in the Euro area. Section 2.7 examines finite-sample coverage accuracy of RGIV and GIV through simulations. Section 2.8 concludes.

2.2 Simple illustration

The remainder of this paper will consider the sovereign yield spillovers application as a running example. An idiosyncratic shock to one Euro area country raises yields in that country, as well as those of other Euro area countries since losses from default are partially shared. Country-specific spillovers are permitted.

In this section, I use a simple three-country setting to illustrate the construction and properties of GIV and RGIV. Section 2.3 formalizes the discussion and generalizes to n countries.

Notation Throughout this paper, I will follow the notation of Gabaix and Koijen (2024) for convenience. For a vector $X = (X_i)_{i=1,\ldots,n}$ and size S_i satisfying $\sum_{i=1}^n S_i = 1$, the equal-weighted sum X_E and size-weighted sum X_S are defined below:

$$X_E \equiv \frac{1}{n}\sum_{i=1}^n X_i, \quad X_S \equiv \sum_{i=1}^n S_i X_i.$$

Similarly, the equal-weighted and size-weighted cross-sectional averages for time series data $X_t = (X_{it})_{i=1,\ldots,n}$ are defined as $X_{Et} \equiv \frac{1}{n}\sum_{i=1}^n X_{it}$ and $X_{St} \equiv \sum_{i=1}^n S_i X_{it}$ respectively.

2.2.1 Baseline three country setting

I will outline my "baseline" setting for the three country case using sovereign yield spillovers in the Euro area as a running example. For countries $i = 1, 2, 3$, let y_{it} be the yield spread relative to some comparison country. Yield spread growth is $r_{it} = \frac{y_{it} - y_{i,t-1}}{y_{i,t-1}}$. Size S_i is observed and corresponds to a country's "debt at risk." Here, $S_i \in (0, 1)$ is taken to be time-invariant and sums to 1.

Suppose the researcher is interested in estimating elasticity ϕ_i, which will be called the "spillover coefficient." Specifically, yield spread growth r_{it} is determined by a country-specific spillover coefficient ϕ_i, size-aggregated yield spread r_{St}, and idiosyncratic shock u_{it}:

$$r_{it} = \phi_i r_{St} + u_{it}, \quad i = 1, ..., n, \quad \phi_S < 1. \tag{2.2.1}$$

The size-weighted spillover coefficient ϕ_S is taken to be less than 1. Moreover, shocks u_{it} are mutually and serially independent, mean zero, and have country-specific variance $\mathbb{E}(u_{it}^2) = \sigma_i^2 > 0$.

The propagation of an idiosyncratic shock depends on the size of the shock's origin country, the recipient country's spillover coefficient, and the size-weighted average spillover coefficient. To illustrate, consider a unit idiosyncratic shock to country 1 ($u_{1t} = 1$) holding the idiosyncratic shocks of countries 2 and 3 at zero ($u_{2t} = u_{3t} = 0$). Then, the size-weighted yield spread growth r_{St} increases by $S_1 \times \frac{1}{1-\phi_S}$, which is computed from taking a size-weighted average of Equation 2.2.1:

$$\sum_{i=1}^{3} S_i r_{it} = \sum_{i=1}^{3} S_i(\phi_i r_{St} + u_{it}) \implies r_{St} = \frac{u_{St}}{1 - \phi_S}. \tag{2.2.2}$$

The increase in r_{St} is larger when the size of country 1 is large and when countries are, on average, sensitive to spillovers (from multiplier $\frac{1}{1-\phi_S}$). Then, idiosyncratic shock u_{1t} spills over to r_{2t} and r_{3t}, giving rise to increases of $\phi_j r_{St} = \phi_j \frac{S_1}{1-\phi_S}$ for $j = 2, 3$. The total increase in r_{1t} is $1+\phi_1 \frac{S_1}{1-\phi_S}$, a composition of the direct effect of idiosyncratic shock u_{1t} and a spillover effect.

Granularity rules out the ordinary least squares regression of r_{it} on r_{St} as a method for estimating ϕ_i. Equation 2.2.2 highlights granularity in the baseline setting, showing that idiosyncratic shocks are responsible for movements in the aggregated yield spread r_{St} since $S_i > 0$ and $\sigma_i^2 > 0$. Therefore, regressing r_{it} on r_{St} to estimate ϕ_i as $T \to \infty$ suffers from endogeneity bias since $\mathbb{E}(r_{St}u_{1t}) = \frac{S_1\sigma_1^2}{1-\phi_S} \neq 0$. An alternative estimator is needed.

2.2.2 Illustration of GIV

In this section, I review the baseline GIV estimator of Gabaix and Koijen (2024) applied to this simple setting and discuss the conditions for its validity. I show that a skewed size distribution is necessary for the relevance condition to hold. Shock orthogonality and homogeneous idiosyncratic shock variances are necessary for the instrumental variables exclusion restriction to hold. Moreover, the exclusion restriction fails under heterogeneity of spillover coefficients across countries. As discussed later, there are GIV extensions that individually accommodate spillover coefficient heterogeneity and heterogeneous (and unknown) shock variances, but these require additional conditions.

Take the setting of Section 2.2.1 and further assume homogeneous spillover coefficients ($\phi_1 = \phi_2 = \phi_3 = \phi$), homogeneous shock variances ($\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma^2$), and skewed unit sizes (ruling out $S_1 = S_2 = S_3$). Just as in Section 2.2.1, granularity induces endogeneity bias in the regression of r_{it} on r_{St} for regression coefficient $\widehat{\phi}^i_{OLS}$

$$\widehat{\phi}^i_{OLS} - \phi \xrightarrow{p} \frac{\mathbb{E}(r_{St}r_{it})}{\mathbb{E}(r_{St}^2)} - \phi = \frac{S_i\sigma^2}{\mathbb{E}(r_{St}^2)[1-\phi]} > \phi.$$

Gabaix and Koijen (2024) propose estimating ϕ using instrumental variables. An equal-weighted aggregation of the country-level yield spread growth gives rise to the following IV regression model

$$r_{Et} = \phi r_{St} + u_{Et}.$$

The equal-weighted yield spread growth responds to the size-weighted yield spread growth r_{St} (through spillover coefficient ϕ) and equal-weighted idiosyncratic shocks u_{Et}. For the IV regression model in the above display, "granular instrument" z_t is constructed as the difference between the size- and equal-weighted yield spread growth

$$z_t := r_{St} - r_{Et} = \Big(\phi r_{St} + u_{St}\Big) - \Big(\phi r_{St} + u_{Et}\Big) = u_{St} - u_{Et}.$$

From the homogeneous spillovers assumption, the endogeneity from the size-weighted yield spread growth r_{St} is differenced away.

Skewness in country sizes is crucial for the instrumental variables relevance condition to hold and z_t reflects variation coming from idiosyncratic shocks to relatively large countries. To illustrate, first consider the extreme case of no skewness ($S_1 = S_2 = S_3$). Then the size- and equal-weighted yield spread growths are identical, producing a granular instrument that is identically zero $z_t = 0$. More generally, the instrumental variables relevance condition can be explicitly computed:

$$\begin{aligned}
\mathbb{E}\left[z_t r_{St}\right] &= \frac{1}{1-\phi}\mathbb{E}\left[u_{St}\left(u_{St} - u_{Et}\right)\right] \\
&= \frac{\sigma^2}{1-\phi}\left[S_1\left(S_1 - \frac{1}{3}\right) + S_2\left(S_2 - \frac{1}{3}\right) + S_3\left(S_3 - \frac{1}{3}\right)\right].
\end{aligned}$$

When the size of country 1 is large relative to other countries (when $S_1 \gg S_2, S_3$), the relevance condition is further from zero reflected by the $S_1(S_1 - 1/3)$ term.

The shock orthogonality and homogeneous (or known) shock variance assumptions are crucial for the exclusion restriction to hold. Explicitly, the covariance between the instrument and IV regression model error u_{Et} is

$$\begin{aligned}\mathbb{E}[u_{Et}(u_{St}-u_{Et})] &= \frac{1}{3}\Big[S_1\sigma^2+S_2\sigma^2+S_3\sigma^2\Big]-\frac{1}{3}\Big[\frac{1}{3}\sigma^2+\frac{1}{3}\sigma^2+\frac{1}{3}\sigma^2\Big] \\ &= \frac{1}{3}\sigma^2-\frac{1}{3}\sigma^2=0.\end{aligned}$$

In the first line of the above display, the uncorrelated shock assumption ensures that covariance terms $\mathbb{E}(u_{it}u_{jt})$ are zero. The homogeneous shock variance assumption ensures that the difference between the first line's bracketed terms is zero. Note that any instrument of the form $z_t' = W_1r_{1t}+W_2r_{2t}+W_3r_{3t}-\frac{1}{3}r_{1t}-\frac{1}{3}r_{2t}-\frac{1}{3}r_{3t}$ where $W_1+W_2+W_3=1$ satisfies the instrumental variables exclusion restriction. Gabaix and Koijen (2024) show that weighting by size gives a variance-minimizing estimator for the IV regression model. The spillover coefficient can also be consistently estimated when a time fixed effect is included, as the time fixed effect is differenced away in the construction of z_t.

The exclusion restriction argument described in the preceding paragraph can easily be extended to settings in which the shock variances are known to the econometrician rather than being homogeneous across units. Gabaix and Koijen (2024) show that identical computations hold after replacing the equal weights in $z_t = r_{St}-r_{Et}$ with inverse variance weights.

Spillover coefficient heterogeneity leads to a failure in the instrumental variables exclusion restriction. Allowing for unit-specific spillover coefficients, the granular instrument contains variation from r_{St}:

$$z_t = r_{St}-r_{Et} = (\phi_S-\phi_E)r_{St}+u_{St}-u_{Et}. \tag{2.2.3}$$

From the $(\phi_S-\phi_E)$ term, the magnitude of the exclusion restriction's violation is greater when spillover coefficients vary systematically by size. Moreover, I show later in Proposition 2.3.1 that there is no guarantee that the GIV spillover coefficient estimand is a non-negative weighted average of spillover coefficients. In this sense, the GIV spillover coefficient estimand could be far from the potentially heterogeneous true spillover coefficients.

2.2.3 RGIV illustration: Environment as GMM moment conditions

This subsection introduces the estimator proposed in this paper, robust granular instrumental variables (RGIV). RGIV exploits the uncorrelatedness of idiosyncratic shocks through the generalized method of moments. The estimator's identifying variation comes from individual country-level idiosyncratic shocks.

Returning to the baseline environment described in Section 2.2.1—which again features unit-specific spillover coefficients ϕ_i and unit-specific shock variances σ_i^2—the RGIV estimator encodes the uncorrelatedness of idiosyncratic shocks through GMM moment conditions. Taking sizes S_1, S_2, S_3 as given, store data in the vector $\mathbf{r}_t = [r_{1t}, r_{2t}, r_{3t}]'$ and parameters in $\boldsymbol{\phi} = [\phi_1, \phi_2, \phi_3]'$. Letting $u_i(\mathbf{r}_t, \boldsymbol{\phi}) = r_{it} - \phi_i r_{St}$, moment function $g(\mathbf{r}_t, \boldsymbol{\phi})$ encodes the condition that idiosyncratic shocks are uncorrelated

$$g(\mathbf{r}_t, \boldsymbol{\phi}) = \Big[u_1(\mathbf{r}_t, \boldsymbol{\phi})u_2(\mathbf{r}_t, \boldsymbol{\phi}) \quad u_1(\mathbf{r}_t, \boldsymbol{\phi})u_3(\mathbf{r}_t, \boldsymbol{\phi}) \quad u_2(\mathbf{r}_t, \boldsymbol{\phi})u_3(\mathbf{r}_t, \boldsymbol{\phi})\Big]'$$

where $\mathbb{E}[g(\mathbf{r}_t, \boldsymbol{\phi}_0)] = 0$ for true parameter $\boldsymbol{\phi}_0$. The parameter-dependent weight matrix is $\widehat{W}(\boldsymbol{\phi}) = \mathrm{diag}\Big(\frac{1}{\widehat{\sigma}_1^2(\boldsymbol{\phi})\widehat{\sigma}_2^2(\boldsymbol{\phi})}, \frac{1}{\widehat{\sigma}_1^2(\boldsymbol{\phi})\widehat{\sigma}_3^2(\boldsymbol{\phi})}, \frac{1}{\widehat{\sigma}_2^2(\boldsymbol{\phi})\widehat{\sigma}_3^2(\boldsymbol{\phi})}\Big)$ for $\widehat{\sigma}_i^2(\boldsymbol{\phi}) = \frac{1}{T}\sum_{t=1}^T u_i(\mathbf{r}_t, \boldsymbol{\phi})^2$. Then, for parameter space $\boldsymbol{\Phi}$ where $\phi_S < 1$ for $\phi \in \Phi$, the robust granular instrumental variables (RGIV) estimator is defined as a continuously updating GMM estimator:

$$\widehat{\boldsymbol{\phi}}^{RGIV} = \arg\min_{\phi \in \boldsymbol{\Phi}} \Big(\frac{1}{T}\sum_{t=1}^T g(\mathbf{r}_t, \boldsymbol{\phi})\Big)' \widehat{W}(\boldsymbol{\phi}) \Big(\frac{1}{T}\sum_{t=1}^T g(\mathbf{r}_t, \boldsymbol{\phi})\Big). \tag{2.2.4}$$

$\widehat{W}(\boldsymbol{\phi})$ is an efficient GMM weight matrix when idiosyncratic shocks are independent. Avoiding inversion of a potentially non-diagonal matrix, $\widehat{W}(\boldsymbol{\phi})$ also ensures numerical stability if the initialization of $\boldsymbol{\phi}$ is far from the GMM objective function's minimum.

Equivalently, the continuously updating GMM formulation minimizes the average squared correlation coefficients between pairs of estimated idiosyncratic shocks. To see this, define the estimated correlation coefficient of idiosyncratic shocks as $\widehat{\rho}_{ij}(\boldsymbol{\phi}) = \frac{\frac{1}{T}\sum_{t=1}^T u_i(\mathbf{r}_t, \boldsymbol{\phi})u_j(\mathbf{r}_t, \boldsymbol{\phi})}{\sqrt{\widehat{\sigma}_i^2(\boldsymbol{\phi})\widehat{\sigma}_j^2(\boldsymbol{\phi})}}$.

Then, the RGIV estimator is

$$\widehat{\boldsymbol{\phi}}^{RGIV} = \arg\min_{\boldsymbol{\phi}\in\boldsymbol{\Phi}} \frac{1}{3}\left[\widehat{\rho}_{12}(\boldsymbol{\phi})^2 + \widehat{\rho}_{13}(\boldsymbol{\phi})^2 + \widehat{\rho}_{23}(\boldsymbol{\phi})^2\right]$$

after multiplying the objective function in Equation 2.2.4 by $\frac{1}{3}$. Intuitively, RGIV chooses the spillover coefficient vector $\boldsymbol{\phi}$ that makes the estimated shocks the least correlated.

RGIV also admits an instrumental variables interpretation, as a country's spillover coefficient is estimated using information from internally estimated idiosyncratic shocks to other countries. The asymptotic variance matrix of the RGIV estimator is $V = (G'\Sigma^{-1}G)^{-1}$ for Jacobian matrix $G = \mathbb{E}[\nabla_{\phi} g(\mathbf{r}_t, \boldsymbol{\phi}_0)]$ and moment covariance matrix $\Sigma = \mathbb{E}[g(\mathbf{r}_t, \boldsymbol{\phi}_0)g(\mathbf{r}_t, \boldsymbol{\phi}_0)']$. Then the diagonal entries of V are

$$\text{Avar}\left(\widehat{\phi}_i^{RGIV}\right) = \frac{\sigma_i^2}{\prod_{j\neq i}\left(S_j^2\sigma_j^2\right)} \frac{\left(1-\phi_S\right)^2\left(S_1^2\sigma_1^2 + S_2^2\sigma_2^2 + S_3^2\sigma_3^2\right)}{4}.$$

The above display illustrates that the identifying variation of the RGIV estimator doesn't require skewness in the unit size distribution, as the identifying variation comes from individual idiosyncratic shocks. Estimates for $\widehat{\phi}_i^{RGIV}$ are more precise when idiosyncratic shocks to countries $j \neq i$ have higher variance. Intuitively, the estimated idiosyncratic shocks can be viewed as sequentially estimated internal instruments. Guessing the spillover coefficient of country 1, the resulting estimated idiosyncratic shock to country 1 can be used as an instrument for the estimation of the spillover coefficients for countries 2 and 3. If one or more pairs of estimated idiosyncratic shocks are too correlated, the procedure is repeated for a new spillover coefficient.

2.3 General model and main results

This section describes the assumptions needed for the robust GIV estimator for $n < \infty$ countries.

2.3.1 Assumptions

Assumption 2.3.1. *(Baseline model)*

(i) ***Model**: For fixed $n \geq 3$ units, let known sizes $S_i \in (0,1)$ sum to 1. Outcome $\mathbf{r}_t = [r_{1t}, \ldots, r_{nt}]'$ responds to the size-aggregated outcome r_{St} according to spillover coefficient ϕ_i and unobserved shocks $\mathbf{u}_t = [u_{1t}, u_{2t}, \ldots, u_{nt}]'$*

$$r_{it} = \phi_i r_{St} + u_{it}, \quad \forall i = 1, ..., n, \quad \phi_S < 1.$$

(ii) ***Shock moments**: For $\sigma_i^2 > 0$, shocks $\mathbf{u}_t$ are i.i.d. with moments $\mathbb{E}(\mathbf{u}_t) = 0$, $\mathbb{E}(\mathbf{u}_t \mathbf{u}_t') = \mathrm{diag}(\sigma_1^2, \ldots, \sigma_n^2)$, and $\mathbb{E}(\|\mathbf{u}_t\|^4) < \infty$. Moreover, u_{it} is independent of u_{jt} for $i \neq j$.*

(iii) ***Parameter space**: For spillover coefficient $\boldsymbol{\phi} = [\phi_1, \ldots, \phi_n]'$, the true parameter $\boldsymbol{\phi}_0$ is in the interior of parameter space $\boldsymbol{\Phi}$. $\boldsymbol{\Phi}$ is compact and for any $\widetilde{\boldsymbol{\phi}} \in \boldsymbol{\Phi}$, $\widetilde{\phi}_S < 1$.*

In Assumption 2.3.1(i), the outcome variable r_{it} responds to the size-aggregated outcome r_{St} according to spillover coefficient ϕ_i and idiosyncratic shock u_{it}. Through $\phi_S < 1$, positive idiosyncratic shocks increase the size-weighted outcome r_{St}. As will be outlined below, the RGIV estimator is just-identified for $n = 3$ and over-identified for $n > 3$. Moreover, size S_i is taken to be *known* by the econometrician. Hence, the model presented in 2.3.1(i) can be modified to allow for time-varying size (for $S_{it} \in (0,1)$ and $\sum_{i=1}^{n} S_{it} = 1$) without changing the proofs to follow. In Assumption 2.3.1(iii), the parameter space is assumed to be compact and restricted to encode the sign restriction of $\phi_S < 1$. Compactness is a standard technical assumption for the consistency of extremum estimators (Newey and McFadden, 1994).

2.3.2 RGIV estimator

The GMM estimator established in Definition 2.3.1 below (called RGIV) exploits the uncorrelated unit-specific shock condition established in Assumption 2.3.1(ii). RGIV is constructed as a continuously updating GMM estimator (Hansen et al., 1996). The moment function

$g(\mathbf{r}_t, \boldsymbol{\phi})$ contains the pairwise products of each of the estimated shocks, and the weight matrix inversely weights each moment by the product of the respective estimated shock variances. Exploiting shock uncorrelatedness can be best understood as being consistent with the tradition of exploiting second moment information in the traditional SVAR identification setting (Kilian and Lütkepohl, 2017; Leeper et al., 1996).

Definition 2.3.1. *For outcome variable $\mathbf{r}_t = [r_{1t}, \ldots, r_{nt}]'$, let $u_i(\mathbf{r}_t, \boldsymbol{\phi}) = r_{it} - \phi_i r_{St}$ for $i = 1, ..., n$. The moment function $g(\mathbf{r}_t, \boldsymbol{\phi})$ is*

$$g(\mathbf{r}_t, \boldsymbol{\phi}) = [u_1(\mathbf{r}_t, \boldsymbol{\phi})u_2(\mathbf{r}_t, \boldsymbol{\phi}), \ldots,\ u_1(\mathbf{r}_t, \boldsymbol{\phi})u_n(\mathbf{r}_t, \boldsymbol{\phi}),\ u_2(\mathbf{r}_t, \boldsymbol{\phi})u_3(\mathbf{r}_t, \boldsymbol{\phi}), \ldots, u_{n-1}(\mathbf{r}_t, \boldsymbol{\phi})u_n(\mathbf{r}_t, \boldsymbol{\phi})]'.$$

For $\widehat{\sigma}_i^2(\boldsymbol{\phi}) = \frac{1}{T}\sum_{t=1}^{T} u_i(\mathbf{r}_t, \boldsymbol{\phi})^2$, define the sample weight matrix

$$\widehat{W}(\boldsymbol{\phi}) = \mathrm{diag}\Big(\frac{1}{\widehat{\sigma}_1^2(\boldsymbol{\phi})\widehat{\sigma}_2^2(\boldsymbol{\phi})}, \ldots, \frac{1}{\widehat{\sigma}_1^2(\boldsymbol{\phi})\widehat{\sigma}_n^2(\boldsymbol{\phi})}, \frac{1}{\widehat{\sigma}_2^2(\boldsymbol{\phi})\widehat{\sigma}_3^2(\boldsymbol{\phi})}, \ldots, \frac{1}{\widehat{\sigma}_{n-1}^2(\boldsymbol{\phi})\widehat{\sigma}_n^2(\boldsymbol{\phi})}\Big).$$

Then, for GMM objective function $\widehat{Q}_T(\boldsymbol{\phi}) = \Big[\frac{1}{T}\sum_{t=1}^{T} g(\mathbf{r}_t, \boldsymbol{\phi})\Big]'\widehat{W}(\boldsymbol{\phi})\Big[\frac{1}{T}\sum_{t=1}^{T} g(\mathbf{r}_t, \boldsymbol{\phi})\Big]$, the ***robust granular instrumental variables (RGIV)*** *estimator is*
$\widehat{\boldsymbol{\phi}}^{RGIV} = \arg\min_{\boldsymbol{\phi}\in\boldsymbol{\Phi}} \widehat{Q}_T(\boldsymbol{\phi})$.

The RGIV estimator is robust in that the estimator allows for unit-specific spillover coefficient heterogeneity while also allowing for unknown shock variances and equal unit sizes. When the researcher is *a priori* certain of homogeneous spillover coefficients $\phi_i = \phi$, the RGIV moment vector can accommodate this parameter restriction by restricting the elements of parameter vector $\boldsymbol{\phi}$ to be homogeneous across units yielding possible gains in efficiency. Moreover, the assumption of parameter homogeneity is formally testable as will be discussed in Section 2.4.

Lemma 2.3.1 (Identification)**.** *Impose Assumption 2.3.1. For $g_0(\boldsymbol{\phi}) = \mathbb{E}[g(\mathbf{r}, \boldsymbol{\phi})]$, $g_0(\boldsymbol{\phi}_0) = 0$ for the true parameter $\boldsymbol{\phi}_0$ and $g_0(\widetilde{\boldsymbol{\phi}}) \neq 0$ for $\widetilde{\boldsymbol{\phi}} \in \boldsymbol{\Phi}$ such that $\widetilde{\boldsymbol{\phi}} \neq \boldsymbol{\phi}_0$.*

Proof. See Appendix B.1.4. ☐

For Lemma 2.3.1, the parameter space restriction $\widetilde{\phi}_S < 1$ in Assumption 2.3.1(iii) rules out the false solution to the population moment condition. The proof of Lemma 2.3.1 shows that there is a second parameter $\check{\boldsymbol{\phi}}$ such that $g_0(\check{\boldsymbol{\phi}}) = 0$. However, $\check{\boldsymbol{\phi}}$ is not in the parameter space because $\check{\phi}_S = 1 + (1 - \phi_S) > 1$ and thus is not a candidate solution. In economic terms, Assumption 2.3.1(iii) represents the researcher's knowledge of the sign of an idiosyncratic shock's effect on the size-weighted outcome r_{St}. This knowledge can come from an application's institutional details; for the case of the running example of Eurozone yield spread spillovers, a positive idiosyncratic shock to one country gives rise to an increase in aggregated yield spreads since losses from the default of government debt are partially shared. The identification lemma can also be adapted for the opposite case $\phi_S > 1$ by reversing the inequality specified in Assumption 2.3.1(iii) to $\widetilde{\phi}_S > 1$.

Mimicking Assumption 1 of Gabaix and Koijen (2024), the Lemma is derived under the weaker condition that the correlation structure of shocks is known (see Condition (ii') in Appendix B.1.4 for details). Theorems 2.3.1 and 2.3.2 (below), which establish consistency and asymptotic normality of the RGIV estimator, are shown under the assumption of uncorrelated shocks but can analogously be adapted to the known shock correlation case. These results follow from a standard application of arguments provided in Pakes and Pollard (1989).

Theorem 2.3.1 (Consistency of RGIV)**.** *Impose Assumption 2.3.1. The RGIV estimator is consistent* $\widehat{\boldsymbol{\phi}}^{RGIV} \xrightarrow{p} \boldsymbol{\phi}_0$ *for the true parameter* $\boldsymbol{\phi}_0$ *as* $T \to \infty$*.*

Proof. See Appendix B.1.1. □

Theorem 2.3.2 (Asymptotic normality of RGIV)**.** *Impose Assumption 2.3.1. The RGIV estimator is asymptotically normal*

$$\sqrt{T}(\widehat{\boldsymbol{\phi}}^{RGIV} - \boldsymbol{\phi}_0) \xrightarrow{d} \mathcal{N}(0, (G'WG)^{-1}G'W\Sigma WG(G'WG)^{-1})$$

for RGIV population weight matrix $W = \operatorname{diag}(\frac{1}{\sigma_1^2\sigma_2^2}, \dots, \frac{1}{\sigma_1^2\sigma_n^2}, \frac{1}{\sigma_2^2\sigma_3^2}, \dots, \frac{1}{\sigma_{n-1}^2\sigma_n^2})$*, moment covariance matrix* $\Sigma = \mathbb{E}[g(\mathbf{r}_t, \boldsymbol{\phi}_0)g(\mathbf{r}_t, \boldsymbol{\phi}_0)']$ *and* $G = \mathbb{E}[\nabla_{\boldsymbol{\phi}} g(\mathbf{r}_t, \boldsymbol{\phi}_0)]$ *as* $T \to \infty$*.*

Proof. See Appendix B.1.2. □

Remarks.

1. The assumptions of granularity ($S_i > 0$) and non-zero shock variances ($\sigma_i^2 > 0$) guarantee that $G'WG$ is full rank. Intuitively, these conditions guarantee that idiosyncratic shocks contribute to fluctuations in the endogenous variable r_{St} and can be exploited as identifying variation. Granularity and non-zero shock variances can loosely be viewed as the corresponding "relevance conditions" for the RGIV setting.

2. Diagonal weight matrix $\widehat{W}(\widehat{\boldsymbol{\phi}}^{RGIV})$ is an estimator for the efficient GMM weight matrix when idiosyncratic shocks are not only uncorrelated, but also independent. Its specific form allows the RGIV estimator to be interpreted as the $\boldsymbol{\phi}$ that minimizes the average squared correlation coefficient between shocks. Practically, a continuously updating GMM estimator with diagonal weight matrix $\widehat{W}(\boldsymbol{\phi})$ is one way to ensure the stability of the GMM objective function when T is relatively small. The choice of a diagonal $\widehat{W}(\boldsymbol{\phi})$ hedges against potential numerical instability relative to choosing an alternative weight matrix constructed as $\overline{W}(\boldsymbol{\phi}) = [\frac{1}{T}\sum_{t=1}^{T} g(\mathbf{r}_t, \boldsymbol{\phi})g(\mathbf{r}_t, \boldsymbol{\phi})']^{-1}$. Evaluating an objective function using a weight matrix of the form $\overline{W}(\boldsymbol{\phi})$ would require inverting the potentially ill-conditioned matrix $\frac{1}{T}\sum_{t=1}^{T} g(\mathbf{r}_t, \boldsymbol{\phi})g(\mathbf{r}_t, \boldsymbol{\phi})'$, which could arise if the initial conditions for the sample GMM objective function are far from the true value.

3. In Section 2.7, the procedure is implemented in MATLAB using the constrained nonlinear optimizer `fmincon()`. The sample GMM objective function $\widehat{Q}_T(\boldsymbol{\phi})$ described in Definition 2.3.1 is used as the objective function with the constraint $\phi_S < 1$. As is standard for GMM estimators, the asymptotic variance for $\widehat{\boldsymbol{\phi}}^{RGIV}$ can be computed using "plug-in" estimators for the components of the asymptotic variance expression in Theorem 2.3.2. Specifically, I use the weight matrix estimator $\widehat{W}(\widehat{\boldsymbol{\phi}}^{RGIV})$ from Definition 2.3.1 for W, the sample analogue of the moment covariance matrix $\widehat{\Sigma} =$

$\frac{1}{T}\sum_{t=1}^{T} g(\mathbf{r}_t, \widehat{\boldsymbol{\phi}}^{RGIV}) g(\mathbf{r}_t, \widehat{\boldsymbol{\phi}}^{RGIV})'$ for Σ, and the sample analogue of the Jacobian matrix $\widehat{G} = \frac{1}{T}\sum_{t=1}^{T} \nabla_{\boldsymbol{\phi}} g(\mathbf{r}_t, \widehat{\boldsymbol{\phi}}^{RGIV})$ for G.

4. Under the weaker condition of cross-sectional shock uncorrelatedness (imposing $\mathbb{E}[u_{it}u_{jt}] = 0$ for $i \neq j$ rather than the stronger condition of $u_{it} \perp\!\!\!\perp u_{jt}$), the RGIV estimator is still consistent and asymptotically normal. However, the GMM weight matrix $\widehat{W}(\widehat{\boldsymbol{\phi}}^{RGIV})$ would no longer converge in probability to the efficient GMM weight matrix $W = \mathbb{E}[g(\mathbf{r}_t, \boldsymbol{\phi}_0) g(\mathbf{r}_t, \boldsymbol{\phi}_0)']^{-1}$. Shock independence guarantees that off-diagonal elements of W are zero and that the on-diagonal elements are multiplicatively separable i.e. that $\mathbb{E}(u_{it}^2 u_{jt}^2) = \mathbb{E}(u_{it}^2)\mathbb{E}(u_{jt}^2)$ for $i \neq j$. Under the weaker condition of uncorrelated shocks, higher order dependence (like that arising from a shared volatility term) is permissible.[5] Here, the weight matrix is no longer efficient but the estimator is still consistent and asymptotically normal.

5. $\widehat{\boldsymbol{\phi}}^{RGIV}$ can still be used as a plug-in estimator for the efficient GMM weight matrix when idiosyncratic shocks are uncorrelated but not independent. The following gives an example of one such procedure. First, as noted in the previous bullet, RGIV can be used to obtain a preliminary (consistent) estimate of the spillover coefficient. Second, GMM can be used to estimate $\widehat{\boldsymbol{\phi}}^{\text{Step 2}}$ with sample GMM weight matrix $\widehat{W}^{\text{Step 2}} = [\frac{1}{T}\sum_{t=1}^{T} g(\mathbf{r}_t, \widehat{\boldsymbol{\phi}}^{RGIV}) g(\mathbf{r}_t, \widehat{\boldsymbol{\phi}}^{RGIV})']^{-1}$. Since $\widehat{W}^{\text{Step 2}}$ is an estimator for the efficient GMM weight matrix under uncorrelated but not independent shocks, $\widehat{\boldsymbol{\phi}}^{\text{Step 2}}$ is an efficient estimator for $\boldsymbol{\phi}_0$.

2.3.3 GIV under spillover coefficient heterogeneity

In general, the baseline Gabaix and Koijen (2024) GIV estimand ϕ^{GIV} under spillover coefficient heterogeneity cannot be interpreted as a weighted average of unit-specific spillover

[5] Serial correlation is another form of dependence of interest for applied work. Here, Theorem 2.3.2 can be adapted to serially correlated shocks by exchanging the currently used central limit theorem for independent and identically distributed data for one applicable to data with dependence, like those found in Davidson (1994).

coefficients. From Equation 2.2.3, the spillover coefficient homogeneity assumption is necessary for the instrumental variable exclusion restriction to hold, as it ensures the endogenous term r_{St} is differenced away. When shock variances are homogeneous across countries, Proposition 2.3.1 (below) decomposes the GIV estimand into an equal-weighted spillover coefficient term and a term that depends on $(\phi_S - \phi_E)$. The bias is larger when the spillover coefficient varies systematically with the size distribution, giving rise to a larger gap between ϕ_S and ϕ_E. Moreover, as also discussed in Appendix D.8 of Gabaix and Koijen (2024), the bias is smaller when the number of units is large.

Proposition 2.3.1. *Impose Assumption 2.3.1 and shock variance homogeneity ($\sigma_i^2 = \sigma^2$). Then the GIV estimand is* $\boldsymbol{\phi}^{GIV} = \frac{\mathbb{E}(z_t r_{Et})}{\mathbb{E}(z_t r_{St})} = \phi_E + \frac{\phi_S - \phi_E}{n} \cdot \frac{1}{\frac{\phi_S - \phi_E}{1-\phi_S}[\sum_{i=1}^n S_i^2] - \frac{1}{n} + \sum_{i=1}^n S_i^2}$.

Proof. See B.1.3. □

Proposition 2.3.1 implies that the GIV estimand doesn't admit a weighted average interpretation of unit-specific spillover coefficients. To see this, consider the following example. Suppose $n = 3$, $\mathbf{S} = \begin{bmatrix}0.2 & 0.3 & 0.5\end{bmatrix}'$, and $\boldsymbol{\phi} = \begin{bmatrix}0.6 & 0.3 & 0.3\end{bmatrix}'$. Applying Proposition 2.3.1, $\phi^{GIV} = -0.18 \notin [0.3, 0.6]$, so ϕ^{GIV} is not a positive weighted average of individual spillover coefficients. Section 2.7 investigates the practical implications of GIV under spillover coefficient heterogeneity using an empirically relevant DGP.

While Proposition 2.3.1 highlights the potential pitfalls from mistakenly applying the baseline GIV estimator, Gabaix and Koijen (2024) also give guidance for estimating unit-specific spillover coefficients under additional restrictions. Their Proposition 6 and Appendix D.9 present alternative procedures that require heterogeneity to depend on observables and for the shock variance to be known respectively. In contrast, these restrictions are unnecessary for RGIV.

2.4 Testing

In this section, I propose two tests derived from standard GMM results (Newey and McFadden, 1994): a test of over-identifying restrictions that evaluates the uncorrelatedness of idiosyncratic shocks and a test that evaluates the homogeneous spillover coefficient condition of Gabaix and Koijen (2024). For what follows, impose Assumption 2.3.1.

2.4.1 RGIV specification test

The uncorrelatedness of idiosyncratic shocks is directly testable when there are four or more units, as the number of moment conditions exceeds the number of unit-specific spillover coefficients. Recall that the RGIV estimator is a GMM estimator for moment function $g(\mathbf{r}_t, \boldsymbol{\phi})$, which encodes the pairwise uncorrelatedness of idiosyncratic shocks. For $n \geq 4$ countries, the number of moments exceeds the number of estimated parameters allowing for the use of the Sargan–Hansen test. The null hypothesis H_0: $\mathbb{E}[g(\mathbf{r}_t, \boldsymbol{\phi}_0)] = \mathbf{0}$ for true parameter $\boldsymbol{\phi}_0$ is rejected for large values of the J-statistic $J_T = T \cdot \widehat{Q}_T(\widehat{\boldsymbol{\phi}}^{RGIV})$. Intuitively, J_T is large when one or more pairs of estimated idiosyncratic shocks are correlated.

2.4.2 Spillover coefficient homogeneity test

Spillover coefficient homogeneity across units is not only a subject of potential substantive interest—particularly given its role in the Gabaix and Koijen (2024) GIV estimator—but also can be formally evaluated within the framework of RGIV. Since the null hypothesis of coefficient homogeneity H_0: $\phi_1 = \phi_2 = ... = \phi_n$ is a special case of RGIV's unit-specific spillover coefficients, H_0 can be tested with the distance metric test. Here, estimator $\overline{\boldsymbol{\phi}}$ minimizes $\widehat{Q}_T(\boldsymbol{\phi})$ subject to the constraints of null hypothesis H_0. Then, spillover coefficient homogeneity is rejected when the distance metric test statistic $DM_T = T(\widehat{Q}_T(\overline{\boldsymbol{\phi}}) - \widehat{Q}_T(\widehat{\boldsymbol{\phi}}^{RGIV}))$ is large since $DM_T \xrightarrow{d} \chi^2_{n-1}$ under the null hypothesis. A large value of DM_T indicates that the constraints of null hypothesis H_0 bind.

2.5 Extensions to additional explanatory variables

Motivated by the requirements of empirical applications, this section discusses two extensions that relax the condition of uncorrelated idiosyncratic shocks discussed in Section 2.3 through observed and unobserved explanatory variables. I show that when observable time-varying explanatory variables determine the shocks' correlation structure, RGIV can be used after residualizing the outcome variables with respect to these observables. When the correlation structure is instead determined by unobserved factors with unknown loadings, the global identification condition fails.

2.5.1 RGIV with observed explanatory variables

In practice, observable characteristics can determine the correlation structure among shocks as described by the following two situations. First, the outcome variable could have unit-specific exposures to a particular observed variable—take country-specific exposures to the USD-EUR exchange rate in the Euro area sovereign yields example. Second, observable characteristics could also be used to account for a correlation structure driven by unobserved explanatory variables—like country-specific exposures to a "global financial conditions" factor—so long as such unobserved factors are in the span of the observed explanatory variables.

Observed variable $\mathbf{x}_t$ $(k \times 1)$ affects outcome variable r_{it} through a direct effect and an indirect effect. Modifying Assumption 2.3.1(i), unit-specific coefficients $\boldsymbol{\beta}_i$ determine the cross-sectional correlation of v_{it}

$$r_{it} = \phi_i r_{St} + \underbrace{\boldsymbol{\beta}_i'\mathbf{x}_t + u_{it}}_{v_{it}}, \quad \mathbf{x}_t \perp\!\!\!\perp u_{it}, \quad i = 1, \dots, n \tag{2.5.1}$$

where u_{it} is still idiosyncratic in the sense that u_{it} is independent of u_{jt} for $i \neq j$. The orthogonality condition $u_{it} \perp\!\!\!\perp \mathbf{x}_t$ can be interpreted as a "selection-on-observables" assumption; the cross-sectional correlation of v_{it} is entirely determined by the observed explanatory variables. Holding r_{St} and u_{it} constant, $\boldsymbol{\beta}_i$ can be interpreted as the direct effect of $\mathbf{x}_t$ on r_{it}.

Additionally, mediated through changes in $r_{St} = \frac{1}{1-\phi_S}[\boldsymbol{\beta}_S'\mathbf{x}_t + u_{St}]$, $\frac{\phi_i}{1-\phi_S}\beta_S$ determines the indirect effect of $\mathbf{x}_t$ on r_{it}. Analogous to the discussion in Section 2.2.1, the indirect effect of a change of $\mathbf{x}_t$ on unit i is larger when units are on average sensitive to $\mathbf{x}_t$ (through $\boldsymbol{\beta}_S$), unit i is sensitive to spillovers (through ϕ_i), and when units are on average sensitive to spillovers (through ϕ_S).

The spillover coefficients in Equation 2.5.1 can be estimated by using RGIV after residualizing the outcome variable r_{it} with respect to observed variables $\mathbf{x}_t$. Residualizing purges r_{it} of the variation induced by the direct and indirect effects of $\mathbf{x}_t$ on r_{it}. Concretely, the procedure has two steps: (1) Compute the residual $\dot{r}_{it}$ of the regression of r_{it} on $\mathbf{x}_t$, (2) treating $\dot{r}_{it}$ as data, estimate $\boldsymbol{\phi}$ using the RGIV estimator described in Definition 2.3.1. Conveniently, estimation uncertainty of the first step's regression coefficients has no effect on the asymptotic variance of the RGIV estimator in the second step. Thus, treating $\dot{r}_{it}$ as data in the second step produces valid standard errors for $\widehat{\boldsymbol{\phi}}^{RGIV}$. See Appendix B.1.6 for formal results.

2.5.2 Non-identification when explanatory variables are unobserved

In this section, I describe the tradeoff in assumptions between allowing spillover coefficient heterogeneity/unknown shock variances and a factor structure for the shocks. I show that global identification of unit-specific spillover coefficients is lost when a single unobserved factor is included in the error term. Such a case is empirically relevant because there is typically no a priori reason to believe that spillover coefficients are homogeneous across units and because estimated latent factors are commonly used as control variables in the applied GIV literature.

GIV regressions with estimated latent factors as control variables are ubiquitous in the applied GIV literature (Flynn and Sastry (2022); Gabaix and Koijen (2024, 2021); Camanho et al. (2022); Baumeister and Hamilton (2023); Adrian et al. (2022) among others). Typically, latent factors are estimated using principal components on the demeaned outcome variable before being included as control variables in the GIV regression. Such an approach is attractive because it enables practitioners to apply GIV to applications where the correlations between unit shocks are driven by a small number of latent factors. These estimated factors however are subject to measurement error, so their inclusion as control variables in subsequent GIV regressions give rise to attenuation bias.

Banafti and Lee (2022) address this concern by extending GIV with homogeneous spillover coefficients to a large time and panel dimension framework. When the size distribution of units is very skewed (more skewed than Zipf's law), the sampling uncertainty arising from latent factors and loadings is negligible under their procedure. Homogeneity of spillover coefficients across units is crucial for the procedure's validity. When spillover coefficients are heterogeneous, cross-sectionally demeaning each unit no longer differences away the (no longer constant) contribution of spillovers. As a result, PCA estimates of the latent factors are polluted by the presence of spillovers.

Given the empirical relevance of unit-level heterogeneity and latent factors, I consider a heterogeneous spillover coefficient latent factor model. Taking n to be fixed, restrictions on the skewness of the unit size distribution are unneeded. Investigating identification, I augment Assumption 2.3.1(i) to include a single latent factor f_t with unknown unit-specific loading λ_i

$$r_{it} = \phi_i r_{St} + \lambda_i f_t + u_{it}, \quad \lambda_i \neq 0, \quad f_t \perp\!\!\!\perp u_{it}, \quad n \geq 5 \tag{2.5.2}$$

where the number of units n is fixed and time $T \to \infty$. Factor f_t is normalized so that $\mathbb{E}(f_t) = 0$, $\mathbb{E}(f_t^2) = 1$, and $\lambda_1 > 0$. Shocks u_{it} are idiosyncratic in that they are independent of latent factor f_t and $u_{it} \perp\!\!\!\perp u_{jt}$ for $i \neq j$. Then, extending the logic of the RGIV estimator to the single factor case, the moment function between units $i \neq j$ is

$$g_{ij}^{\text{factor}}(\mathbf{r}_t, \boldsymbol{\theta}) = (r_{it} - \phi_i r_{St})(r_{jt} - \phi_j r_{St}) - \lambda_i \lambda_j$$

where $\boldsymbol{\theta} = [\boldsymbol{\phi}', \boldsymbol{\lambda}']'$ for $\boldsymbol{\lambda} = [\lambda_1, \ldots, \lambda_n]'$. $g_{ij}^{\text{factor}}(\mathbf{r}_t, \boldsymbol{\theta})$ can then be stored in moment vector $g^{\text{factor}}(\mathbf{r}_t, \boldsymbol{\theta})$. When $n \geq 5$, the number of moments (a total of $n(n-1)/2$) is greater than or equal to the number of parameters to be estimated (a total of $2n$).

Lemma 2.5.1 (below) however shows that the population moment condition $g_0^{\text{factor}}(\boldsymbol{\theta}) = \mathbb{E}[g(\mathbf{r}, \boldsymbol{\theta})]$ has multiple roots, so the model described in Equation 2.5.2 is not identified.

Lemma 2.5.1. *Consider Equation 2.5.2 and let $\boldsymbol{\theta}_0$ be the true parameter. There exists $\widetilde{\boldsymbol{\theta}} \neq \boldsymbol{\theta}_0$ such that $g_0^{\text{factor}}(\widetilde{\boldsymbol{\theta}}) = 0$.*

Proof. See Appendix B.1.5. □

The failure in the global identification condition comes from the assumptions of unknown shock variances and unknown factor loadings. If instead shock variances and factor loadings were taken to be known—mirroring Assumption 1 of Gabaix and Koijen (2024) and as is the case for the proof of Lemma 2.3.1—then the unit-specific spillover coefficients are identified.[6] In the proof of Lemma 2.5.1, I show that factor loadings can compensate for incorrect guesses for the spillover coefficient. To see this, let $\widetilde{\boldsymbol{\theta}} = [\widetilde{\boldsymbol{\phi}}', \widetilde{\boldsymbol{\lambda}}']'$ be a candidate root to the population moment condition. In the proof, I consider the set of solutions where the first unit's spillover coefficient and loading equal their true values ($\widetilde{\phi}_1 = \phi_1$ and $\widetilde{\lambda}_1 = \lambda_1$). I then show that a subset of the population moment conditions imply that $\widetilde{\phi}_k - \phi_k = -\frac{\lambda_1(1-\phi_S)}{\lambda_1\lambda_S + S_1\sigma_1^2}(\widetilde{\lambda}_k - \lambda_k)$ for $k \geq 2$. In words, $\widetilde{\lambda}_k$ can compensate for an incorrect spillover coefficient $\widetilde{\phi}_k \neq \phi_k$. Formally, the proof shows that there is at least one $\widetilde{\boldsymbol{\theta}} \neq \boldsymbol{\theta}_0$ such that $g_0^{\text{factor}}(\widetilde{\boldsymbol{\theta}}) = 0$. Moreover, $\widetilde{\boldsymbol{\theta}}$ need not be "close" to the true parameter, as there is no guarantee that $\max_i \widetilde{\phi}_i \geq \min_i \phi_i$ or $\min_i \widetilde{\phi}_i \leq \max_i \phi_i$. In this sense, $\widetilde{\boldsymbol{\phi}}$ is potentially far from the true spillover coefficient $\boldsymbol{\phi}_0$.

Lemma 2.5.1 also implies a tradeoff between modeling unit-level heterogeneity and allowing for correlated shocks. Recall that Proposition 7 of Gabaix and Koijen (2024) shows that a *homogeneous* spillover coefficient is identified when shocks admit a factor structure with unknown loadings and if shock variances are homogeneous across units. In contrast, Lemma

[6] In a related case, Appendix B.2.2 shows that differencing RGIV moments can account for factor loadings determined by unit-specific observables.

2.5.1 shows that global identification is lost under unrestricted heterogeneity on spillover coefficients and shock variances. Taken together, these results suggest that practitioners face a tradeoff between two empirically-relevant models.

2.6 Application

Applying the robust granular instrumental variables (RGIV) methodology to the Euro area sovereign yield spillovers application of a working paper version of Gabaix and Koijen (2024), this section finds strong evidence of country-level heterogeneity in the spillovers of idiosyncratic shocks.

Just as in Gabaix and Koijen (2024), the sample consists of daily data on 10-year zero coupon yields from Bloomberg from September 1, 2009 to May 31, 2018 giving a total of 2283 observations. The included countries are Austria, Belgium, Finland, France, Germany, Greece, Ireland, Italy, Netherlands, Portugal, Slovenia, and Spain. For the yield spread of country i (relative to Germany) y_{it}, the outcome variable r_{it} is defined as $r_{it} = \frac{y_{it}-y_{it-1}}{0.01+y_{i,t-1}}$ just as in Gabaix and Koijen (2024). Since shocks are linearly related to the outcome variable as outlined in Assumption 2.3.1(i), r_{it} is winsorized (over time) at the 0.5 and 99.5th percentiles. Following Gabaix and Koijen (2024), size is time-varying and computed as "debt-at-risk" $S_{i,t-1} = \frac{B_{i,t-1}y_{i,t-1}}{\sum_j B_{j,t-1}y_{j,t-1}}$ where $B_{i,t-1}$ is the outstanding government debt of country i.

Observed explanatory variables are included to account for unit-specific exposures to latent aggregate shocks. In particular, I include the STOXX 50 Volatility Index (differences), the STOXX Europe 600 Index (growth), the EUR-USD exchange rate (growth), the United States 10 Year Treasury yield (growth), BBB/Baa-10Y spread (differences), and the European Fama-French 5 factors (Fama and French, 2015). These observed explanatory variables account for differential exposure of countries to uncertainty, exchange rates, equity prices, and risk. All data are downloaded from Bloomberg.

Even with observed explanatory variables, well-documented regional correlations among Europe's core and periphery countries likely still drive correlations between shocks (Bayoumi and Eichengreen, 1992). To address this concern, I size-aggregate r_{it} to form larger country blocks; even if shocks *within* blocks are correlated, RGIV is still valid so long as shocks *between* blocks are uncorrelated. Hence, uncorrelatedness between blocks is a weaker condition than uncorrelatedness between countries. Specifically, I consider the following four blocks:

- Block 1 (*Core*): Austria, Belgium, Finland, France, Netherlands;
- Block 2 (*Western periphery*): Ireland, Portugal, and Spain;
- Block 3 (*Eastern periphery*): Greece and Italy;
- Block 4: Slovenia.

Block 1 includes countries typically classified as being members of the EU "core." Blocks 2 and 3 contain countries typically classified as being members of the EU "periphery." Block 4 contains Slovenia, which is typically left uncategorized on account of its distinct institutional structure as a former member of Yugoslavia.

I report RGIV point estimates and standard errors for individual spillover coefficients. Conservatively, standard errors are computed using a HAC GMM weight matrix to account for the possible serial correlation of idiosyncratic shocks.[7] Size-weighted spillover coefficients are constructed using the Delta method, using the average block size over the estimation sample.

I also present GIV results based on the shock variance approximation and factor estimation procedures detailed in Section 5.3 of the July 2021 working paper version of Gabaix and Koijen (2024). I include this procedure for comparison, as it is used in the original application. This procedure approximates the variance of the shocks with $\mathrm{Var}(r_{it})$, valid when spillovers are small. As a diagnostic for GIV instrument strength, the first stage F-statistic is also reported.

[7] Specifically, a Newey-West kernel with the Lazarus et al. (2018) truncation parameter rule of $1.3\sqrt{T}$.

I find strong evidence of spillovers in the aggregate and of spillover heterogeneity across countries. In the preferred specification (Column 1 of Table 2.6.1), the null hypothesis of uncorrelated idiosyncratic shocks isn't rejected suggesting that the RGIV moment conditions are consistent with the data. Evidence of spillovers in the aggregate, the size-weighted spillover coefficient is 0.54 (with a standard error of 0.08). Moreover, with a p-value of < 0.001 for the spillover coefficient homogeneity test, the null hypothesis of spillover coefficient homogeneity is rejected at conventional significance levels. Turning to individual spillover coefficients, the spillover coefficient for the western periphery block (0.83) is more than twice that of the core countries (0.40).

The qualitative features of the preferred specification are robust to omitting observed explanatory variables and using an estimator that is efficient under higher-order dependence among shocks. Omitting explanatory variables, Column 2 of Table 2.6.1 shows spillover coefficient heterogeneity across blocks. Relative to the preferred specification, the estimated size-weighted spillover coefficient is slightly larger (at 0.63). Recall that the RGIV estimator is efficient under the independence of idiosyncratic shocks, but isn't guaranteed to be efficient under the weaker condition of idiosyncratic shock uncorrelatedness.[8] To address, Column 3 shows the results of a "higher-order efficient" estimator, which uses the procedure outlined in Remark 5 of Section 2.3.2. The point estimates and standard errors are nearly identical, suggesting that higher order dependence of idiosyncratic shocks play a minor role in this application. The spillover coefficient estimated with GIV procedure featured in a working paper version of Gabaix and Koijen (2024) is comparable to the size-weighted spillover coefficient computed with RGIV, but is unable to speak to country-level heterogeneity. Column 4 of Table 2.6.1 presents results of the baseline GIV methodology applied to this section's core-periphery-aggregated panel (distinct from the country-level panel of Gabaix and Koijen (2024)). The spillover coefficient of 0.57 is close to the size-aggregated spillover coefficient computed in the preferred RGIV specification.

[8]For example, a shared volatility term would induce dependence of higher order moments.

	Preferred	No controls	Higher-order efficient	1-factor GIV
RGIV results:				
ϕ_S	0.54	0.63	0.54	
	(0.08)	(0.07)	(0.08)	
$\phi_{\text{IRL,PRT,ESP}}$	0.83	0.86	0.83	
	(0.06)	(0.05)	(0.06)	
$\phi_{\text{GRC, ITA}}$	0.44	0.56	0.44	
	(0.16)	(0.13)	(0.16)	
ϕ_{Core}	0.39	0.47	0.40	
	(0.03)	(0.02)	(0.03)	
ϕ_{SVN}	0.33	0.49	0.33	
	(0.06)	(0.05)	(0.05)	
ϕ^{GIV}				0.56
				(0.02)
Tests (p-values):				
Specification	0.950	0.923	0.935	
Homogeneity	<0.001	<0.001	<0.001	
First-stage F-statistic				592

Table 2.6.1: Spillover coefficient estimation results. RGIV coefficient estimates for ϕ_S, $\phi_{\text{IRL,PRT,ESP}}$, $\phi_{\text{GRC, ITA}}$, ϕ_{Core} and ϕ_{SVN} are listed above standard errors, which are provided in parentheses. p values are provided in the bottom section of the table for the specification test and parameter homogeneity tests. Estimates for the GIV spillover coefficient and standard error are listed in the ϕ^{GIV} row, and the instrumental variables first-stage F-statistic is given in the table's last row. See the main text for details on the table's columns.

Section B.2.3 gives additional RGIV results under alternative winsorizations, the more widely-used Andrews (1991) truncation parameter, omitting the Fama-French factors as observed explanatory variables, and alternative choices of country blocking. Section B.2.3 also contains results for the 0-factor and 2-factor GIV specifications.

Summarizing, RGIV finds strong evidence of country-level differences in sovereign yield spillovers. In response to a 1% increase in the size-weighted relative yield spread, the relative yield spread of "core" countries increases by 0.4% compared to an increase of 0.8% for countries in the western "periphery." Substantively, these estimates point to the importance of understanding the role of country-level characteristics in the heterogeneous propagation of idiosyncratic shocks during sovereign debt crises.

2.7 Simulation study

I close by showing that RGIV has good finite sample performance in simulation using four DGPs based on the preferred specification studied in Section 2.6. In contrast, I show that two versions of the GIV procedures—the "feasible" GIV procedure (featured in this paper's application application) and "oracle" GIV procedure that takes idiosyncratic shock variances σ_i^2 to be known—can severely under-cover under spillover coefficient heterogeneity.

Setup The "homogeneous spillovers" DGP is loosely based on the preferred specification of Section 2.6. Under spillover coefficient homogeneity across units, the DGP serves as a baseline, as both the oracle GIV and RGIV estimators are correctly specified. For $n = 4$ units and a sample length of $T = 2283$, the model parameters are below:

$$\boldsymbol{\phi} = \begin{bmatrix} 0.54 & 0.54 & 0.54 & 0.54 \end{bmatrix}', \quad \boldsymbol{\sigma} = \begin{bmatrix} 0.014 & 0.014 & 0.014 & 0.014 \end{bmatrix}'$$
$$\mathbf{S} = \begin{bmatrix} 0.29 & 0.56 & 0.14 & 0.01 \end{bmatrix}'.$$

Having established an environment for which both the oracle GIV and RGIV estimators are valid, I study three additional DGPs that deviate from the homogeneous spillovers DGP. These illustrate the finite sample properties of RGIV and GIV. First, the "coefficient outlier" DGP takes the homogeneous spillovers specification and sets the spillover coefficient of the fourth unit to 0.75. Second, the "shock variance outlier" DGP takes the homogeneous spillovers specification and sets the shock standard deviation of the first unit to 0.03. Third, the "application" DGP allows for heterogeneous spillover coefficients and shock variances by assigning them to be the estimated values from the application's preferred specification.

For each DGP, I consider results from the RGIV estimator, "feasible" GIV estimator, and "oracle" estimator. For the RGIV estimator, I compute confidence intervals for the spillover coefficients of individual units, size-weighted spillover coefficients, and equal-weighted spillover coefficients. These confidence intervals highlight how RGIV can be used for researchers interested in individual spillover coefficients and for those interested in a single, aggregated parameter. The "feasible" GIV estimator is computed using the $\mathrm{Var}(r_{it}) \approx \mathrm{Var}(u_{it})$ approximation (valid when spillovers are small) while the "oracle" GIV estimator is computed as if the true idiosyncratic shock variance $\mathrm{Var}(u_{it})$ were known. Both GIV estimators are included to show the practical implications of mistakenly assuming spillover coefficient homogeneity across units. The feasible GIV estimator, in particular, is included to evaluate the approximation $\mathrm{Var}(r_{it}) \approx \mathrm{Var}(u_{it})$ and for completeness (it is featured in the application section).

A conservative notion of coverage is used for the GIV estimators. Since the GIV estimators require spillover coefficient homogeneity across units, I report the proportion of GIV confidence intervals that contain *any* positive-weighted average of potentially heterogeneous spillover coefficients—measured as the proportion of GIV confidence intervals with a nonempty intersection with the closed interval $[\min_{1\leq i\leq n} \phi_i, \max_{1\leq i\leq n} \phi_i]$.

Results While the empirical coverage is near the nominal level for RGIV, feasible RGIV can severely under-cover. Table 2.7.1 shows that nearly 95% of the confidence intervals for $\widehat{\phi}_S$ and $\widehat{\phi}_E$ contain the true estimands ϕ_S and ϕ_E in all four DGPs. In contrast, the true spillover coefficient is contained in none of the feasible GIV confidence intervals in the homogeneous spillovers DGP. Considering that the true coefficient is contained in 95% of the oracle estimator confidence intervals, together these results suggest that the approximation $\mathrm{Var}(r_{it}) \approx \mathrm{Var}(u_{it})$ is inappropriate for this DGP.

Under spillover coefficient heterogeneity, the GIV estimators can substantially under-cover—even with the conservative notion of coverage featured in this simulation study. For the coefficient outlier DGP, none of the feasible GIV confidence intervals contain *any* positive weighted average of heterogeneous spillovers, compared to 15% for the oracle estimator. Hence, for certain DGPs, mistakenly assuming spillover coefficient homogeneity and applying versions of the baseline GIV estimator can result in unreliable inference.

DGP	RGIV		GIV		Testing ($\alpha = 0.05$)	
	$\widehat{\phi}_S$	$\widehat{\phi}_E$	Feasible	Oracle	Spec.	Homog.
Homogeneous spillovers	0.94	0.97	0	0.95	0.054	0.042
	(0.12)	(0.038)	(0.058)	(0.046)		
Coefficient outlier	0.94	0.97	0	0.15	0.047	0.998
	(0.12)	(0.037)	(0.078)	(0.071)		
Variance outlier	0.97	0.94	0.0068	0.95	0.045	0.052
	(0.045)	(0.067)	(0.037)	(0.033)		
Application	0.95	0.97	1.00	1.00	0.064	0.996
	(0.14)	(0.052)	(0.083)	(0.10)		

Table 2.7.1: Simulation results (5000 simulations, $T = 2283$, 95% confidence intervals). Empirical coverage probabilities are listed above median confidence interval lengths, which are provided in parentheses. RGIV: Coverage probabilities are computed as the proportion of confidence intervals containing the estimand (ϕ_S for $\widehat{\phi}_S$ and ϕ_E for $\widehat{\phi}_E$). GIV: Coverage probabilities are computed as the proportion of confidence intervals containing any positive-weighted average of ϕ_i for the feasible and oracle GIV estimators. Computed at the 5% level, the RGIV specification test (Spec.) and coefficient homogeneity test (Homog.) are provided in the final two columns.

Evaluating the properties of the testing procedures featured in Section 2.4, the empirical size of the RGIV specification and coefficient homogeneity tests are near their nominal level. The empirical rejection rate of the coefficient homogeneity test is 4.2% and 5.2% for the homogeneous spillovers and variance outlier DGPs respectively, where the null hypothesis of spillover coefficient homogeneity across units is true. When spillover coefficients are not homogeneous, as is the case under the elasticity outlier and application DGPs, the null hypothesis of homogeneous coefficients is rejected in more than 99% of Monte Carlo replications. Turning to the specification test, note that the RGIV framework is correctly specified in all four DGPs. At worst, the null hypothesis of idiosyncratic shock uncorrelatedness is rejected in 6.4% of Monte Carlo replications under the application DGP.

The RGIV estimator presents a modest (if any) power tradeoff relative to the oracle GIV estimator, and unit-specific RGIV spillover coefficients have good coverage properties despite applying to a more general environment. In the homogeneous spillovers application, the equal-weighted RGIV spillover coefficient has a comparable median confidence interval length

(0.038) to that of GIV (0.046).[9] Turning to unit-specific spillover coefficients, Table 2.7.2 shows that the empirical coverage across all four DGPs and RGIV spillover coefficients is near 0.95.

DGP	ϕ_1	ϕ_2	ϕ_3	ϕ_4
Homogeneous spillovers	0.96	0.95	0.95	0.95
	(0.16)	(0.3)	(0.075)	(0.058)
Coefficient outlier	0.95	0.95	0.95	0.94
	(0.16)	(0.3)	(0.075)	(0.058)
Variance outlier	0.95	0.96	0.95	0.95
	(0.43)	(0.18)	(0.046)	(0.044)
Application	0.96	0.95	0.96	0.95
	(0.12)	(0.32)	(0.091)	(0.14)

Table 2.7.2: Simulation results for the RGIV unit-specific spillover coefficients (5000 simulations, $T = 2283$, 95% confidence intervals). Empirical coverage probabilities are listed above median confidence interval lengths, which are provided in paretheses.

2.8 Discussion

Gabaix and Koijen (2024) introduces granularity-based identification, a substantial step forward for credible spillover estimates in macroeconomics and finance. Its baseline granular instrumental variables procedure, however, requires strong assumptions—namely homogeneous spillovers across units, homogeneous shock variances, skewed unit size, and idiosyncratic shocks (after accounting for common factors with known loadings). I build on this innovative approach by showing that unit-specific spillover coefficient heterogeneity with heterogeneous (and unknown) shock variances can jointly be accounted for without further restrictions. My estimator, called robust granular instrumental variables, also allows for homogeneous unit sizes unlike GIV. Intuitively, my approach uses internally estimated individual idiosyncratic shocks as instruments. I give results on identification and inference, showing that the required GIV assumption of coefficient homogeneity is directly testable. Relaxing the idiosyncratic shock

[9] In general, researchers interested in *any* positive-weighted average of potentially heterogeneous spillover coefficients could estimate unit-specific spillover coefficients using RGIV and compute the weights that minimize the confidence interval length of the resulting aggregated estimator.

assumption, I highlight a tradeoff: practitioners must choose between allowing unrestricted unit-level heterogeneity and a general shock covariance structure. Studying Euro zone sovereign yields, I find strong evidence of country-level heterogeneity in spillovers. I find that my proposed estimator has good finite sample properties through simulation.

There are several directions for future work. First, my approach builds on the baseline framework proposed by Gabaix and Koijen (2024). Here, an individual unit's outcome is determined by the size-weighted aggregate outcome. A structure where spillover responses are allowed to differ by the shock's source could be of interest for empirical work. Second, while the core-periphery structure of Euro area countries serves as a natural basis for grouping units in the empirical application, grouping could be automated. Third, as discussed in Section 2.5.2, practitioners face a tradeoff between unrestricted unit-level heterogeneity in spillover coefficients and correlated shocks. Further work on alternative economically-motivated conditions that preserve global identification would be of considerable interest to applied users.

Chapter 3

Double Robustness of Local Projections and Some Unpleasant VARithmetic*

3.1 Introduction

In recent years, local projection (LP) estimators of impulse response functions have become a very popular alternative to structural vector autoregressions (henceforth interchangeably referred to as VAR or SVAR, Sims, 1980). In addition to their simplicity, one potential explanation for the popularity of LPs is their perceived robustness to misspecification, as claimed by Jordà (2005) in his seminal article that proposed the estimation method:

*I thank my coauthors José Luis Montiel Olea, Mikkel Plagborg-Møller, and Christian Wolf for their permission to reproduce our work here. The paper has been presented at the following: University of Michigan, Boston University, European Central Bank, University of Texas at Austin, London School of Economics, Columbia University, Federal Reserve Bank of Philadelphia, Princeton University, Federal Reserve Bank of Cleveland, Conference in Honor of Christopher Sims (Federal Reserve Bank of Atlanta), 2024 NBER Summer Institute, 2024 NBER-NSF Time Series Conference, University of Maryland, 17th Greater New York Area Econometrics Colloquium, Harvard University/Massachusetts Institute of Technology joint seminar, University of Michigan, University of Notre Dame, 2025 World Congress of the Econometric Society, Federal Reserve Bank of St. Louis, and XV Workshop in Time Series Econometrics (University of Zaragoza). The paper is conditionally accepted at Econometrica. The latest public working paper version of the article has been reproduced in this chapter.

> "*[T]hese projections are local to each forecast horizon and therefore more robust [than VARs] to misspecification of the unknown DGP.*"

While this sentiment has been echoed in influential reviews (e.g., Ramey, 2016; Nakamura and Steinsson, 2018; Jordà, 2023), there so far exist essentially no theoretical results on the relative robustness of LP and VAR inference procedures to misspecification. Plagborg-Møller and Wolf (2021) and Xu (2023) show that the two estimators are in fact asymptotically equivalent—and thus equally robust to misspecification—in a general VAR(∞) model if the estimation lag length diverges to infinity with the sample size. However, this result does not directly speak to the empirically relevant case where researchers employ small-to-moderate lag lengths to preserve degrees of freedom. Applied researchers must therefore base their choice of inference procedure on empirically calibrated simulation studies (Kilian and Kim, 2011; Li et al., 2024).

In this paper we provide a formal proof of Jordà's claim that conventional LP confidence intervals for impulse responses are surprisingly robust to misspecification. On the other hand, VAR confidence intervals are robust if, *and only if*, they are as wide as LP intervals asymptotically, as is the case when they control for a large number of lags. If the confidence interval is shorter, then it is *necessarily* unreliable.

We consider a large class of stationary data generating processes (DGPs) that are well approximated by a finite-order SVAR model, but subject to local misspecification in the form of an asymptotically vanishing moving average (MA) process, of potentially infinite order. This class is consistent with essentially all linearized structural macroeconomic models and covers many types of dynamic misspecification, such as under-specification of the lag length, failure to include relevant control variables, inappropriate aggregation, and measurement error. Intuitively, with this set-up we capture the idea that finite-order VAR models provide a good but imperfect approximation of reality.

In this setting, we prove that the conventional LP confidence interval has correct (pointwise) asymptotic coverage even for local misspecification that is of such a large magnitude that it can be detected with probability 1 in large samples. This robustness property requires that we control for those lags of the data that are strong predictors of the outcome or impulse variables, but—crucially for applied work—the omission of lags with small-to-moderate predictive power does not threaten coverage. We argue that our result can be interpreted as a consequence of the *double robustness* of the LP estimator, which is analogous to the double robustness of modern partially linear regression estimators in the literature on debiased machine learning.[2]

In stark contrast to LP, even small amounts of misspecification can cause conventional VAR confidence intervals for impulse responses to suffer from severe undercoverage. We first derive analytically the worst-case bias and coverage of VARs over all possible misspecification processes, subject to a constraint on the overall magnitude of the misspecification. The worst-case bias and coverage distortion are small if, *and only if*, the asymptotic variance is close to that of LP. In general, the only way to guarantee robustness of conventional VAR inference is thus to include so many lags that the VAR estimator is asymptotically equivalent with LP. If instead the VAR confidence interval is much shorter (as is typically the case in applied practice), then it will severely undercover even for a misspecification term that: (i) is small in magnitude; (ii) has dynamic properties that cannot be ruled out *ex ante* based on economic theory; and (iii) is difficult to detect *ex post* with model specification tests. Instead of increasing the lag length, coverage can also be restored by using a larger bias-aware critical value (Armstrong and Kolesár, 2021), but we show that the resulting confidence intervals are so wide that one may as well report the LP interval.

We demonstrate the practical relevance of our theoretical results through a comprehensive review of current practice in the applied VAR literature, together with a simulation study. In papers published in top economics journals, researchers tend to select small to moderate lag lengths, and often report impulse responses at horizons far exceeding the lag length.

[2]Important contributions include Robins et al. (1992), Robins and Rotnitzky (1995), Robins et al. (2000), Robins and Rotnitzky (2001), Bang and Robins (2005), Ai and Chen (2007), Chernozhukov et al. (2018), and Chernozhukov et al. (2022).

Our theory suggests this practice is likely to render inference vulnerable to misspecification. To substantiate this conclusion, we simulate data calibrated to the oil shock application in Känzig (2021). The DGP is taken to be a VAR estimated on the paper's actual data, but with 18 lags rather than 12. The VAR confidence interval materially undercovers—particularly at medium and long horizons—if the lag length is set to 12 or selected by AIC, in line with applied practice, while the LP interval attains close to nominal coverage. Increasing the estimation lag length beyond the conventional choices ameliorates the VAR undercoverage, at the cost of delivering confidence intervals as wide as those of LP.

Literature. Relative to the previously cited simulation studies of LPs and VARs, we here derive *analytical* results on the worst-case asymptotic properties of these two inference procedures that hold for a wide range of stationary, locally misspecified VAR models. The simulations in Li et al. (2024) suggest a stark bias-variance trade-off between LP (low bias, high variance) and moderate-lag VAR estimators (moderate bias, low variance). The reason behind the theoretical superiority of LP proved in this paper is that, if the objective is to construct confidence intervals with robust coverage for a wide range of DGPs, then even a moderate amount of VAR bias cannot be tolerated, as it causes the VAR confidence interval to be poorly centered. A concern for correct confidence interval coverage thus effectively induces a large weight on bias in the researcher's objective function, justifying the use of LP despite its higher variance.

The robustness of LPs to misspecification discussed here—with stationary data and at fixed horizons—is conceptually and theoretically distinct from the robustness of LPs to the persistence in the data and the length of the impulse response horizon shown by Montiel Olea and Plagborg-Møller (2021). Nevertheless, it turns out that controlling for lags ("lag augmentation") is key to all the robustness properties established in Montiel Olea and Plagborg-Møller (2021) and in the present paper.

We also build upon previous research into misspecified VAR models, uncovering novel results about the robustness of LPs and the worst-case properties of VAR procedures. Braun and Mittnik (1993) derive expressions for the probability limits of VAR estimators under global MA misspecification; however, since bias always dominates variance asymptotically in their framework, they do not characterize the properties of LP and VAR inference procedures, which is the focus of our paper. Schorfheide (2005) characterizes the asymptotic mean squared errors of iterated and direct multi-step forecasts in a reduced-form VAR model with MA terms of order $T^{-1/2}$, and González-Casasús and Schorfheide (2025) use this framework to select hyperparameters for VAR forecasts. Müller and Stock (2011) construct Bayesian forecast intervals in a locally misspecified univariate AR model. Relative to these papers, we here contribute by: (i) focusing on structural impulse responses rather than forecasting; (ii) allowing for more general rates of local misspecification, which is key to uncovering the double robustness of LP; and (iii) deriving simple analytical formulae for worst-case bias and coverage of VARs. As such, our results formalize concerns by applied practitioners about the lack of VAR robustness and sensitivity to lag length (Chari et al., 2008; Nakamura and Steinsson, 2018; see also Inoue and Kilian, 2002, and Kilian and Lütkepohl, 2017, Chapters 2.6.5 and 6.2).

Whereas our paper deals with bias imparted by dynamic misspecification, the analysis does not capture other familiar sources of small-sample bias. In particular, our asymptotics abstract from the order-T^{-1} biases that arise from (i) persistence in the data (Pope, 1990; Kilian, 1998; Herbst and Johannsen, 2024) and (ii) the nonlinearity of the impulse response transformation of the VAR parameters (Jensen's inequality).

Outline. Section 3.2 defines the local-to-SVAR model as well as the LP and VAR estimators. Section 3.3 proves the robustness of LP and the fragility of VAR confidence intervals. Section 3.4 derives analytically the worst-case bias and coverage of VARs, and shows that bias-aware VAR confidence intervals tend to be wider than the LP interval. Section 3.5

demonstrates the practical relevance of our theoretical results through a review of the applied literature and a simulation study. Section 3.6 concludes. Replication materials are available online.[3]

Notation. All asymptotic limits are taken as the sample size $T \to \infty$ and are *pointwise* in the sense of fixing the true model parameters and the impulse response horizon. A sum $\sum_{\ell=a}^{b} c_\ell$ is defined to equal 0 when $a > b$.

3.2 Framework

We start out by defining the model and estimators.

3.2.1 Model and assumptions

Extending the forecasting model of Schorfheide (2005), we consider a multivariate, stationary structural VARMA$(1, \infty)$ model that is local to an SVAR(1) model:

$$y_t = Ay_{t-1} + H[I + T^{-\zeta}\alpha(L)]\varepsilon_t, \quad \text{for all } t, \tag{3.2.1}$$

where the data vector $y_t = (y_{1,t}, \dots, y_{n,t})'$ is n-dimensional, the shock vector $\varepsilon_t = (\varepsilon_{1,t}, \dots, \varepsilon_{m,t})'$ is m-dimensional, A is an $n \times n$ matrix, H is an $n \times m$ matrix, $\alpha(L) = \sum_{\ell=1}^{\infty} \alpha_\ell L^\ell$ is an $m \times m$ lag polynomial, and T denotes the sample size. We allow the number of shocks m to potentially exceed the number of variables n, and *vice versa*. We show below that equation (3.2.1) encompasses local-to-SVAR models with $p > 1$ lags by writing them in companion form.

The model (3.2.1) captures the idea that the time series dynamics of the data are well approximated by an autoregressive model driven by unobserved white noise shocks ε_t, but with a small amount of misspecification in the form of an MA process $T^{-\zeta}\alpha(L)\varepsilon_t$. The

[3] https://github.com/ckwolf92/lp_var_inference

misspecification is asymptotically small in the sense that the MA coefficients converge to zero at the rate $T^{-\zeta}$, though the misspecification may still affect the properties of estimators, as shown by Schorfheide (2005) and as demonstrated below. We argue below that MA misspecification of this form can capture many empirically relevant types of dynamic misspecification. We consider local rather than global misspecification in the spirit of local power analysis (e.g., Rothenberg, 1984), since this makes the bias-variance trade-off between the VAR and LP estimators matter even asymptotically as the sample size T diverges, allowing us to make tractable analytical comparisons between these two procedures.

The parameter of interest is the response at horizon h of the variable $y_{i^*,t}$ with respect to the shock $\varepsilon_{j^*,t}$ for some indices $i^* \in \{1, \dots, n\}$ and $j^* \in \{1, \dots, m\}$, to be defined below.

Assumption 3.2.1. *For each T, $\{y_t\}_{t\in\mathbb{Z}}$ is the stationary solution to equation (3.2.1), given the following restrictions on parameters and shocks:*

i) $\varepsilon_t \overset{i.i.d.}{\sim} (0_{m\times 1}, D)$*, where* $D \equiv \operatorname{diag}(\sigma_1^2, \dots, \sigma_m^2)$*, and the elements of* ε_t *are mutually independent. For all* $j = 1, \dots, m$, $\sigma_j^2 > 0$ *and* $E(\varepsilon_{j,t}^4) < \infty$.

ii) All eigenvalues of A are strictly below 1 in absolute value.

iii) The first j^* rows of H are of the form $(\tilde{H}, 0_{j^*\times(m-j^*)})$, where $\tilde{H}$ is a $j^* \times j^*$ lower triangular matrix with 1's on the diagonal. In particular, we require $j^* \le n$.

iv) $S \equiv \operatorname{var}(\tilde{y}_t)$ *is non-singular, where* $\tilde{y}_t \equiv (I - AL)^{-1}H\varepsilon_t$ *is the stationary solution to (3.2.1) when* $\alpha(L) = 0$*. Specifically,* $\boldsymbol{e}(S) = (I - A \otimes A)^{-1}\boldsymbol{e}(\Sigma)$*, where* $\Sigma \equiv HDH'$.

v) $\alpha(L)$ is absolutely summable.

vi) $\zeta > 0$.

The assumptions that the shocks are mutually and serially independent are made to simplify the exposition; we prove formally in Appendix C.3.1 that our results on the robustness of LP and on the asymptotic bias of VAR go through for a large class of conditionally heteroskedastic shock processes. The assumptions on H correspond to recursive (also known as Cholesky) identification of the shock of interest $\varepsilon_{j^*,t}$, with a unit effect normalization $H_{j^*,j^*} = 1$. A special case is when the shock is directly observed, which corresponds to ordering it first (i.e., $j^* = 1$). Appendix C.3.2 shows that our results extend also to identification via external instruments or proxies (Stock and Watson, 2018). Absolute summability of $\alpha(L)$ is a weak regularity condition ensuring the vector MA(∞) process $\alpha(L)\varepsilon_t$ is well-defined (Brockwell and Davis, 1991, Proposition 3.1.1). Finally, the assumption that $\zeta > 0$ restricts attention to local misspecification, as discussed earlier.

The impulse response of interest is defined as

$$\theta_{h,T} \equiv e'_{i^*,n}\left(A^h H + T^{-\zeta}\sum_{\ell=1}^{h} A^{h-\ell} H \alpha_\ell\right) e_{j^*,m} = E[y_{i^*,t+h} \mid \varepsilon_{j^*,t} = 1] - E[y_{i^*,t+h} \mid \varepsilon_{j^*,t} = 0],$$

where $e_{i,n}$ denotes the n-dimensional unit vector with a 1 in position i. The first term in the parenthesis is the usual VAR impulse response formula, while the second term arises from the MA component. Importantly, and consistent with our focus on the consequences of dynamic misspecification, we do not treat the VAR misspecification as non-classical measurement error that should be ignored for structural analysis; instead, the true causal model has a VARMA form (with small but potentially non-zero MA terms), and we care about the full transmission mechanism of shocks in this model.

Additional lags. Our framework covers local-to-SVAR(p) models of the form

$$\check{y}_t = \sum_{\ell=1}^{p} \check{A}_\ell \check{y}_{t-\ell} + \check{H}[I + T^{-\zeta}\alpha(L)]\varepsilon_t, \tag{3.2.2}$$

where $\check{y}_t$ is $\check{n}$-dimensional, the $\check{A}_\ell$ matrices are $\check{n} \times \check{n}$, and $\check{H}$ is $\check{n} \times m$ and satisfies Assumption 3.2.1(iii)). This fits into the original model (3.2.1) if we set $n = \check{n}p$ and define the companion form representation

$$y_t = \begin{pmatrix} \check{y}_t \\ \check{y}_{t-1} \\ \check{y}_{t-2} \\ \vdots \\ \check{y}_{t-p+1} \end{pmatrix}, \quad A = \begin{pmatrix} \check{A}_1 & \check{A}_2 & \dots & \check{A}_{p-1} & \check{A}_p \\ I & 0 & \dots & 0 & 0 \\ 0 & I & \dots & 0 & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & \dots & I & 0 \end{pmatrix}, \quad H = \begin{pmatrix} \check{H} \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

In particular, we can allow the estimation lag length p to exceed the true minimal lag length p_0 of the model by setting $\check{A}_\ell = 0$ for $\ell > p_0$. This fact will prove useful when we consider what happens as the lag length of the estimated VAR is increased.

Types of misspecification. Our local-to-SVAR model (3.2.1) with MA misspecification covers several empirically relevant types of model misspecification. While essentially all modern discrete-time, linearized macro models have VARMA representations, they usually cannot be represented exactly as finite-order VAR models (e.g., Kilian and Lütkepohl, 2017, Chapter 6.2). Even if the true DGP were a finite-order VAR, dynamic misspecification of the estimation model can give rise to MA terms, for example due to under-specification of the lag length or failing to control for some of the variables in the true system. Relatedly, MA terms may appear because of a failure of invertibility of the shocks (Alessi et al., 2011). VARMA representations can also arise from temporal or cross-sectional aggregation of finite-order VAR models, including contamination by classical measurement error (Granger and Morris, 1976; Lütkepohl, 1984). In all of these cases, if the number of lags used for estimating the VAR is chosen to be sufficiently large, then the MA remainder will be small, consistent with the spirit of our locally misspecified model (3.2.1).

In terms of structural shock identification, our framework accommodates both the case of a well-identified shock (or instrument/proxy, see Appendix C.3.2) but misspecification in other parts of the model, as well as misspecification in the structural shock identification itself. Key to this generality is that we allow the $m \times m$ MA polynomial $\alpha(L)$ to be arbitrary. To see this, consider the case $j^* = 1$, so interest centers on the dynamic causal effects of the first shock $\varepsilon_{1,t}$. If the first row of $\alpha(L)$ is zero, then $\varepsilon_{1,t}$ is well-identified as the reduced-form residual in the first equation of the VAR. If the first row of $\alpha(L)$ is non-zero, then the reduced-form residual will be contaminated by lagged shocks, thus allowing for the possibility that shock identification itself is not entirely accurate.

3.2.2 Estimators

We consider two estimators of the impulse response $\theta_{h,T}$ using the data $\{y_t\}_{t=1}^T$:

1. The *LP* estimator is the coefficient $\hat{\beta}_h$ in a regression of $y_{i^*,t+h}$ on $y_{j^*,t}$, controlling for $\underline{y}_{j^*,t} \equiv (y_{1,t}, \dots, y_{j^*-1,t})'$ (i.e., the variables ordered before $y_{j^*,t}$, if any) and lagged data:

$$y_{i^*,t+h} = \hat{\beta}_h y_{j^*,t} + \hat{\omega}_h' \underline{y}_{j^*,t} + \hat{\gamma}_h' y_{t-1} + \hat{\xi}_{i^*,h,t}, \tag{3.2.3}$$

   where $\hat{\xi}_{i^*,h,t}$ is the least-squares residual. Recall from the previous subsection that if we are estimating an SVAR(p) specification in the data $\check{y}_t$, then the vector y_{t-1} actually contains p lags $\check{y}_{t-1}, \dots, \check{y}_{t-p}$.

2. The *VAR* estimator is defined as the response of $y_{i^*,t+h}$ with respect to the j^*-th recursively orthogonalized innovation, where the magnitude of the innovation is normalized such that $y_{j^*,t}$ increases by one unit on impact:

$$\hat{\delta}_h \equiv e_{i^*,n}' \hat{A}^h \hat{\nu},$$

   where

$$\hat{A} \equiv \left(\sum_{t=2}^{T} y_t y_{t-1}' \right) \left(\sum_{t=2}^{T} y_{t-1} y_{t-1}' \right)^{-1}, \quad \hat{\nu} \equiv \hat{C}_{j^*,j^*}^{-1} \hat{C}_{\bullet,j^*},$$

and $\hat{C}_{\bullet,j^*}$ is the j^*-th column of the lower triangular Cholesky factor $\hat{C}$ of the covariance matrix $\hat{\Sigma} \equiv \frac{1}{T}\sum_{t=1}^{T} \hat{u}_t \hat{u}_t' = \hat{C}\hat{C}'$ of the residuals $\hat{u}_t \equiv y_t - \hat{A}y_{t-1}$. Again, in the case of an SVAR(p) specification, the above formulae operate on the companion form.

Note that the two estimators coincide at the impact horizon: $\hat{\beta}_0 = \hat{\delta}_0$ (see Lemma C.5.5 in Appendix C.5).

It is well known that conventional confidence intervals based on both these estimators would have correct asymptotic coverage in a well-specified VAR model. However, the presence of the additional MA term in the model (3.2.1) means that, in principle, both the LP and VAR estimators ought to control for infinitely many lags of the data, rather than just one. Nevertheless, as we will now establish, this dynamic misspecification has much more serious consequences for the VAR procedure than for LP.

3.3 Robust local projections, fragile VARs

This section shows that the conventional LP confidence interval is robust to large amounts of misspecification. In contrast, the conventional VAR confidence interval has fragile coverage, except when it is asymptotically as wide as the LP interval, as will be the case with sufficiently large lag length.

3.3.1 Large-sample distributions and confidence interval coverage

We begin by characterizing the large-sample distributions of the LP and VAR estimators.

The robustness of LPs. Our first main result establishes that the large-sample distribution of the LP estimator is unaffected by large amounts of misspecification.

Proposition 3.3.1. *Under Assumption 3.2.1,*

$$\hat{\beta}_h - \theta_{h,T} = \frac{1}{\sigma_{j^*}^2}\frac{1}{T}\sum_{t=1}^{T} \xi_{i^*,h,,t}\varepsilon_{j^*,t} + O_p(T^{-2\zeta}) + o_p(T^{-1/2}),$$

where

$$\xi_{h,t} = (\xi_{1,h,t}, \ldots, \xi_{n,h,t})' \equiv A^h \overline{H}_{j^*} \overline{\varepsilon}_{j^*,t} + \sum_{\ell=1}^{h} A^{h-\ell} H \varepsilon_{t+\ell},$$

with $\overline{H}_{j^*} \equiv (H_{\bullet,j^*+1}, \ldots, H_{\bullet,m})$ *and* $\overline{\varepsilon}_{j^*,t} \equiv (\varepsilon_{j^*+1,t}, \ldots, \varepsilon_{m,t})'$.

Proof. See Section C.2.1. □

The result implies that the first-order asymptotic behavior of LP does not depend on the misspecification parameter $\alpha(L)$, provided $\zeta > 1/4$ so that $O_p(T^{-2\zeta}) = o_p(T^{-1/2})$. Though this robustness property of LP is with respect to local (i.e., asymptotically vanishing) misspecification, it is still quantitatively meaningful, given that MA terms of order $T^{-\zeta}$ with $\zeta \in (1/4, 1/2)$ in the model (3.2.1) can be detected with probability 1 asymptotically by conventional VAR model specification tests, such as the Hausman test considered in Section 3.3.2.

Why is LP robust to misspecification of such large magnitude? We will offer two mathematically equivalent pieces of intuition, with our discussion throughout deliberately heuristic. The classic omitted variable bias (OVB) formula suggests that the bias of the LP impulse response estimator $\hat{\beta}_h$ in the regression (3.2.3) is proportional to the product of two factors: (i) the direct effect of omitted lags on $y_{i^*,t+h}$, and (ii) the covariance of the residualized regressor of interest $y_{j^*,t} - E[y_{j^*,t} \mid \underline{y}_{j^*,t}, y_{t-1}]$ with the omitted lags. The factor (i) is of order $T^{-\zeta}$ in our local-to-SVAR model (3.2.1). The factor (ii) is also of order $T^{-\zeta}$, since the residualized regressor equals $\varepsilon_{j^*,t} + O_p(T^{-\zeta})$ under Assumption 3.2.1(iii)), and the shock $\varepsilon_{j^*,t}$ is uncorrelated with any lagged data. Hence, the OVB is of order $T^{-2\zeta}$, so when $\zeta > 1/4$, the bias of the estimator is negligible relative to the standard deviation (which is of order $T^{-1/2}$, as in the correctly specified case). This argument relies on the LP regression controlling for the most important lags of the data (i.e., y_{t-1}); without lagged controls, one or both factors in the OVB formula may not be small (González-Casasús and Schorfheide, 2025).

The preceding intuition is a special case of the *double robustness* property of partially linear regressions, see Example 1.1 in Chernozhukov et al. (2018) and Example 1 in Chernozhukov et al. (2022). We will now argue that this property applies also to LP, again settling for a heuristic argument. For notational simplicity, set $j^* = 1$ so $\underline{y}_{j^*,t} = 0$. Consider any dynamic model (for example a VARMA(p, q)) that implies the following LP representation:

$$y_{i^*,t+h} = \theta_{0,h} y_{1,t} + \gamma_0(y^{t-1}) + \xi_{i^*,h,t}, \quad \text{where} \quad \xi_{i^*,h,t} \perp\!\!\!\perp y^t \equiv (y_t, y_{t-1}, \dots).$$

Here $\theta_{0,h}$ is the true impulse response, $\gamma_0(\cdot)$ is a function of lagged data, and "$\perp\!\!\!\perp$" signifies independence. Define $\nu_0(y^{t-1}) \equiv E[y_{1,t} \mid y^{t-1}]$. By applying the Frisch-Waugh lemma to the regression (3.2.3), we see that the LP estimator $\hat{\beta}_h$ is the sample analogue of the solution $\theta_{0,h}$ to the moment condition

$$E[\{y_{i^*,t+h} - \theta_{0,h} y_{1,t} - \gamma_0(y^{t-1})\}\{y_{1,t} - \nu_0(y^{t-1})\}] = 0.$$

If we evaluate the moment on the left-hand side at arbitrary functions $\gamma(\cdot)$ and $\nu(\cdot)$ rather than at the true ones $\gamma_0(\cdot)$ and $\nu_0(\cdot)$, a simple calculation shows that it equals $E[\{\gamma_0(y^{t-1}) - \gamma(y^{t-1})\}\{\nu_0(y^{t-1}) - \nu(y^{t-1})\}]$.[4] Hence, the moment condition is satisfied at the true impulse response parameter $\theta_{0,h}$ as long as *either* $\gamma = \gamma_0$ or $\nu = \nu_0$, making the LP estimator *doubly robust*: it is consistent if we correctly specify either the controls $\gamma(y^{t-1})$ in the outcome equation or the controls $\nu(y^{t-1})$ in the implicit first-stage regression that isolates the shock $\varepsilon_{j^*,t} = y_{j^*,t} - \nu(y^{t-1})$. Because of double robustness, and as argued more generally by Chernozhukov et al. (2018) (and confirmed by our proof), it turns out that estimation error in γ_0 and ν_0 only affects the asymptotic distribution of $\hat{\beta}_h$ through the *product* of the estimation errors $\|\hat{\gamma} - \gamma_0\| \times \|\hat{\nu} - \nu_0\|$. In our local-to-SVAR model (3.2.1), both terms in this product

[4] We can write the moment as $E[\{y_{i^*,t+h} - \theta_{0,h} y_{1,t} - \gamma_0(y^{t-1}) + \gamma_0(y^{t-1}) - \gamma(y^{t-1})\}\{y_{1,t} - \nu(y^{t-1})\}] = E[\{\gamma_0(y^{t-1}) - \gamma(y^{t-1})\}\{y_{1,t} - \nu_0(y^{t-1}) + \nu_0(y^{t-1}) - \nu(y^{t-1})\}]$, since $y_{i^*,t+h} - \theta_{0,h} y_{1,t} - \gamma_0(y^{t-1}) = \xi_{i^*,h,t}$ is independent of y^t (orthogonality would suffice if $\nu(\cdot)$ were linear). The claim now follows from $E[y_{1,t} - \nu_0(y^{t-1}) \mid y^{t-1}] = 0$ by definition of $\nu_0(\cdot)$.

are of order $T^{-\zeta}$ due to the omitted lags. The product is then of order $T^{-2\zeta}$ and thus asymptotically negligible when $\zeta > 1/4$, consistent with our earlier intuition.

The fragility of VARs. In contrast to LP, the VAR estimator is fragile.

Proposition 3.3.2. *Under Assumption 3.2.1,*

$$\hat{\delta}_h - \theta_{h,T} = \operatorname{tr}\left\{S^{-1}\Psi_h H T^{-1}\sum_{t=1}^{T}\varepsilon_t\tilde{y}'_{t-1}\right\} + \frac{1}{\sigma_{j^*}^2}e'_{i^*,n}A^h T^{-1}\sum_{t=1}^{T}\xi_{0,t}\varepsilon_{j^*,t}$$
$$+ T^{-\zeta}\operatorname{aBias}(\hat{\delta}_h) + o_p(T^{-1/2} + T^{-\zeta}),$$

where

$$\operatorname{aBias}(\hat{\delta}_h) \equiv \operatorname{tr}\left\{S^{-1}\Psi_h H\sum_{\ell=1}^{\infty}\alpha_\ell D H'(A')^{\ell-1}\right\} - e'_{i^*,n}\sum_{\ell=1}^{h}A^{h-\ell}H\alpha_\ell e_{j^*,m},$$

$$\Psi_h \equiv \sum_{\ell=1}^{h}A^{h-\ell}H_{\bullet,j^*}e'_{i^*,n}A^{\ell-1},$$

and $\{\tilde{y}_t\}$ and S are defined in Assumption 3.2.1.

Proof. See Section C.2.2. □

The convergence rate $T^{-\min\{1/2,\zeta\}}$ of the VAR estimator is weakly slower than the $T^{-\min\{1/2,2\zeta\}}$ rate achieved by LP. This is because the VAR estimator suffers from bias of order $T^{-\zeta}$, while the stochastic terms of order $T^{-1/2}$ are the same as they would be in a correctly specified SVAR(p) model.[5] The VAR bias is only asymptotically negligible if $\zeta > 1/2$, a much smaller degree of robustness than shown above for LP. The case $\zeta = 1/2$ is of particular interest, as then the bias and standard deviation are of the same asymptotic order (see also Schorfheide, 2005). MA terms of order $T^{-1/2}$ can be detected with asymptotic probability strictly between 0 and 1 by specification tests, as will be shown in Section 3.3.2.

The asymptotic bias is due to two forces: first, the coefficient matrix $\hat{A}$ is biased due to the endogeneity caused by the MA terms, and second, the VAR estimator extrapolates the

[5]The first stochastic term captures sampling uncertainty in the reduced-form impulse responses $\hat{A}^h$, while the second term captures uncertainty in the structural impact response vector $\hat{\nu}$.

horizon-h impulse response based on a parametric formula $\hat{A}^h$ that does not hold exactly in the true VARMA model (3.2.1). This is more easily seen in the special case of a univariate model $y_t = \rho y_{t-1} + [1 + T^{-\zeta}\alpha(L)]\varepsilon_t$ with $n = m = 1$, in which case

$$\text{aBias}(\hat{\delta}_h) \equiv \underbrace{h\rho^{h-1}}_{\frac{\partial(\rho^h)}{\partial\rho}} \underbrace{(1-\rho^2)\sum_{\ell=1}^{\infty}\rho^{\ell-1}\alpha_\ell}_{\text{aBias}(\hat{\rho})=\frac{\text{cov}(\alpha(L)\varepsilon_t,\tilde{y}_{t-1})}{\text{var}(\tilde{y}_{t-1})}} - \underbrace{\sum_{\ell=1}^{h}\rho^{h-\ell}\alpha_\ell}_{\theta_{h,T}-\rho^h},$$

where $\hat{\rho} = \hat{A}$ is the AR(1) coefficient from an OLS regression of y_t on y_{t-1}.[6] While the dynamic responses estimated by the VAR are prone to bias, the shock identification *per se* is not, in the sense that the VAR's estimated impact (horizon-0) response is identical to the doubly robust LP estimate (see Section 3.2.2).

Confidence intervals. The preceding results imply that the conventional LP confidence interval is robust to misspecification while the conventional VAR interval is not. We define the level-$(1-a)$ LP and VAR confidence intervals using the standard formulae:

$$\text{CI}(\hat{\beta}_h) \equiv \left[\hat{\beta}_h \pm z_{1-a/2}\sqrt{\text{aVar}(\hat{\beta}_h)/T}\right], \quad \text{CI}(\hat{\delta}_h) \equiv \left[\hat{\delta}_h \pm z_{1-a/2}\sqrt{\text{aVar}(\hat{\delta}_h)/T}\right]. \tag{3.3.1}$$

Here $z_{1-a/2}$ is the $1-a/2$ quantile of the standard normal distribution, and $\text{aVar}(\hat{\beta}_h)$ and $\text{aVar}(\hat{\delta}_h)$ are the asymptotic variances of the leading (order-$T^{-1/2}$) stochastic terms in the representations of the LP and VAR estimators in Propositions 3.3.1 and 3.3.2; explicit formulae for the asymptotic variances are given in Corollary C.1.2 in Section C.1.3, which also implies that $\text{aVar}(\hat{\beta}_h) \geq \text{aVar}(\hat{\delta}_h)$. None of the results below would change if we replaced the asymptotic variances with the conventional consistent estimates of these (that assume correct specification, as implemented in standard econometric software packages).[7]

[6] Lag augmentation of the VAR impulse response estimator as in Inoue and Kilian (2020) may reduce the first term in the bias formula, but it does not affect the second term.

[7] Under Assumption 3.2.1, homoskedastic standard errors suffice. For LP, Heteroskedasticity and Autocorrelation Robust inference would generally be required under the weaker assumptions in Appendix C.3.1, though simple heteroskedasticity-robust standard errors suffice under the assumptions discussed by Montiel Olea and Plagborg-Møller (2021) and Xu (2023).

Corollary 3.3.1. *Under Assumption 3.2.1 and* $\zeta > 1/4$, $\lim_{T\to\infty} P(\theta_{h,T} \in \mathrm{CI}(\hat{\beta}_h)) = 1 - a$. *If moreover* $\mathrm{aVar}(\hat{\delta}_h) > 0$ *and* $\mathrm{aBias}(\hat{\delta}_h) \neq 0$, *then* $\lim_{T\to\infty} P(\theta_{h,T} \in \mathrm{CI}(\hat{\delta}_h)) = \lim_{T\to\infty}\{1 - r\left(T^{1/2-\zeta} b_h; z_{1-a/2}\right)\}$, *where* $b_h \equiv \mathrm{aBias}(\hat{\delta}_h)/\sqrt{\mathrm{aVar}(\hat{\delta}_h)}$, $r(b;c) \equiv P_{Z\sim N(0,1)}(|Z+b| > c) = \Phi(-c-b) + \Phi(-c+b)$, *and* $\Phi(\cdot)$ *is the standard normal distribution function.*

Proof. Considering separately the three cases $\zeta \in (1/4, 1/2)$, $\zeta = 1/2$, and $\zeta > 1/2$, the result is an immediate consequence of Propositions 3.3.1 and 3.3.2. □

LP robustly controls coverage when $\zeta > 1/4$, while the VAR confidence interval generically has coverage converging to zero for $\zeta \in (1/4, 1/2)$, and strictly below the nominal level $1 - a$ for $\zeta = 1/2$. Intuitively, the VAR confidence interval has the right width (the same as in the correctly specified case) but the wrong location due to the bias.

3.3.2 Hausman misspecification test

To aid in interpreting the magnitude of the local misspecification in our set-up, we consider a Hausman (1978) test of correct specification of the VAR model that compares the VAR and LP impulse response estimates. This test rejects for large values of $\sqrt{T}|\hat{\beta}_h - \hat{\delta}_h|/\sqrt{\mathrm{aVar}(\hat{\beta}_h) - \mathrm{aVar}(\hat{\delta}_h)}$. A test of this kind was proposed by Stock and Watson (2018) in the context of testing for invertibility.

Proposition 3.3.3. *Impose Assumption 3.2.1,* $\zeta > 1/4$, *and* $\mathrm{aVar}(\hat{\beta}_h) > \mathrm{aVar}(\hat{\delta}_h) > 0$. *Then the asymptotic rejection probability of the Hausman test equals*

$$\lim_{T\to\infty} P\left(\frac{\sqrt{T}|\hat{\beta}_h - \hat{\delta}_h|}{\sqrt{\mathrm{aVar}(\hat{\beta}_h) - \mathrm{aVar}(\hat{\delta}_h)}} > z_{1-a/2}\right) = \lim_{T\to\infty} r\left(\frac{T^{1/2-\zeta} b_h}{\sqrt{\mathrm{aVar}(\hat{\beta}_h)/\,\mathrm{aVar}(\hat{\delta}_h) - 1}}; z_{1-a/2}\right),$$

where b_h *and* $r(\cdot,\cdot)$ *were defined in Corollary 3.3.1.*

Proof. Considering separately the three cases $\zeta \in (1/4, 1/2)$, $\zeta = 1/2$, and $\zeta > 1/2$, the result follows from Propositions 3.3.1 and 3.3.2 as well as Corollary C.1.2 in Section C.1.3. □

As claimed previously, the Hausman test is consistent against MA misspecification of order $T^{-\zeta}$ with $\zeta \in (1/4, 1/2)$, except in the knife-edge case where $\text{aBias}(\hat{\delta}_h) = 0$. When $\zeta = 1/2$ and $\text{aBias}(\hat{\delta}_h) \neq 0$, the asymptotic rejection probability is strictly between the significance level a and 1. In Section 3.4 we will use the Hausman test to quantify the difficulty of detecting especially pernicious types of model misspecification.

3.3.3 Lag length selection

We now elaborate further on the role of the estimation lag length. We first discuss the properties of LP when the lag length is selected using standard information criteria. We then show that increasing the lag length robustifies VAR inference by rendering it equivalent with LP inference.

Lag length selection for LP. Appendix C.3.3 shows that the LP confidence interval maintains correct asymptotic coverage when the lag length p is selected via the Bayesian Information Criterion (BIC), provided that $\zeta \geq 1/2$. Specifically, the BIC should be applied to an auxiliary VAR in the observed data series $\check{y}_t$; the selected lag length then determines the number p of lags to control for in subsequent LP inference (which otherwise discards the auxiliary VAR). The resulting LP inference is robust to model selection errors that are known to cause difficulties for VAR inference (Leeb and Pötscher, 2005; Kilian and Lütkepohl, 2017, chapter 2.6.5). At a high level, the greater reliability of LP inference with data-dependent lag length is a consequence of the double robustness discussed earlier, see Belloni and Chernozhukov (2013) and Chernozhukov et al. (2018).

While the BIC suffices in theory for valid local projection inference, we follow Kilian and Lütkepohl (2017) and recommend that researchers employ the more conservative Akaike Information Criterion (AIC) in practice. The reason is that, in finite samples, the downsides of under-specifying the lag length outweigh the slight inefficiency associated with over-selecting the lag length. Section 3.5 demonstrates that this lag length selection procedure delivers LP confidence intervals with accurate coverage in realistic DGPs.

Long-lag VARs. One simple way to remove the asymptotic bias of the VAR estimator is to control for sufficiently many lags—typically many more lags than indicated by conventional information criteria. This is because in this case the estimator is asymptotically equivalent with the LP estimator. See Plagborg-Møller and Wolf (2021) and Xu (2023) for related results in models without explicit MA misspecification.

Corollary 3.3.2. *Suppose the model* (3.2.2) *written in companion form* (3.2.1) *satisfies Assumption 3.2.1 and* $\zeta > 1/4$. *Let* $\tilde{\check{y}}_t$ *denote the stationary solution to equation* (3.2.2) *when* $\alpha(L) = 0$. *If* $\varepsilon_{j^*,t-\ell} \in \operatorname{span}(\tilde{\check{y}}_{t-1}, \dots, \tilde{\check{y}}_{t-p})$ *for all* $\ell = 1, \dots, h$, *then* $\operatorname{aBias}(\hat{\delta}_h) = 0$ *and* $\operatorname{aVar}(\hat{\delta}_h) = \operatorname{aVar}(\hat{\beta}_h)$. *In particular, these results obtain if either of the following two sufficient conditions hold:*

i) The model is a local-to-SVAR(p_0) model (i.e., $\check{A}_\ell = 0$ *for* $p_0 < \ell \leq p$*) and* $h \leq p - p_0$, *where* p *is the estimation lag length.*

ii) The shock of interest is directly observed and ordered first (i.e., $j^* = 1$ *and* $\check{A}_{1,j,\ell} = 0$ *for all* j, ℓ*), and* $h \leq p$.

Proof. See Section C.2.3. □

We see that, the larger the impulse horizon h of interest, the larger is the estimation lag length p required for bias reduction. In fact, Section 3.4 shows that the *only* way to guarantee that the asymptotic bias of the VAR estimator is zero is to control for so many lags that LP and VAR are asymptotically equivalent.

3.4 VAR inference under bounded misspecification

To show that the fragility of VARs is likely to matter in practice, we now investigate the worst-case properties of VAR procedures under a tight constraint on the amount of misspecification.

We prove that the conventional VAR confidence interval is robust if, *and only if*, LP and VAR intervals coincide asymptotically. VARs with short-to-moderate lag lengths suffer from severe coverage distortions even for small amounts of misspecification that are hard to rule out either economically or statistically. Beyond increasing the lag length, an alternative strategy to fix VAR undercoverage is to use a larger bias-aware critical value; however, we show that the resulting confidence interval is usually wider than the LP interval. Finally, we show that all conclusions extend to the case of joint inference on multiple impulse responses.

Throughout this section we set $\zeta = 1/2$ so that the asymptotic bias-variance trade-off between LP and VAR is non-trivial.

3.4.1 Worst-case bias and mean-squared error

Building towards our main results on VAR coverage distortions, we begin by deriving the worst-case bias and mean-squared error of the VAR estimator.

Misspecification bound. To quantify the amount of misspecification in the local-to-SVAR model (3.2.1) with $\zeta = 1/2$, we define the *noise-to-signal ratio*

$$\operatorname{tr}\left\{\operatorname{var}(T^{-1/2}\alpha(L)\varepsilon_t)\operatorname{var}(\varepsilon_t)^{-1}\right\} = \operatorname{tr}\left\{\left(T^{-1}\sum_{\ell=1}^{\infty}\alpha_\ell D\alpha_\ell'\right)D^{-1}\right\} = T^{-1}\|\alpha(L)\|^2,$$

where we define the norm $\|\alpha(L)\| \equiv \sqrt{\sum_{\ell=1}^{\infty}\operatorname{tr}\{D\alpha_\ell' D^{-1}\alpha_\ell\}}$. Suppose we are willing to impose *a priori* that the noise-to-signal ratio is at most M^2/T for some constant $M \in (0,\infty)$. For small M^2/T, this roughly means that a fraction M^2/T of the variance of the model's error term is due to the misspecification. This corresponds to restricting the parameter space for $\alpha(L)$ to all absolutely summable lag polynomials that satisfy $\|\alpha(L)\| \leq M$. In the following we will consider the worst-case properties of the VAR estimator over this parameter space, treating the other (consistently estimable) parameters (A, H, D) as fixed.

Worst-case bias.

Proposition 3.4.1. *Impose Assumption 3.2.1, $\zeta = 1/2$, and* $\text{aVar}(\hat{\delta}_h) > 0$*. Then*

$$\max_{\alpha(L)\colon \|\alpha(L)\| \le M} |b_h| = M \sqrt{\frac{\text{aVar}(\hat{\beta}_h)}{\text{aVar}(\hat{\delta}_h)} - 1},$$

where we recall the definition $b_h = \text{aBias}(\hat{\delta}_h)/\sqrt{\text{aVar}(\hat{\delta}_h)}$*. Recall also that* $\text{aVar}(\hat{\beta}_h)$ *and* $\text{aVar}(\hat{\delta}_h)$ *do not depend on* $\alpha(L)$*.*

Proof. The claim is a special case of Proposition C.3.2 in Appendix C.3.4. □

Under our bound M^2/T on the noise-to-signal ratio, the worst-case (scaled) VAR bias is a simple function of M and of the relative asymptotic precision $\text{aVar}(\hat{\beta}_h)/\,\text{aVar}(\hat{\delta}_h)$ of the VAR estimator vs. LP. These two quantities are "sufficient statistics" for the worst-case bias regardless of the number n of variables in the VAR, the lag length p, the specific VAR parameters (A, H, D), and the horizon h. Hence, our subsequent analysis of the worst-case properties of VAR procedures depends only on M and on the relative precision, allowing us to concisely present analytical results that cover a wide range of local-to-SVAR models without having to resort to simulations that inevitably only cover a finite number of DGPs.

Proposition 3.4.1 shows that VAR estimators must trade off efficiency and robustness: the worst-case VAR bias is small precisely when the VAR estimator has nearly the same variance as LP. While the worst-case bias can be reduced by increasing the VAR estimation lag length p, the proposition shows that this can *only* happen at the expense of increasing the variance. If we include so many lags that the worst-case bias is zero (cf. Corollary 3.3.2), then the VAR estimator must *necessarily* be asymptotically equivalent with LP.

Worst-case mean squared error. For future reference we briefly discuss how the worst-case mean squared error (MSE) of the VAR estimator depends on the imposed bound on misspecification. Based on Propositions 3.3.1 and 3.3.2 as well as Corollary C.1.2, we define the asymptotic MSE of the VAR and LP estimators as follows:

$$\text{aMSE}(\hat{\beta}_h) \equiv \text{aVar}(\hat{\beta}_h), \quad \text{aMSE}(\hat{\delta}_h) \equiv \text{aBias}(\hat{\delta}_h)^2 + \text{aVar}(\hat{\delta}_h).$$

Corollary 3.4.1. *Impose Assumption 3.2.1 and $\zeta = 1/2$. Then*

$$\sup_{\alpha(L)\colon \|\alpha(L)\| \leq M} \{\text{aMSE}(\hat{\delta}_h) - \text{aMSE}(\hat{\beta}_h)\} = (M^2 - 1)\{\text{aVar}(\hat{\beta}_h) - \text{aVar}(\hat{\delta}_h)\}.$$

Proof. See Section C.2.4. □

In words, the worst-case MSE regret of VAR relative to LP is proportional to the variance reduction of VAR relative to LP, with a proportionality constant of $M^2 - 1$. If $M > 1$ (corresponding to a noise-to-signal ratio greater than $1/T$), the worst-case MSE of VAR thus strictly exceeds the MSE of LP. From here it is also straightforward to recover the minimax optimal way to average LP and VAR estimates.

Corollary 3.4.2. *Impose Assumption 3.2.1, $\zeta = 1/2$, and* $\text{aVar}(\hat{\beta}_h) > \text{aVar}(\hat{\delta}_h)$*. Consider the model-averaging estimator $\hat{\theta}_h(\omega) \equiv \omega\hat{\beta}_h + (1-\omega)\hat{\delta}_h$, and denote its asymptotic MSE by* $\text{aMSE}(\hat{\theta}_h(\omega))$*. Then*

$$\underset{\omega \in \mathbb{R}}{\text{argmin}} \sup_{\alpha(L)\colon \|\alpha(L)\| \leq M} \text{aMSE}(\hat{\theta}_h(\omega)) = \frac{M^2}{1 + M^2}.$$

Proof. See Section C.2.5. □

If $M = 1$, it is minimax optimal to weight the LP and VAR estimates equally. If $M = 2$ (corresponding to a noise-to-signal ratio of $4/T$), the LP estimator receives 80% weight.

3.4.2 Worst-case coverage

We now turn to our main area of interest: the worst-case asymptotic coverage of the conventional VAR confidence interval under our bound on the amount of misspecification. This turns out to take a very simple form.

Corollary 3.4.3. *Impose Assumption 3.2.1, $\zeta = 1/2$, and $\operatorname{aVar}(\hat{\delta}_h) > 0$. Then*

$$\inf_{\alpha(L)\colon \|\alpha(L)\| \leq M} \lim_{T\to\infty} P(\theta_{h,T} \in \mathrm{CI}(\hat{\delta}_h)) = 1 - r\left(M\sqrt{\operatorname{aVar}(\hat{\beta}_h)/\operatorname{aVar}(\hat{\delta}_h) - 1};\, z_{1-a/2}\right).$$

Proof. This is an immediate consequence of Corollary 3.3.1 and Proposition 3.4.1. □

Based on this corollary, Figure 3.4.1 provides a complete characterization of the robustness-efficiency trade-off for VAR confidence intervals. It plots the worst-case coverage probability as a function of the ratio of standard errors for VAR and LP, given significance level $a = 10\%$ and different values of M. The shaded area depicts an empirically relevant range of standard error ratios obtained in four empirical applications from Ramey (2016).[8] We see that, even for $M = 1$ (corresponding to a noise-to-signal ratio of $1/T$), the worst-case coverage probability is below 48% whenever the asymptotic standard deviation of the VAR estimator is less than half that of LP—a value that is typical in applied work. Further, at the bottom end of the empirically relevant range, the worst-case coverage probability is essentially zero as soon as $M \geq 1$. It is only at the very right side of the figure—when the VAR includes enough lags to remove nearly all bias, thus increasing the standard error almost to that of LP—that the VAR confidence interval has coverage close to the nominal level.

The potential for VAR undercoverage documented here may not be so concerning if the worst-case misspecification can be ruled out on economic theory grounds, or if it is easily detectable statistically. We now argue that neither appears to be the case.

[8] We replicate Ramey's identification schemes for monetary, tax, government spending, and technology shocks. The shaded area shows the 10th to 90th percentiles of standard error ratios at horizons exceeding 1 year. See the online replication materials for details.

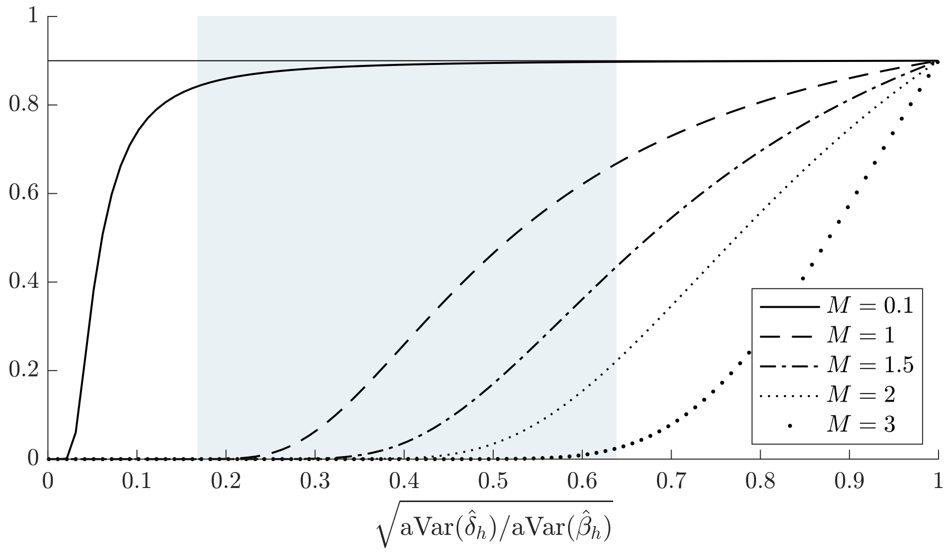

Figure 3.4.1: Worst-case asymptotic coverage probability of the conventional 90% VAR confidence interval. Horizontal axis: relative asymptotic standard deviation of VAR vs. LP. Different lines: different bounds M on $\|\alpha(L)\|$. Shaded area: empirical 10th–90th percentile range of relative standard errors based on Ramey (2016), see the online replication materials for details. The solid horizontal line marks the nominal coverage probability $1 - a = 90\%$.

Economic theory. The shape and magnitude of the least favorable misspecification is difficult to rule out generally based on economic theory. The least favorable MA polynomial $\alpha^\dagger(L; h, M) = \sum_{\ell=1}^{\infty} \alpha^\dagger_{\ell,h,M} L^\ell$ for VAR coverage is the same as the least favorable one for bias (i.e., the $\alpha(L)$ that achieves the maximum in Proposition 3.4.1). Since $\mathrm{aBias}(\hat{\delta}_h)$ is linear in $\alpha(L)$, the least favorable choice given the constraint $\|\alpha(L)\| \leq M$ follows from the Cauchy-Schwarz inequality (see the proof of Proposition C.3.2 in Appendix C.3):

$$\alpha^\dagger_{\ell,h,M} \propto D^{1/2} H' \Psi_h' S^{-1} A^{\ell-1} H D^{1/2} - \mathbb{1}(\ell \leq h) \sigma_{j^*}^{-1} D^{1/2} H' (A')^{h-\ell} e_{i^*,n} e_{j^*,m}', \quad \ell \geq 1, \quad (3.4.1)$$

where the constant of proportionality (which does not depend on the lag ℓ) is chosen so that $\|\alpha^\dagger(L; h, M)\| = M$. Note that the shape of the least favorable MA polynomial depends on the particular horizon h of interest but not on M; i.e., the bound M^2/T on the noise-to-signal ratio only scales the polynomial up or down.

We note two main properties of the least favorable misspecification. First, the magnitude of the MA coefficients $\alpha^\dagger_{\ell,h,M}$ decays exponentially as $\ell \to \infty$. In other words, not only is the overall magnitude of the least favorable model misspecification small (as imposed in the noise-to-signal bound), the MA coefficients at long lags are in fact particularly small. Second, numerical examples shown in Section C.1.1 suggest that the MA coefficients tend to be largest in magnitude at horizon h, displaying either a hump-shaped pattern as a function of ℓ—consistent with economic theories of adjustment costs or learning—or a single zig-zag pattern—consistent with theories of overshooting or lumpy adjustment. We thus view MA dynamics of the worst-case form as empirically and theoretically relevant.[9]

Statistical tests. The least favorable misspecification is also difficult to detect statistically. Propositions 3.3.3 and 3.4.1 imply that, for $\alpha(L) = \alpha^\dagger(L; h, M)$, the asymptotic rejection probability of the Hausman test of correct VAR specification equals $r(M; z_{1-a/2})$. When $M = 1$ (corresponding to a noise-to-signal ratio of $1/T$), the odds of the Hausman test *failing* to reject the misspecification are nearly 3-to-1 at significance level $a = 10\%$, since $r(1; z_{0.95}) = 26\%$. At significance level $a = 5\%$, the odds are nearly 5-to-1, since $r(1; z_{0.975}) = 17\%$. Standard *ex post* model misspecification tests are thus unlikely to indicate a problem even if the potential for undercoverage is severe.

Rather than committing *a priori* to a parameter space for $\alpha(L)$ through choice of M, we can also ask a different question: across all possible types and magnitudes of misspecification, what is the worst-case probability that the conventional VAR confidence interval fails to cover the true impulse response, yet we fail to reject correct specification of the VAR model?

Corollary 3.4.4. *Impose Assumption 3.2.1, $\zeta = 1/2$, and $\operatorname{aVar}(\hat{\beta}_h) > \operatorname{aVar}(\hat{\delta}_h) > 0$. Consider the joint event $\mathcal{A}_T$ that $\theta_{h,T} \notin \operatorname{CI}(\hat{\delta}_h)$ and the Hausman test in Proposition 3.3.3 fails to reject misspecification. Then*

$$\sup_{\alpha(L)} \lim_{T\to\infty} P(\mathcal{A}_T) = \sup_{b \geq 0} r(b; z_{1-a/2}) \left\{ 1 - r\left(\frac{b}{\sqrt{\operatorname{aVar}(\hat{\beta}_h)/\operatorname{aVar}(\hat{\delta}_h) - 1}}; z_{1-a/2} \right) \right\},$$

[9] However, the least favorable MA polynomial derived above need not be of interest to researchers who trust that some equations in their SVAR specification are exactly correctly specified, as this imposes the additional restrictions that some linear combinations of the rows of the MA polynomial $\alpha(L)$ equal zero.

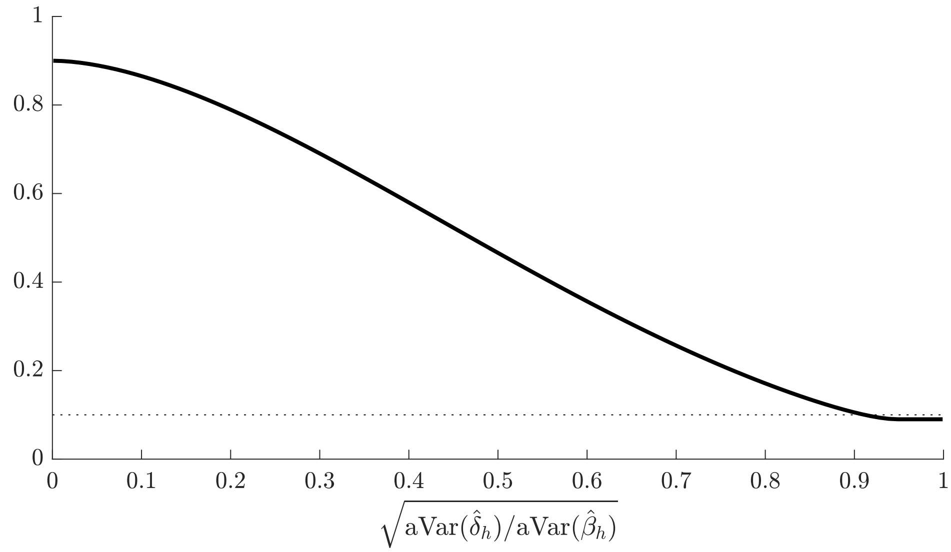

Figure 3.4.2: Worst-case asymptotic probability of the joint event that the conventional VAR confidence interval fails to cover the true impulse response and yet the Hausman test fails to reject misspecification. Horizontal axis: relative asymptotic standard deviation of VAR vs. LP. The dotted horizontal line marks the nominal significance level $a = 10\%$.

where the supremum on the left-hand side is taken over all absolutely summable lag polynomials $\alpha(L)$.

Proof. See Section C.2.6. $\square$

Figure 3.4.2 plots this worst-case probability for a significance level of $a = 10\%$, which by Corollary 3.4.4 depends only on the ratio $\mathrm{aVar}(\hat{\delta}_h)/\,\mathrm{aVar}(\hat{\beta}_h)$. Under correct specification, the probability of the joint event is equal to $a(1-a)$ ($= 9\%$ when $a = 10\%$). With misspecification, the joint probability instead exceeds 46% when the asymptotic standard deviation of the VAR estimator is less than half that of the LP estimator. As $\mathrm{aVar}(\hat{\delta}_h)/\,\mathrm{aVar}(\hat{\beta}_h) \to 0$, the worst-case joint probability approaches $1 - a$. We thus again see that statistical tests may fail to warn against the potential for severe VAR coverage distortions.

3.4.3 Bias-aware inference

Rather than removing bias by increasing the lag length (thus ensuring equivalence with LP), an alternative way to fix the undercoverage of the conventional VAR confidence interval is to adjust the critical value upward to compensate for the bias, as suggested in a general setting by Armstrong and Kolesár (2021). Suppose again that we restrict the misspecification $\alpha(L)$ to satisfy $\|\alpha(L)\| \leq M$. Then we define the *bias-aware* VAR confidence interval

$$\mathrm{CI}_B(\hat{\delta}_h; M) \equiv \left[\hat{\delta}_h \pm \mathrm{cv}_{1-a}\left(M\sqrt{\frac{\mathrm{aVar}(\hat{\beta}_h)}{\mathrm{aVar}(\hat{\delta}_h)} - 1}\right)\sqrt{\mathrm{aVar}(\hat{\delta}_h)/T}\right],$$

where the bias-aware critical value $\mathrm{cv}_{1-a}(b)$ is given by the number such that $r(b; \mathrm{cv}_{1-a}(b)) = a$, and $r(\cdot, \cdot)$ is defined in Corollary 3.3.1. By construction, this bias-aware confidence interval has correct (but potentially conservative) asymptotic coverage.

Corollary 3.4.5. *Impose Assumption 3.2.1, $\zeta = 1/2$, and $\mathrm{aVar}(\hat{\delta}_h) > 0$. Then*

$$\inf_{\alpha(L)\colon \|\alpha(L)\| \leq M} \lim_{T\to\infty} P(\theta_{h,T} \in \mathrm{CI}_B(\hat{\delta}_h; M)) = 1 - a.$$

Proof. The result follows immediately from Propositions 3.3.2 and 3.4.1. □

It turns out, however, that a very tight bound M on the signal-to-noise ratio is required for the bias-aware VAR interval to be shorter than the LP interval. Figure 3.4.3 plots the relative interval length as a function of the relative asymptotic standard deviation of VAR and LP, for a significance level of $a = 10\%$ and for different misspecification bounds M. The figure shows that M has to be quite small—apparently below 1—for the bias-aware VAR length to dominate the LP length regardless of the DGP and horizon. Even for $M = 1.5$, bias-aware VAR is at best only moderately shorter than LP. Finally, for values of M above 2 (corresponding to a noise-to-signal ratio above $4/T$), bias-aware VAR is dominated by LP.

In Section C.1.2 we furthermore show that the conventional LP confidence interval is at worst slightly wider than a more efficient bias-aware confidence interval centered at the model averaging estimator $\hat{\theta}_h(\omega) = \omega\hat{\beta}_h + (1-\omega)\hat{\delta}_h$, introduced in Corollary 3.4.2 above. Even if the weight ω is chosen to optimize confidence interval length, the gains relative to the LP interval are very small when $M \geq 2$ (corresponding to a noise-to-signal ratio above $4/T$).

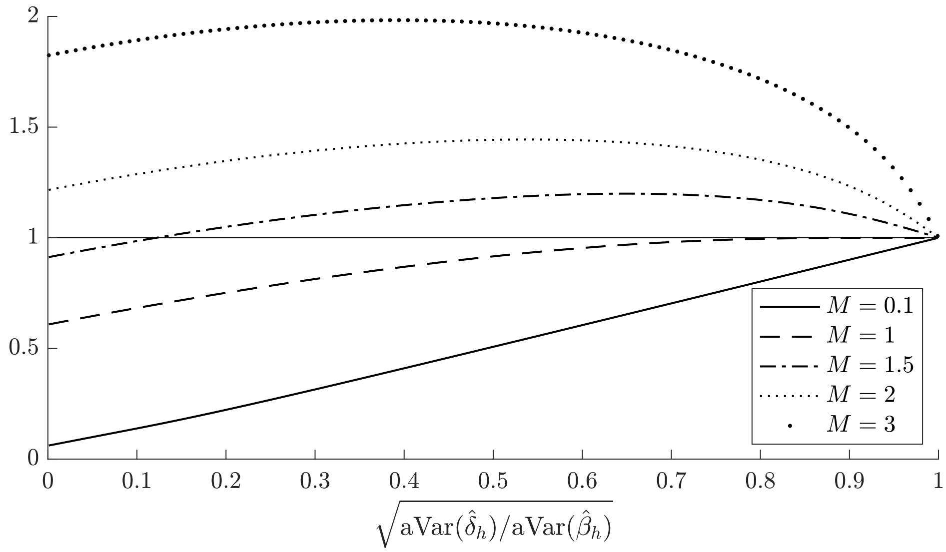

Figure 3.4.3: Relative length of bias-aware VAR confidence interval vs. conventional LP interval. Significance level $a = 10\%$. Horizontal axis: relative asymptotic standard deviation of VAR vs. LP. Different lines: different bounds M on $\|\alpha(L)\|$. The solid horizontal line marks the value 1.

We thus conclude that, while bias-aware VAR inference is possible in theory, in practice the gains relative to the simpler LP interval are small at best, unless we put an extremely tight bound on the noise-to-signal ratio.

3.4.4 Inference on multiple impulse responses

Since the least favorable MA polynomial derived in Section 3.4.2 depends on the horizon h of interest, one might hope that VARs would not be as prone to bias and thereby undercoverage if interest centers on *multiple* impulse responses. Unfortunately, Appendix C.3.4 shows that this is not the case. There we consider inference on a vector of impulse responses for any combination of response variables i, shocks j, and horizons h. Generalizing Proposition 3.4.1, we show that the worst-case norm of the bias is non-negligible if the VAR offers efficiency gains for *any* linear combination of the parameters of interest. This implies that the conventional VAR confidence interval has fragile coverage even if the target parameter is a linear combination of impulse responses (such as the integral or sum across multiple horizons, as in the fiscal multiplier applications reviewed in Ramey, 2016). The conventional Wald confidence ellipsoid centered at the VAR estimator is similarly fragile.

3.5 Practical relevance

This section establishes the practical relevance of our theoretical conclusions by comprehensively reviewing current practice for lag length selection in the applied VAR literature, coupled with a simulation study calibrated to the application in Känzig (2021).

3.5.1 Review of current practice for VAR lag length selection

To evaluate lag length selection practice in the applied VAR literature, we created a comprehensive list of articles published between January 2015 and June 2025 in six top economics journals: *American Economic Review*, *American Economic Journal: Macroeconomics*, *Econometrica*, *Journal of Political Economy*, *Quarterly Journal of Economics*, and *Review of Economic Studies*. We restrict attention to papers that use VARs on time series data (not panel data) to estimate *structural* impulse response functions. This yields 81 papers total. For each paper, we picked a single specification as the "main" one.[10] To be conservative, if multiple specifications received equal attention in the main analysis, we picked the one with the largest lag length. The full list of papers, together with the recorded information about the VAR specifications, is provided in the online replication materials.

Our findings suggest that the theoretical results of the preceding sections are likely to have bite in practice, as typical VAR estimation lag lengths in applied papers are short or moderate. The modal lag length is 4 in quarterly data and 12 in monthly data, the mean is slightly below the mode, and less than 10% of papers employ lag lengths greater than or equal to twice the modal values.[11] 20% of papers select the lag length in a data-dependent way, often using information criteria. On average across papers, the estimation lag length is only 28% as large as the longest reported impulse horizon. We conclude that few applied papers follow the recommendation of Kilian and Lütkepohl (2017, pp. 58–66) to use long lag lengths and avoid information criteria. This suggests that VAR inference results reported in

[10] In a few cases, the main use of VARs in a paper was only reported in the appendix.

[11] Additionally, all the various empirical VAR specifications reported in the handbook chapters of Ramey (2016) and Stock and Watson (2016) are consistent with these patterns.

much of the applied literature could be subject to the fragility we highlighted in the preceding sections. In fact, since around 40% of papers employ Bayesian shrinkage, the estimation bias could be even larger than indicated by the lag length alone.

3.5.2 Empirically calibrated simulation study

We now show through simulations that our asymptotic results are informative about the performance of LPs and VARs in an empirically relevant finite-sample setting when the lag length is selected as in current applied practice. Our DGP is calibrated to the oil news shock application in Känzig (2021).

Set-up. The DGP is a VAR estimated on the dataset of Känzig (2021), using a somewhat longer lag length than that used in the paper. The data series are that paper's oil shock proxy, the real price of oil, world oil production, world oil inventories, world industrial production, U.S. industrial production, and the U.S. consumer price index (CPI). Whereas Känzig employs 12 lags, we estimate a recursively identified VAR(18) by OLS on his data and use this as the simulation DGP, with i.i.d. Gaussian shocks. We do not claim that the VAR(18) DGP is more "realistic" than a VAR(12) estimated on the same data, but we contend that it is desirable that confidence intervals should have reliable coverage in both these DGPs.

The parameters of interest are the impulse responses of CPI to the observed oil shock. To be consistent with our theory, we use an "internal instruments" specification that orders the proxy first in the VAR; this differs from the "external instruments" specification used by Känzig. The sample size is $T = 720$ months. We will entertain different choices of the estimation lag length p for the LP and VAR estimators. We report results for both delta method and bootstrap confidence intervals. Results are based on 10,000 Monte Carlo simulations. See Appendix C.4 for implementation details.

Results. Figure 3.5.1 shows that our theoretical results on LP robustness and VAR fragility are practically relevant. The figure depicts the coverage probabilities (left panel) and median confidence interval length (right panel) for VARs (in red, solid and dashed) and LPs (in blue, solid and dashed). The top panel fixes the estimation lag length at $p = 12$, while the bottom panel selects the lag length by AIC. Both these choices are frequently encountered in the applied literature, as documented earlier. Given these conventional lag lengths, VAR confidence intervals tend to be shorter than LP intervals, but quite materially undercover, with coverage falling below 60% at medium and long horizons. LP instead attains close to the nominal coverage level of 90% throughout, as expected. The mean lag length selected by AIC is 9.7, evidently insufficient to guard against dynamic misspecification.

Appendix C.4 illustrates that the VAR under-coverage can be ameliorated by increasing the lag length beyond what is typically used in current applied practice, at the expense of higher variance, consistent with Section 3.3.3. The supplement also reports that the VAR estimator achieves lower MSE than LP, suggesting that—despite the poor performance of the VAR confidence interval—the larger bias of the VAR estimator may not compromise its usefulness as a *point estimator* (see also Li et al., 2024).

Montiel Olea et al. (2025) find that the qualitative conclusions above extend to a wide range of empirically calibrated simulation DGPs based on richly specified dynamic factor models: VAR confidence intervals fail to adequately control coverage in a sizable fraction of DGPs, while LP confidence intervals robustly maintain accurate coverage. They document that LP is not only more robust to lag length selection, but also to the choice of control variables, consistent with the theory in this paper.

3.6 Conclusion

Our theoretical results suggest the following practical take-aways:

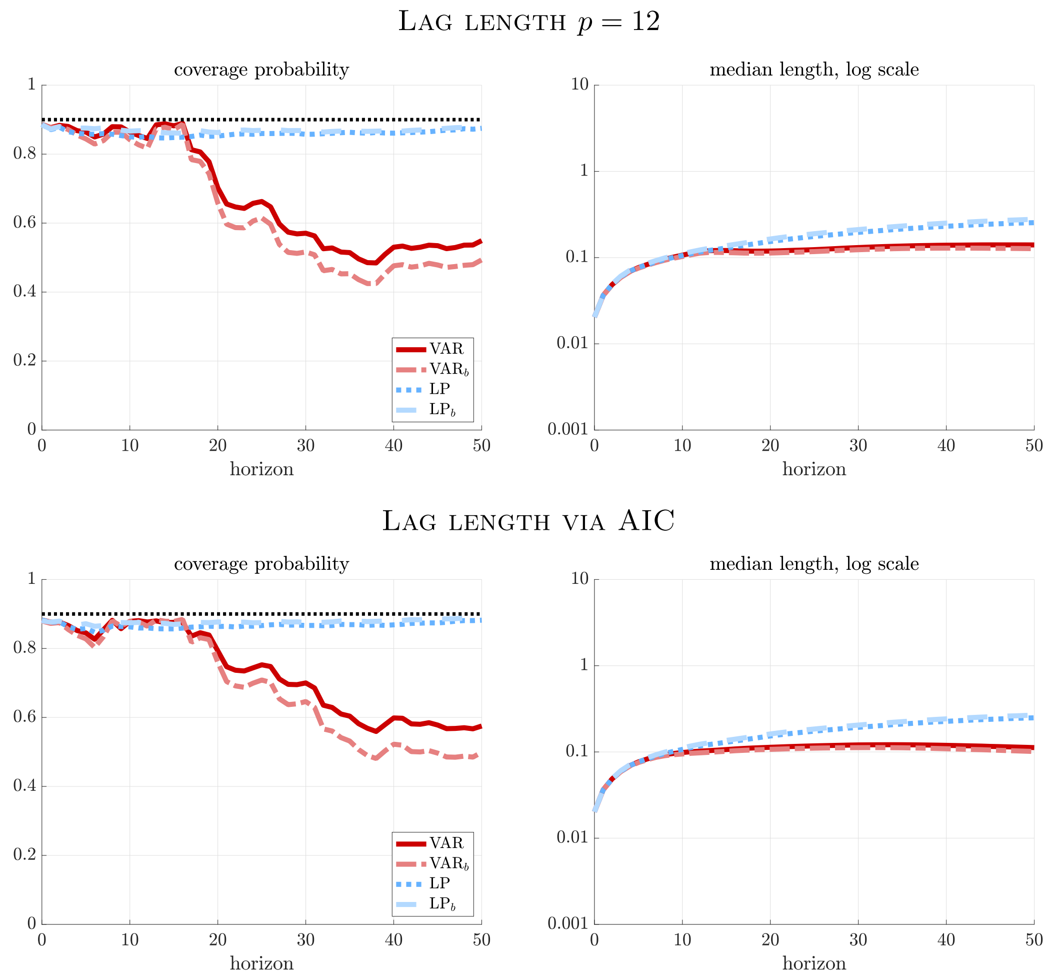

Figure 3.5.1: Coverage probability (left) and median length (right) for VAR (red) and LP (blue) nominal 90% confidence intervals computed via the delta method or bootstrap (the latter are indicated with subscript "b" in the legends). Lag length: fixed at $p = 12$ in the top panel, and selected using AIC in the bottom panel.

1. When the goal is to construct confidence intervals for impulse responses that have accurate coverage in a wide range of empirically relevant DGPs—as opposed to minimizing MSE—then the smaller bias of LPs documented in simulations by Li et al. (2024) is more valuable than the smaller variance enjoyed by VAR estimators.

2. Researchers who use LP should control for those lags of the data that are strong predictors of the outcome or impulse variables. This is important not only when the shock is recursively identified, but even if the researcher directly observes a near-perfect proxy for the shock of interest. However, unlike for VAR inference, it is *not* necessary to get the lag length or set of control variables exactly right to achieve correct coverage. To

select the number of lags to control for in the LP, we recommend running an auxiliary VAR in all variables used in the analysis and selecting the lag length to minimize the AIC; the auxiliary VAR is only used as a device to select the lag length and is otherwise discarded.

3. With the moderate lag lengths typical in current applied practice, VAR confidence intervals will only have accurate coverage at short horizons, and only because they are approximately equivalent with LP intervals at these horizons. If, at some horizons of interest, an estimated VAR yields confidence intervals that are substantially narrower than the corresponding LP intervals, we recommend increasing the VAR lag length until that is no longer the case, to guarantee robust confidence interval coverage. Conventional tests of correct VAR specification do not suffice to guard against coverage distortions.

Is there a way forward for VAR inference, beyond just including a large number of lags? We showed how to construct a VAR confidence interval with a bias-aware critical value that robustly controls coverage, but found that it will typically lead to wider confidence intervals than LP. Another option would be to estimate VARMA models rather than pure VARs, though this would be computationally expensive, and the bias-variance trade-off relative to LPs is unclear. In principle, VAR procedures may work better under additional restrictions on the misspecification, such as shape restrictions on the impulse response functions.[12] However, it appears that detailed application-specific restrictions would be required to generate a negligible worst-case bias, since we have shown that the least favorable misspecification in our baseline analysis cannot generally be ruled out based on economic theory. Rather than restricting the parameter space, future research could instead investigate weakening the coverage requirement, e.g., only requiring a certain coverage probability *on average* over a set of horizons (Armstrong et al., 2022), or by changing the target for inference from the true impulse response function to a smooth projection of this function (Genovese and Wasserman,

[12]Given any convex parameter space for the misspecification MA polynomial $\alpha(L)$, the worst-case bias of the VAR estimator (see Proposition 3.3.2) can be computed using convex programming.

2008). Finally, a subjectivist Bayesian VAR modeler need only worry about our negative results if their prior on potential misspecification attaches significant weight to MA processes that imply large VAR biases.

Appendix A

Appendix to Chapter 1

A.1 Theoretical results

Section A.1.1 contains matrix results used in subsequent proofs. Section A.1.2 contains the assumptions and results for the test of multivariate heterogeneity. Section A.1.3 contains results on semiparametric efficiency.

A.1.1 Matrix results

For this subsection, let $\mathbf{\Sigma}$ be a positive definite matrix where $\mathbf{G}'\mathbf{\Sigma}^{-1}\mathbf{G}$ is invertible and $\mathbf{M} = \mathbf{I} - \mathbf{G}(\mathbf{G}'\mathbf{\Sigma}^{-1}\mathbf{G})^{-1}\mathbf{G}'\mathbf{\Sigma}^{-1}$. Similarly, suppose $\mathbf{W}$ is positive semidefinite, $\mathbf{G}'\mathbf{W}\mathbf{G}$ is invertible, and let $\mathbf{M}_{\mathbf{W}} = \mathbf{I} - \mathbf{G}(\mathbf{G}'\mathbf{W}\mathbf{G})^{-1}\mathbf{G}'\mathbf{W}$. Let † be the generalized matrix inverse.

Proposition A.1.1. $\mathbf{M}'_{\mathbf{W}}(\mathbf{M}_{\mathbf{W}}\mathbf{\Sigma}\mathbf{M}_{\mathbf{W}}')^{\dagger}\mathbf{M}_{\mathbf{W}}$ *is invariant to the choice of* $\mathbf{W}$*.*

Proof. Begin by showing $\text{col}(\mathbf{M}) = \text{col}(\mathbf{G})$. From the rank-nullity theorem, it suffices to show that $\text{null}(\mathbf{M}') = \text{null}(\mathbf{G}')$. Define projection matrix $\mathbf{P}_{\mathbf{W}} = \mathbf{G}(\mathbf{G}'\mathbf{W}\mathbf{G})^{-1}\mathbf{G}'\mathbf{W}$ and consider $\mathbf{x}$ such that $\mathbf{P}'_{\mathbf{W}}\mathbf{x} = 0$. Since $\mathbf{G}'\mathbf{W}\mathbf{G}$ is invertible,

$$\mathbf{W}\mathbf{G}(\mathbf{G}'\mathbf{W}\mathbf{G})^{-1}\mathbf{G}'\mathbf{x} = 0 \implies \mathbf{G}'\mathbf{x} = 0.$$

Now consider $\mathbf{x}$ such that $\mathbf{G}'\mathbf{x} = 0$. Then $\mathbf{P}'_{\mathbf{W}}\mathbf{x} = 0$. Hence, $\text{null}(\mathbf{M_W}') = \text{null}(\mathbf{G}')$ and $\text{col}(\mathbf{M_W}) = \text{col}(\mathbf{G})$.

Since $\text{col}(\mathbf{M_W})$ is invariant to the choice of weight matrix, Theorem 4.8 of Rao and Mitra (1972) implies $\mathbf{M}'_{\mathbf{W}}(\mathbf{M_W}\boldsymbol{\Sigma}\mathbf{M_W}')^{\dagger}\mathbf{M_W}$ is invariant to $\mathbf{W}$. □

Lemma A.1.1. *$\boldsymbol{\Sigma}^{-1}\mathbf{M}$ is a reflexive g-inverse of $\mathbf{M}\boldsymbol{\Sigma}\mathbf{M}'$.*

Proof. $\mathbf{M}\boldsymbol{\Sigma} = \boldsymbol{\Sigma} - \mathbf{G}(\mathbf{G}'\boldsymbol{\Sigma}^{-1}\mathbf{G})^{-1}\mathbf{G}' = \mathbf{M}\boldsymbol{\Sigma}\mathbf{M}'$. Then, from idempotency of $\mathbf{M}$, the g-inverse and reflexive properties hold

$$(\mathbf{M}\boldsymbol{\Sigma}\mathbf{M}')\boldsymbol{\Sigma}^{-1}\mathbf{M}(\mathbf{M}\boldsymbol{\Sigma}\mathbf{M}') = \mathbf{M}\boldsymbol{\Sigma}$$
$$(\boldsymbol{\Sigma}^{-1}\mathbf{M})(\mathbf{M}\boldsymbol{\Sigma}\mathbf{M}')(\boldsymbol{\Sigma}^{-1}\mathbf{M}) = (\boldsymbol{\Sigma}^{-1}\mathbf{M}).$$

□

Proposition A.1.2. *For duplication matrix $\mathbf{D}$, let $\mathbf{H}$ and $\boldsymbol{\Lambda}$ be matrices such that*

$$\mathbf{X} \sim \mathcal{N}(-\frac{1}{2}\mathbf{M_W}\mathbf{H}\mathbf{D}'\mathbf{D}\text{vech}(\boldsymbol{\Lambda}), \mathbf{M_W}\boldsymbol{\Sigma}\mathbf{M_W}')$$

Then, $\mathbf{D}'\mathbf{D}(\mathbf{M_W}\mathbf{H})'(\mathbf{M_W}\boldsymbol{\Sigma}\mathbf{M_W}')^{\dagger}\mathbf{X}$ is identical in distribution to $\mathbf{D}'\mathbf{D}\mathbf{H}'\boldsymbol{\Sigma}^{-1}\mathbf{X} \sim \mathcal{N}(-\frac{1}{2}\boldsymbol{\Omega}\text{vec}(\boldsymbol{\Lambda}), \boldsymbol{\Omega})$ where $\boldsymbol{\Omega} = (\mathbf{D}'\mathbf{D})\mathbf{H}'\boldsymbol{\Sigma}^{-1}\mathbf{M}\mathbf{H}(\mathbf{D}'\mathbf{D})$.

Proof. Let $\boldsymbol{\Omega}_{\mathbf{W}} = (\mathbf{D}'\mathbf{D})(\mathbf{M_W}\mathbf{H})'(\mathbf{M_W}\boldsymbol{\Sigma}\mathbf{M_W}')^{\dagger}(\mathbf{M_W}\mathbf{H})(\mathbf{D}'\mathbf{D})$. Then $(\mathbf{D}'\mathbf{D})(\mathbf{M_W}\mathbf{H})'(\mathbf{M_W}\boldsymbol{\Sigma}\mathbf{M_W}')^{\dagger}\mathbf{X} \sim \mathcal{N}(-\frac{1}{2}\boldsymbol{\Omega}_{\mathbf{W}}\text{vec}(\boldsymbol{\Lambda}), \boldsymbol{\Omega}_{\mathbf{W}})$.

Proposition A.1.1 implies $\mathbf{M_W}'(\mathbf{M_W}\boldsymbol{\Sigma}\mathbf{M_W}')^{\dagger}\mathbf{M_W}$ is invariant to the choice of $\mathbf{W}$, so $\boldsymbol{\Omega}_{\mathbf{W}}$ is also invariant to $\mathbf{W}$.

Without loss of generality, set $\mathbf{W} = \boldsymbol{\Sigma}^{-1}$. Lemma A.1.1 shows that $\boldsymbol{\Sigma}^{-1}\mathbf{M}$ is a reflexive g-inverse of $\mathbf{M_W}\boldsymbol{\Sigma}\mathbf{M_W}'$. Then

$$(\mathbf{D}'\mathbf{D})(\mathbf{M}\mathbf{H})'\boldsymbol{\Sigma}^{-1}\mathbf{X} \sim \mathcal{N}(-\frac{1}{2}(\mathbf{D}'\mathbf{D})(\mathbf{M}\mathbf{H})'\boldsymbol{\Sigma}^{-1}\mathbf{M}\mathbf{H}(\mathbf{D}'\mathbf{D})\text{vech}(\boldsymbol{\Lambda}), (\mathbf{D}'\mathbf{D})(\mathbf{M}\mathbf{H})'\boldsymbol{\Sigma}^{-1}\mathbf{M}\mathbf{H}(\mathbf{D}'\mathbf{D})). \tag{A.1.1}$$

Also,

$$(\mathbf{D}'\mathbf{D})\mathbf{H}'\boldsymbol{\Sigma}^{-1}\mathbf{X} \sim \mathcal{N}(-\frac{1}{2}(\mathbf{D}'\mathbf{D})\mathbf{H}'\boldsymbol{\Sigma}^{-1}\mathbf{M}\mathbf{H}(\mathbf{D}'\mathbf{D})\text{vech}(\boldsymbol{\Lambda}), (\mathbf{D}'\mathbf{D})\mathbf{H}'\boldsymbol{\Sigma}^{-1}\mathbf{M}\mathbf{H}(\mathbf{D}'\mathbf{D})). \tag{A.1.2}$$

The distributions of Equations A.1.1 and A.1.2 are identical since $\boldsymbol{\Sigma}^{-1}\mathbf{M} = \mathbf{M}'\boldsymbol{\Sigma}^{-1}\mathbf{M}$. □

Lemma A.1.2. *Let* $\mathbf{v} \in \mathbb{R}^{n_v}$ *and* $\overline{\mathbf{v}} = \mathbf{E}\mathbf{v}$ *for selection matrix* $\mathbf{E} = (\mathbf{e}_{i_1}, \ldots, \mathbf{e}_{n_{\overline{v}}})'$ *for unit vector* $\mathbf{e}_i$ *and* $\{i_j\}_{j=1}^{n_{\overline{v}}}$ *gives the indices for the non-zero elements of* $\mathbf{v}$*. Then* $\mathbf{E}'\overline{\mathbf{v}} = \mathbf{v}$*.*

Proof. $\mathbf{E}'\overline{\mathbf{v}} = \mathbf{E}'\mathbf{E}\mathbf{v} = [\sum_{j=1}^{n_{\overline{v}}} \mathbf{e}_{i_j}\mathbf{e}_{i_j}']\mathbf{v} = \mathbf{v}$. □

A.1.2 Framework and results for local parameter heterogeneity

Assumption A.1.1 (GMM conditions)**.**

(i) $\mathbf{W}\mathbb{E}[\mathbf{g}(\mathbf{x}, \mathbf{y}^0, \boldsymbol{\theta})] = 0$ *only if* $\boldsymbol{\theta} = \boldsymbol{\theta}^*$*.*

(ii) $\boldsymbol{\theta}_i \in \text{interior}(\boldsymbol{\Theta})$ *for* $\boldsymbol{\Theta}$ *compact.*

(iii) $\mathbb{E}[\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\mathbf{g}(\mathbf{x}, \mathbf{y}^0, \boldsymbol{\theta})\|] < \infty$.

(iv) $\mathbb{E}[\mathbf{g}(\mathbf{x}, \mathbf{y}^0; \boldsymbol{\theta}^*)] = 0$ *and* $\mathbb{E}[\|\mathbf{g}(\mathbf{x}, \mathbf{y}^0; \boldsymbol{\theta}^*)\|^2] < \infty$ *finite.*

(v) $\mathbb{E}[\sup_{\boldsymbol{\theta}\in\mathcal{N}} \|\frac{\partial}{\partial\boldsymbol{\theta}'}\mathbf{g}(\mathbf{x}, \mathbf{y}^0, \boldsymbol{\theta})\|] < \infty$.

(vi) $\mathbf{G}'\mathbf{W}\mathbf{G}$ *non-singular.*

(vii) $\mathbb{E}[\sup_{\boldsymbol{\theta}\in\mathcal{N}} \|\mathbf{g}(\mathbf{x}, \mathbf{y}^0; \boldsymbol{\theta})\|^2] < \infty$.

(viii) *For* $r \in \{1, \ldots, m\}$*,* $\mathbb{E}[\sup_{\boldsymbol{\theta}\in\mathcal{N}} \|\frac{\partial}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} g_r(\mathbf{x}, \mathbf{y}^0; \boldsymbol{\theta})\|] < \infty$.

Assumption A.1.2 (Regularity conditions)**.** *Let* $\mathcal{N}$ *be a neighborhood of* $\boldsymbol{\theta}^*$*.*

i) *For* $r \in \{1, \ldots, m\}$*,* $\mathbb{E}[\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\frac{\partial g_r(\mathbf{x}, \mathbf{f}(\mathbf{x}, \boldsymbol{\varepsilon}; \boldsymbol{\theta}); \boldsymbol{\theta})}{\partial\boldsymbol{\theta}'}\|]$ *and* $\mathbb{E}[\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\frac{\partial^2 g_r(\mathbf{x}, \mathbf{f}(\mathbf{x}, \boldsymbol{\varepsilon}; \boldsymbol{\theta}); \boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\|]$ *are bounded.*

ii) $\mathbb{E}[\|\frac{\partial^2}{\partial\boldsymbol{\tau}\partial\boldsymbol{\tau}'}g_r(\mathbf{x}_i, \mathbf{f}(\mathbf{x}, \boldsymbol{\varepsilon}; \boldsymbol{\tau}); \boldsymbol{\theta}^*)\|^2] \mid_{\boldsymbol{\tau}=\boldsymbol{\theta}^*}$ *is bounded.*

iii) $\mathbb{E}[\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\frac{\partial}{\partial\boldsymbol{\tau}'}\mathbf{g}(\mathbf{x}_i, \mathbf{f}(\mathbf{x}, \boldsymbol{\varepsilon}; \boldsymbol{\tau}); \boldsymbol{\theta})\|^2] \mid_{\boldsymbol{\tau}=\boldsymbol{\theta}^*}$ *and* $\mathbb{E}[\sup_{\boldsymbol{\theta}\in\mathcal{N}} \|\frac{\partial}{\partial\boldsymbol{\tau}'}\mathbf{g}(\mathbf{x}_i, \mathbf{f}(\mathbf{x}, \boldsymbol{\varepsilon}; \boldsymbol{\tau}); \boldsymbol{\theta})\|^4] \mid_{\boldsymbol{\tau}=\boldsymbol{\theta}^*}$ *are bounded.*

iv) $\mathbb{E}[\sup_{\boldsymbol{\theta}\in\mathcal{N}} \|\frac{\partial}{\partial\boldsymbol{\tau}'}\left(\frac{\partial g_r(\mathbf{x}, \mathbf{f}(\mathbf{x}, \boldsymbol{\varepsilon}; \boldsymbol{\tau}); \boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right)\|^2] \mid_{\boldsymbol{\tau}=\boldsymbol{\theta}^*}$ *is bounded.*

v) $\mathbb{E}[\sup_{\boldsymbol{\theta}\in\mathcal{N}} \|\frac{\partial}{\partial\boldsymbol{\tau}'}\left(\frac{\partial g_r(\mathbf{x}, \mathbf{f}(\mathbf{x}, \boldsymbol{\varepsilon}; \boldsymbol{\tau}); \boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right)g_{r'}(\mathbf{x}, \mathbf{y}^0; \boldsymbol{\theta})\|^2] \mid_{\boldsymbol{\tau}=\boldsymbol{\theta}^*}$ *is bounded for* $r, r' \in \{1, \ldots, m\}$.

vi) $\mathbb{E}[\sup_{\boldsymbol{\theta}\in\mathcal{N}} \|\frac{\partial^3}{\partial\theta_j\partial\theta_{j'}\partial\boldsymbol{\tau}'}g_r(\mathbf{x}, \mathbf{f}(\mathbf{x}, \boldsymbol{\varepsilon}; \boldsymbol{\tau}); \boldsymbol{\theta})\|^2] \mid_{\boldsymbol{\tau}=\boldsymbol{\theta}^*}$ *is bounded.*

Proof of Lemma 1.3.1

Proof. Throughout, Assumption 1.3.3 allows for differentiability of the moment function and generating model. Let the r'th moment condition be $g_r(\mathbf{x}, \mathbf{y}; \boldsymbol{\theta})$. Then, consider the moment's Taylor expansion about $\boldsymbol{\theta}^*$:

$$\begin{aligned}\frac{\sqrt{n}}{n}\sum_{i=1}^{n} g_r(\mathbf{x}_i, \mathbf{y}_i; \boldsymbol{\theta}^*) &= \frac{\sqrt{n}}{n}\sum_{i=1}^{n} g_r(\mathbf{x}_i, \mathbf{f}(\mathbf{x}_i, \boldsymbol{\varepsilon}_i; \boldsymbol{\theta}^*); \boldsymbol{\theta}^*)\\ &+ \frac{\partial}{\partial\boldsymbol{\theta'}}\Big[g_r(\mathbf{x}_i, \mathbf{f}(\mathbf{x}_i, \boldsymbol{\varepsilon}_i; \boldsymbol{\theta}^*); \boldsymbol{\theta}^*)\Big]\mathbf{U}_i\mathbf{s}n^{-1/4}\\ &+ \frac{1}{2}\mathbf{s}'\mathbf{U}_i'\frac{\partial^2}{\partial\boldsymbol{\tau}\partial\boldsymbol{\tau'}}\Big[g_r(\mathbf{x}_i, \mathbf{f}(\mathbf{x}_i, \boldsymbol{\varepsilon}_i; \boldsymbol{\tau}); \boldsymbol{\theta}^*)\Big]\mathbf{U}_i\mathbf{s}n^{-1/2}|_{\boldsymbol{\tau}=\widetilde{\boldsymbol{\theta}}}.\end{aligned}$$

Using Assumption 1.3.3, the second order term follows from a second order mean value theorem expansion where $\widetilde{\boldsymbol{\theta}}$ is on the line segment connecting $\boldsymbol{\theta}^*$ and $\boldsymbol{\theta}^* + \mathbf{U}_i\mathbf{s}n^{-1/4}$.

Note that

$$\begin{aligned}&\frac{1}{2n}\sum_{i=1}^{n}\Big(\mathbf{s}'\mathbf{U}_i'\frac{\partial^2}{\partial\boldsymbol{\tau}\partial\boldsymbol{\tau'}}\Big[g_r(\mathbf{x}_i, \mathbf{f}(\mathbf{x}_i, \varepsilon_i; \boldsymbol{\tau}); \boldsymbol{\theta}^*)\Big]\mathbf{U}_i\mathbf{s} - \mathbb{E}[\mathbf{s}'\mathbf{U}_i'\frac{\partial^2}{\partial\boldsymbol{\tau}_0\partial\boldsymbol{\tau}_0'}\Big[g_r(\mathbf{x}_i, \mathbf{f}(\mathbf{x}_i, \varepsilon_i; \boldsymbol{\tau}_0); \boldsymbol{\theta}^*)\Big]\mathbf{U}_i\mathbf{s}]\Big)|_{\boldsymbol{\tau}=\widetilde{\boldsymbol{\theta}}, \boldsymbol{\tau}_0=\boldsymbol{\theta}^*}\\ &= \frac{1}{2n}\sum_{i=1}^{n}\Big(\mathbf{s}'\mathbf{U}_i'\frac{\partial^2}{\partial\boldsymbol{\tau}\partial\boldsymbol{\tau'}}\Big[g_r(\mathbf{x}_i, \mathbf{f}(\mathbf{x}_i, \varepsilon_i; \boldsymbol{\tau}); \boldsymbol{\theta}^*)\Big]\mathbf{U}_i\mathbf{s} - \mathbb{E}[\mathbf{s}'\mathbf{U}_i'\frac{\partial^2}{\partial\boldsymbol{\tau}\partial\boldsymbol{\tau'}}\Big[g_r(\mathbf{x}_i, \mathbf{f}(\mathbf{x}_i, \varepsilon_i; \boldsymbol{\tau}); \boldsymbol{\theta}^*)\Big]\mathbf{U}_i\mathbf{s}]\Big)\\ &+ \Big(\mathbb{E}[\mathbf{s}'\mathbf{U}_i'\frac{\partial^2}{\partial\boldsymbol{\tau}\partial\boldsymbol{\tau'}}\Big[g_r(\mathbf{x}_i, \mathbf{f}(\mathbf{x}_i, \varepsilon_i; \boldsymbol{\tau}); \boldsymbol{\theta}^*)\Big]\mathbf{U}_i\mathbf{s}] - \mathbb{E}[\mathbf{s}'\mathbf{U}_i'\frac{\partial^2}{\partial\boldsymbol{\tau}_0\partial\boldsymbol{\tau}_0'}\Big[g_r(\mathbf{x}_i, \mathbf{f}(\mathbf{x}_i, \varepsilon_i; \boldsymbol{\tau}_0); \boldsymbol{\theta}^*)\Big]\mathbf{U}_i\mathbf{s}]\Big)|_{\boldsymbol{\tau}=\widetilde{\boldsymbol{\theta}}, \boldsymbol{\tau}_0=\boldsymbol{\theta}^*}\\ &\xrightarrow{p} 0.\end{aligned}$$

The first parenthetical term of the above display is $o_p(1)$ from Chebyshev's inequality, Condition (ii) of Assumption A.1.1, and Condition ii) of Assumption A.1.2. Then, the above display can be written as

$$\frac{\sqrt{n}}{n}\sum_{i=1}^{n} g_r(\mathbf{x}_i, \mathbf{y}_i; \boldsymbol{\theta}^*) = \frac{\sqrt{n}}{n}\sum_{i=1}^{n} g_r(\mathbf{x}_i, \mathbf{f}(\mathbf{x}_i, \boldsymbol{\varepsilon}_i; \boldsymbol{\theta}^*); \boldsymbol{\theta}^*) + \frac{1}{2}\mathbb{E}\left[\mathbf{s}'\mathbf{U}_i'\frac{\partial^2}{\partial\boldsymbol{\tau}\partial\boldsymbol{\tau}'}\Big[g_r(\mathbf{x}_i, \mathbf{f}(\mathbf{x}_i, \boldsymbol{\varepsilon}_i; \boldsymbol{\tau}); \boldsymbol{\theta}^*)\Big]\mathbf{U}_i\mathbf{s}|_{\boldsymbol{\tau}=\boldsymbol{\theta}^*}\right] + o_p(1).$$

The second term of the above display simplifies to

$$\begin{aligned}\frac{1}{2}\mathbb{E}\left[\mathbf{s}'\mathbf{U}_i'\frac{\partial^2}{\partial\boldsymbol{\tau}\partial\boldsymbol{\tau}'}\Big[g_r(\mathbf{x}_i, \mathbf{f}(\mathbf{x}_i, \boldsymbol{\varepsilon}_i; \boldsymbol{\tau}); \boldsymbol{\theta}^*)\Big]\mathbf{U}_i\mathbf{s}|_{\boldsymbol{\tau}=\boldsymbol{\theta}^*}\right] &= -\frac{1}{2}\text{trace}\left(\mathbb{E}\Big[\frac{\partial^2 g_r(\mathbf{x}_i, \mathbf{f}(\mathbf{x}_i, \boldsymbol{\varepsilon}_i; \boldsymbol{\theta}^*))}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\Big]\boldsymbol{\Lambda}\right)\\ &= -\frac{1}{2}\text{vec}\left(\mathbb{E}\Big[\frac{\partial^2 g_r(\mathbf{x}_i, \mathbf{f}(\mathbf{x}_i, \boldsymbol{\varepsilon}_i; \boldsymbol{\theta}^*))}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\Big]\right)'\text{vec}(\boldsymbol{\Lambda})\end{aligned}$$

The first equality follows from the cyclicality of the trace operator, independence of $\mathbf{u}_i$ from $(\mathbf{x}_i', \boldsymbol{\varepsilon}_i')'$, and applying Lemma A.1.3. The desired result follows from stacking moment function $g_r(\mathbf{Z}_i, \boldsymbol{\theta}^*)$ and applying Lemma A.1.2.

□

Lemma A.1.3. *Impose Parts i) and ii) of Assumption 1.3.3, Assumption A.1.1, and Assumption A.1.2. Then,*

i) $\mathbb{E}\Big[\frac{\partial g_r(\mathbf{x}_i, \mathbf{y}_i^0; \boldsymbol{\theta}^*)}{\partial\mathbf{y}'}\frac{\partial\mathbf{f}(\mathbf{x}_i, \boldsymbol{\varepsilon}_i; \boldsymbol{\theta}^*)}{\partial\boldsymbol{\theta}}\Big] = -\mathbb{E}\Big[\frac{\partial g_r(\mathbf{x}_i, \mathbf{y}_i^0; \boldsymbol{\theta}^*)}{\partial\boldsymbol{\theta}}\Big]$.

ii) $\mathbb{E}\Big[\frac{\partial^2}{\partial\boldsymbol{\tau}\partial\boldsymbol{\tau}'}g_r(\mathbf{x}_i, \mathbf{f}(\mathbf{x}_i, \boldsymbol{\varepsilon}_i; \boldsymbol{\tau}); \boldsymbol{\theta}^*)\Big]\ |_{\boldsymbol{\tau}=\boldsymbol{\theta}^*} = -\Big\{\mathbb{E}\Big[\frac{\partial^2 g_r(\mathbf{x}_i, \mathbf{y}_i^0; \boldsymbol{\theta}^*)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\Big] + \mathbb{E}\Big[\Big(\frac{\partial\mathbf{f}(\mathbf{x}_i, \boldsymbol{\varepsilon}_i; \boldsymbol{\theta}^*)}{\partial\boldsymbol{\theta}'}\Big)'\frac{\partial^2 g_r(\mathbf{x}_i, \mathbf{y}_i^0; \boldsymbol{\theta}^*)}{\partial\mathbf{y}\partial\boldsymbol{\theta}'}\Big] + \mathbb{E}\Big[\Big(\frac{\partial\mathbf{f}(\mathbf{x}_i, \boldsymbol{\varepsilon}_i; \boldsymbol{\theta}^*)}{\partial\boldsymbol{\theta}'}\Big)'\frac{\partial^2 g_r(\mathbf{x}_i, \mathbf{y}_i^0; \boldsymbol{\theta}^*)}{\partial\mathbf{y}\partial\boldsymbol{\theta}'}\Big]'\Big\}$.

Proof. For the result in the first display, consider the moment condition $\mathbb{E}[g_r(\mathbf{x}_i, \mathbf{f}(\mathbf{x}_i, \boldsymbol{\varepsilon}_i, \boldsymbol{\theta}^*); \boldsymbol{\theta}^*)] = 0$. Computing the total derivative with respect to $\boldsymbol{\theta}$ and evaluating at $\boldsymbol{\theta} = \boldsymbol{\theta}^*$, Condition i) of Assumption A.1.2, Assumption 1.3.3, and the multivariate chain rule imply

$$\mathbf{0}_{1\times p} = \mathbb{E}\Big[\frac{\partial g_r(\mathbf{x}_i, \mathbf{y}_i^0; \boldsymbol{\theta}^*)}{\partial\mathbf{y}'}\frac{\partial\mathbf{f}(\mathbf{x}_i, \boldsymbol{\varepsilon}_i; \boldsymbol{\theta}^*)}{\partial\boldsymbol{\theta}'} + \frac{\partial g_r(\mathbf{x}_i, \mathbf{y}_i^0; \boldsymbol{\theta}^*)}{\partial\boldsymbol{\theta}'}\Big].$$

Analogously, the expected second order total derivative with respect to $\boldsymbol{\theta}$ is where

$$\mathbf{0}_{p\times p} = \mathbb{E}\Big[\Big(\frac{\partial \mathbf{f}(\mathbf{x}_i,\boldsymbol{\varepsilon}_i;\boldsymbol{\theta}^*)}{\partial\boldsymbol{\theta}'}\Big)'\frac{\partial^2 g_r(\mathbf{x}_i,\mathbf{y}_i^0;\boldsymbol{\theta}^*)}{\partial\mathbf{y}\partial\mathbf{y}'}\frac{\partial \mathbf{f}(\mathbf{x}_i,\boldsymbol{\varepsilon}_i;\boldsymbol{\theta}^*)}{\partial\boldsymbol{\theta}'} + \frac{\partial^2 g_r(\mathbf{x}_i,\mathbf{y}_i^0;\boldsymbol{\theta}^*)}{\partial\boldsymbol{\theta}\partial\mathbf{y}'}\frac{\partial \mathbf{f}(\mathbf{x}_i,\boldsymbol{\varepsilon}_i;\boldsymbol{\theta}^*)}{\partial\boldsymbol{\theta}'}$$
$$+\sum_{t=1}^{T}\frac{\partial g_r(\mathbf{x}_i,\mathbf{y}_i^0;\boldsymbol{\theta}^*)}{\partial y_t}\frac{\partial^2 \mathbf{f}_t(\mathbf{x}_i,\boldsymbol{\varepsilon}_i;\boldsymbol{\theta}^*)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} + \frac{\partial^2 g_r(\mathbf{x}_i,\mathbf{y}_i^0;\boldsymbol{\theta}^*)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} + \Big(\frac{\partial \mathbf{f}(\mathbf{x}_i,\boldsymbol{\varepsilon}_i;\boldsymbol{\theta}^*)}{\partial\boldsymbol{\theta}'}\Big)'\frac{\partial^2 g_r(\mathbf{x}_i,\mathbf{y}_i^0;\boldsymbol{\theta}^*)}{\partial\mathbf{y}\partial\boldsymbol{\theta}'}\Big].$$

Rearranging gives the desired result. $\square$

Proposition A.1.3. *Impose Assumptions 1.3.1, 1.3.3, A.1.1, and A.1.2. Then* $\widehat{\boldsymbol{\theta}} \xrightarrow{p} \boldsymbol{\theta}^*$.

Proof. Let $\mathbf{g}_0(\boldsymbol{\theta}) = \mathbb{E}[\mathbf{g}(\mathbf{x},\mathbf{y}^0,\boldsymbol{\theta})]$ and $Q_0(\theta) = -\mathbf{g}_0(\boldsymbol{\theta})\mathbf{W}\mathbf{g}_0(\boldsymbol{\theta})$. Throughout, Assumption 1.3.3 allows for differentiability of the moment function and generating model.

From the mean-value theorem, there exists $\widetilde{\boldsymbol{\theta}}$ on the line segment between $\boldsymbol{\theta}^*$ and $\boldsymbol{\theta}^* + \mathbf{U}_i\mathbf{s}n^{-1/4}$ such that

$$\begin{aligned}
&\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\|\frac{1}{n}\sum_{i=1}^{n}\mathbf{g}(\mathbf{x},\mathbf{y};\boldsymbol{\theta}) - \mathbb{E}[\mathbf{g}(\mathbf{x},\mathbf{y}^0;\boldsymbol{\theta})]\|\\
&\quad= \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\|\frac{1}{n}\sum_{i=1}^{n}\mathbf{g}(\mathbf{x},\mathbf{y}^0;\boldsymbol{\theta}) + \frac{\partial}{\partial\boldsymbol{\tau}'}\mathbf{g}(\mathbf{x},\mathbf{f}(\mathbf{x}_i,\boldsymbol{\varepsilon}_i;\boldsymbol{\tau});\boldsymbol{\theta})\mathbf{U}_i\mathbf{s}n^{-1/4} - \mathbb{E}[\mathbf{g}(\mathbf{x},\mathbf{y}^0;\boldsymbol{\theta})]\|\,|_{\boldsymbol{\tau}=\widetilde{\boldsymbol{\theta}}}\\
&\quad\leq \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\|\frac{1}{n}\sum_{i=1}^{n}\mathbf{g}(\mathbf{x},\mathbf{y}^0;\boldsymbol{\theta}) - \mathbb{E}[\mathbf{g}(\mathbf{x},\mathbf{y}^0;\boldsymbol{\theta})]\|\\
&\quad+ \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\|\frac{1}{n}\sum_{i=1}^{n}\frac{\partial}{\partial\boldsymbol{\tau}'}\mathbf{g}(\mathbf{x},\mathbf{f}(\mathbf{x}_i,\boldsymbol{\varepsilon}_i;\boldsymbol{\tau});\boldsymbol{\theta})\mathbf{U}_i\mathbf{s}n^{-1/4} - \mathbb{E}[\frac{\partial}{\partial\boldsymbol{\tau}'}\mathbf{g}(\mathbf{x},\mathbf{f}(\mathbf{x}_i,\boldsymbol{\varepsilon}_i;\boldsymbol{\tau});\boldsymbol{\theta})\mathbf{U}_i\mathbf{s}n^{-1/4}]\|\\
&\quad+ \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\|\mathbb{E}[\frac{\partial}{\partial\boldsymbol{\tau}'}\mathbf{g}(\mathbf{x},\mathbf{f}(\mathbf{x}_i,\boldsymbol{\varepsilon}_i;\boldsymbol{\tau});\boldsymbol{\theta})\mathbf{U}_i\mathbf{s}n^{-1/4}]\|\,|_{\boldsymbol{\tau}=\widetilde{\boldsymbol{\theta}}}
\end{aligned}$$

The first term of the above display is $o_p(1)$ from Condition of (iii) of Assumption A.1.1. The second and third terms of the above display are $o_p(1)$ from Chebyshev's inequality, Condition (ii) of Assumption A.1.1, and Condition iii) of Assumption A.1.2. Hence the above display is $o_p(1)$.

Since $\boldsymbol{\Theta}$ is compact, the desired result follows from applying the arguments of Theorem 2.6 of Newey and McFadden (1994) with $\mathbf{g}_0(\boldsymbol{\theta})$ defined in the beginning of this proof.

$\square$

Proof of Proposition 1.3.1

Proof. Let $\widehat{\mathbf{G}}(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{g}(\mathbf{x}_i, \mathbf{y}_i; \boldsymbol{\theta})$. Throughout, Assumption 1.3.3 allows for differentiability of the moment function and generating model.

From the mean value theorem and the first order condition (applying Conditions (vi), (ii)of Assumption A.1.1 and Assumption 1.3.3), there exists $\overline{\boldsymbol{\theta}}$ between $\boldsymbol{\theta}^*$ and $\widehat{\boldsymbol{\theta}}$ such that

$$\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*) = -[\widehat{\mathbf{G}}(\widehat{\boldsymbol{\theta}})'\widehat{\mathbf{W}}\widehat{\mathbf{G}}(\overline{\boldsymbol{\theta}})]^{-1}\widehat{\mathbf{G}}(\widehat{\boldsymbol{\theta}})\frac{\sqrt{n}}{n}\sum_{i=1}^{n}\mathbf{g}(\mathbf{x}_i, \mathbf{y}_i, \boldsymbol{\theta}^*).$$

Consider the r'th moment function. From a mean value theorem expansion, there exists $\widetilde{\boldsymbol{\theta}}$ on the line segment between $\boldsymbol{\theta}^*$ and $\boldsymbol{\theta}^* + \mathbf{U}_i\mathbf{s}n^{-1/4}$ such that

$$\begin{aligned}
&\sup_{\boldsymbol{\theta}\in\mathcal{N}} \Big\|\frac{1}{n}\sum_{i=1}^{n}\frac{\partial g_r(\mathbf{x}_i, \mathbf{y}_i; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} - \mathbb{E}[\frac{\partial g_r(\mathbf{x}, \mathbf{y}^0; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}]\Big\| = \\
&\quad \sup_{\boldsymbol{\theta}\in\mathcal{N}} \Big\|\frac{1}{n}\sum_{i=1}^{n}\frac{\partial g_r(\mathbf{x}_i, \mathbf{y}_i^0; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} + \frac{\partial}{\partial \boldsymbol{\tau}'}\Big(\frac{\partial g_r(\mathbf{x}_i, \mathbf{f}(\mathbf{x}_i, \boldsymbol{\varepsilon}_i; \boldsymbol{\tau}); \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\Big)\mathbf{U}_i\mathbf{s}n^{-1/4} - \mathbb{E}[\frac{\partial g_r(\mathbf{x}, \mathbf{y}^0; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}]\Big\| \,\Big|_{\boldsymbol{\tau}=\widetilde{\boldsymbol{\theta}}} \\
\leq &\quad \sup_{\boldsymbol{\theta}\in\mathcal{N}} \Big\|\frac{1}{n}\sum_{i=1}^{n}\frac{\partial g_r(\mathbf{x}_i, \mathbf{y}_i^0; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} - \mathbb{E}[\frac{\partial g_r(\mathbf{x}, \mathbf{y}^0; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}]\Big\| \\
&\quad + \sup_{\boldsymbol{\theta}\in\mathcal{N}} \Big\|\frac{\partial}{\partial \boldsymbol{\tau}'}\Big(\frac{\partial g_r(\mathbf{x}_i, \mathbf{f}(\mathbf{x}_i, \boldsymbol{\varepsilon}_i; \boldsymbol{\tau}); \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\Big)\mathbf{U}_i\mathbf{s}n^{-1/4} - \mathbb{E}[\frac{\partial}{\partial \boldsymbol{\tau}'}\Big(\frac{\partial g_r(\mathbf{x}_i, \mathbf{f}(\mathbf{x}_i, \boldsymbol{\varepsilon}_i; \boldsymbol{\tau}); \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\Big)\mathbf{U}_i\mathbf{s}n^{-1/4}]\Big\| \\
&\quad + \sup_{\boldsymbol{\theta}\in\mathcal{N}} \Big\|\mathbb{E}[\frac{\partial}{\partial \boldsymbol{\tau}'}\Big(\frac{\partial g_r(\mathbf{x}_i, \mathbf{f}(\mathbf{x}_i, \boldsymbol{\varepsilon}_i; \boldsymbol{\tau}); \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\Big)\mathbf{U}_i\mathbf{s}n^{-1/4}]\Big\| \,\Big|_{\boldsymbol{\tau}=\widetilde{\boldsymbol{\theta}}}\,.
\end{aligned}$$

The first term after the inequality is $o_p(1)$ from Condition (v) of Assumption A.1.1. The second and third terms after the inequality are $o_p(1)$ from Condition iv) of Assumption A.1.2, Chebyshev's inequality, and Condition (ii) of Assumption A.1.1. Then $\widehat{\mathbf{G}}(\widehat{\boldsymbol{\theta}}) \xrightarrow{p} \mathbf{G}$ and $\widehat{\mathbf{G}}(\overline{\boldsymbol{\theta}}) \xrightarrow{p} \mathbf{G}$.

The desired result follows from combining with Lemma 1.3.1 and applying Lemma A.1.2 in the final line,

$$\begin{aligned}\sqrt{n}(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}^*) &= -(\mathbf{G}'\mathbf{W}\mathbf{G})^{-1}\mathbf{G}'\mathbf{W}\frac{\sqrt{n}}{n}\sum_{i=1}^{n}\mathbf{g}(\mathbf{x}_i,\mathbf{y}_i,\boldsymbol{\theta}^*)+o_p(1)\\ &= -(\mathbf{G}'\mathbf{W}\mathbf{G})^{-1}\mathbf{G}'\mathbf{W}\frac{\sqrt{n}}{n}\Big[\sum_{i=1}^{n}\mathbf{g}(\mathbf{x}_i,\mathbf{f}(\mathbf{x}_i,\boldsymbol{\varepsilon}_i;\boldsymbol{\theta}^*))\Big]\\ &\quad+\frac{1}{2}(\mathbf{G}'\mathbf{W}\mathbf{G})^{-1}\mathbf{G}'\mathbf{W}\mathbf{H}(\mathbf{D}'\mathbf{D})\mathbf{E}'\boldsymbol{\lambda}+o_p(1).\end{aligned}$$

□

Proof of Proposition 1.3.2

Proof. Throughout, Assumption 1.3.3 allows for differentiability of the moment function and generating model. A mean value theorem expansion about $\widehat{\boldsymbol{\theta}}$ implies there exists $\widetilde{\boldsymbol{\theta}}$ on the line segment between $\boldsymbol{\theta}^*$ and $\widehat{\boldsymbol{\theta}}$ such that

$$\begin{aligned}\frac{\sqrt{n}}{n}\sum_{i=1}^{n}\mathbf{g}(\mathbf{x}_i,\mathbf{y}_i;\widehat{\boldsymbol{\theta}}) &= \frac{\sqrt{n}}{n}\sum_{i=1}^{n}\mathbf{g}(\mathbf{x}_i,\mathbf{y}_i,\boldsymbol{\theta}^*)+\frac{\partial}{\partial\boldsymbol{\tau}'}\mathbf{g}(\mathbf{x}_i,\mathbf{y}_i;\boldsymbol{\tau})(\widehat{\boldsymbol{\theta}}_0-\boldsymbol{\theta}^*)\mid_{\boldsymbol{\tau}=\widetilde{\boldsymbol{\theta}}}\\ &= \Big(\frac{\sqrt{n}}{n}\sum_{i=1}^{n}\mathbf{M}\mathbf{g}(\mathbf{x}_i,\mathbf{f}(\mathbf{x}_i,\boldsymbol{\varepsilon}_i;\boldsymbol{\theta}^*),\boldsymbol{\theta}^*)\Big)-\frac{1}{2}\mathbf{M}\mathbf{H}(\mathbf{D}'\mathbf{D})\mathbf{E}'\boldsymbol{\lambda}.\end{aligned}$$

The second line follows from Proposition 1.3.1. $\sum_{i=1}^{n}\frac{\partial}{\partial\boldsymbol{\tau}'}\mathbf{g}(\mathbf{x}_i,\mathbf{y}_i;\boldsymbol{\tau})\xrightarrow{p}\mathbf{G}$ from Condition iv) of Assumption A.1.2, Chebyshev's inequality, and Condition (ii) of Assumption A.1.1. □

Lemma A.1.4. *Impose Assumptions 1.3.1, 1.3.3, A.1.1, and A.1.2. Then,*

i) $\widehat{\mathbf{H}}\xrightarrow{p}\mathbf{H}$.

ii) $\widehat{\boldsymbol{\Sigma}}\xrightarrow{p}\boldsymbol{\Sigma}$.

Proof. Proposition A.1.3 implies $\widehat{\boldsymbol{\theta}}\xrightarrow{p}\boldsymbol{\theta}$, so it suffices to show uniform convergence over a neighborhood $\mathcal{N}$ of $\boldsymbol{\theta}^*$. Throughout, Assumption 1.3.3 allows for differentiability of the moment function and generating model.

For i), let $j, j' \in \{1, \ldots, p\}$. A mean value theorem decomposition implies

$$\begin{aligned}
&\sup_{\boldsymbol{\theta}\in\mathcal{N}} \|\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2}{\partial\theta_j\partial\theta_{j'}}g_r(\mathbf{x}_i,\mathbf{y}_i;\boldsymbol{\theta}) - \mathbb{E}[\frac{\partial^2}{\partial\theta_j\partial\theta_{j'}}g_r(\mathbf{x},\mathbf{y}^0;\boldsymbol{\theta})]\| \\
&\leq \sup_{\boldsymbol{\theta}\in\mathcal{N}} \|\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2}{\partial\theta_j\partial\theta_{j'}}g_r(\mathbf{x}_i,\mathbf{y}_i^0;\boldsymbol{\theta}) - \mathbb{E}[\frac{\partial^2}{\partial\theta_j\partial\theta_{j'}}g_r(\mathbf{x},\mathbf{y}^0;\boldsymbol{\theta})]\| \\
&+ \sup_{\boldsymbol{\theta}\in\mathcal{N}} \|\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^3}{\partial\theta_j\partial\theta_{j'}\partial\boldsymbol{\tau}'}g_r(\mathbf{x}_i,\mathbf{f}(\mathbf{x}_i,\boldsymbol{\varepsilon}_i,\boldsymbol{\tau});\boldsymbol{\theta})\mathbf{U}_i\mathbf{s}n^{-1/4}\| \,|_{\boldsymbol{\tau}=\widetilde{\boldsymbol{\theta}}}
\end{aligned}$$

The first line after the inequality is $o_p(1)$ from Condition (viii) of Assumption A.1.1. The second term after the inequality is $o_p(1)$ from Condition vi) of Assumption A.1.2, Chebyshev's inequality, and Condition (ii) of Assumption A.1.1.

For ii), a mean value theorem decomposition implies

$$\begin{aligned}
&\sup_{\boldsymbol{\theta}\in\mathcal{N}} \|\frac{1}{n}\sum_{i=1}^{n}\mathbf{g}(\mathbf{x}_i,\mathbf{y}_i;\boldsymbol{\theta})\mathbf{g}(\mathbf{x}_i,\mathbf{y}_i;\boldsymbol{\theta})' - \mathbb{E}[\mathbf{g}(\mathbf{x},\mathbf{y}^0;\boldsymbol{\theta})\mathbf{g}(\mathbf{x},\mathbf{y}^0;\boldsymbol{\theta})']\| \\
&\leq \sup_{\boldsymbol{\theta}\in\mathcal{N}} \|\frac{1}{n}\sum_{i=1}^{n}\mathbf{g}(\mathbf{x}_i,\mathbf{y}_i^0;\boldsymbol{\theta})\mathbf{g}(\mathbf{x}_i,\mathbf{y}_i^0;\boldsymbol{\theta})' - \mathbb{E}[\mathbf{g}(\mathbf{x},\mathbf{y}^0;\boldsymbol{\theta})\mathbf{g}(\mathbf{x},\mathbf{y}^0;\boldsymbol{\theta})']\| \\
&\quad + \sup_{\boldsymbol{\theta}\in\mathcal{N}} \|\frac{1}{n}\sum_{i=1}^{n}\Big(\frac{\partial}{\partial\boldsymbol{\tau}'}\mathbf{g}(\mathbf{x}_i,\mathbf{f}(\mathbf{x}_i,\boldsymbol{\varepsilon}_i;\boldsymbol{\tau});\boldsymbol{\theta})\mathbf{U}_i\mathbf{s}n^{-1/4}\Big)\Big(\frac{\partial}{\partial\boldsymbol{\tau}'}\mathbf{g}(\mathbf{x}_i,\mathbf{f}(\mathbf{x}_i,\boldsymbol{\varepsilon}_i;\boldsymbol{\tau});\boldsymbol{\theta})\mathbf{U}_i\mathbf{s}n^{-1/4}\Big)'\| \\
&\quad + 2\sup_{\boldsymbol{\theta}\in\mathcal{N}} \|\frac{1}{n}\sum_{i=1}^{n}\mathbf{g}(\mathbf{x}_i,\mathbf{y}^0;\boldsymbol{\theta})\Big(\frac{\partial}{\partial\boldsymbol{\tau}'}\mathbf{g}(\mathbf{x}_i,\mathbf{f}(\mathbf{x}_i,\boldsymbol{\varepsilon}_i;\boldsymbol{\tau});\boldsymbol{\theta})\mathbf{U}_i\mathbf{s}n^{-1/4}\Big)'\|
\end{aligned}$$

The first term after the inequality is $o_p(1)$ from Condition (vii) of Assumption A.1.1. The second term after the inequality is $o_p(1)$ from Chebyshev's inequality, Condition (ii) of Assumption A.1.1, and Condition iii) of Assumption A.1.2. Analogously, the third term after the inequality is $o_p(1)$ from Chebyshev's inequality, Condition (ii) of Assumption A.1.1, and Conditions iv) and v) of Assumption A.1.2.

$\square$

A.1.3 Semiparametric

Assumption A.1.3 (Smoothness, semiparametric)**.** *For all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, $p_{xy}(\mathbf{x}, \mathbf{y}; \boldsymbol{\theta})$ is two times continuously differentiable in $\boldsymbol{\theta}$ with probability 1.*

Assumption A.1.4 (Dominance)**.**

i) For any $\mathbf{x}, \mathbf{y}$ in the support, $\frac{\partial}{\partial t}\{p_{xy,t}(\mathbf{x}, \mathbf{y}; \boldsymbol{\theta}_t + \mathbf{U}\boldsymbol{\sigma}_t)p_{u,t}(\mathbf{u})\}$ is dominated by an integrable function.

ii) For all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, $\frac{\partial}{\partial t}\mathbf{g}(\mathbf{x}, \mathbf{y}; \boldsymbol{\theta})p_{xy,t}(\mathbf{x}, \mathbf{y}; \boldsymbol{\theta})$ is dominated by an integrable function.

iii) For all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, $\frac{\partial}{\partial \boldsymbol{\theta}'}\Big[\mathbf{g}(\mathbf{x}, \mathbf{y}; \boldsymbol{\theta})p(\mathbf{x}, \mathbf{y}; \boldsymbol{\theta})\Big]$ is dominated by an integrable function.

iv) For all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, $g_r(\mathbf{x}, \mathbf{y}; \boldsymbol{\theta}^*)\frac{\partial^2}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'}p(\mathbf{x}, \mathbf{y}; \boldsymbol{\theta})$ is dominated by an integrable function.

Definition A.1.1. *Let μ be an arbitrary measure relative to which $P_{\boldsymbol{\theta}_t,\boldsymbol{\sigma}_t,\mathrm{vech}(\mathbf{C}),p_{xy,t},p_{u,t}}$ and $P_{\boldsymbol{\theta}^*,\boldsymbol{\sigma}^*,\mathrm{vech}(\mathbf{C}),p_{xy},p_u}$ possess densities $p_{\boldsymbol{\theta}_t,\boldsymbol{\sigma}_t,\mathrm{vech}(\mathbf{C}),p_{xy,t},p_{u,t}}$ and $p_{\boldsymbol{\theta}^*,\boldsymbol{\sigma}^*,\mathrm{vech}(\mathbf{C}),p_{xy},p_u}$ respectively. Suppose the map $t \mapsto P_{\boldsymbol{\theta}_t,\boldsymbol{\sigma}_t,\mathrm{vech}(\mathbf{C}),p_{xy,t},p_{u,t}}$ from a non-negative neighborhood of $0 \in [0, \infty)$ to $\mathcal{P}$ satisfies*

$$\int \left[\frac{\sqrt{p_{\boldsymbol{\theta}_t,\boldsymbol{\sigma}_t,\mathrm{vech}(\mathbf{C}),p_{xy,t},p_{u,t}}} - \sqrt{p_{\boldsymbol{\theta}^*,\boldsymbol{\sigma}^*,\mathrm{vech}(\mathbf{C}),p_{xy},p_u}}}{t} - \frac{1}{2}(\mathbf{a}'\dot{\boldsymbol{\ell}} + \eta)\sqrt{p_0}\right]^2 d\mu \to 0,$$

for $\dot{\boldsymbol{\ell}}(\mathbf{x}, \mathbf{y}) = \frac{1}{p_{xy}(\mathbf{x},\mathbf{y};\boldsymbol{\theta}^*)}\begin{bmatrix} \frac{\partial}{\partial \boldsymbol{\theta}}p_{xy}(\mathbf{x}, \mathbf{y}; \boldsymbol{\theta}^*) \\ \frac{1}{2}\mathbf{E}(\mathbf{D}'\mathbf{D})\mathrm{vech}(\frac{\partial^2}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'}p_{xy}(\mathbf{x}, \mathbf{y}; \boldsymbol{\theta}^*)) \end{bmatrix}$, $\mathbf{a} = (\mathbf{c}', \boldsymbol{\lambda}')'$, *and measurable function $\eta(\mathbf{x}, \mathbf{y}) = \partial \log p_t(\mathbf{x}, \mathbf{y}; \boldsymbol{\theta}^*)/\partial t$. Then the parametric submodel $P_{\boldsymbol{\theta}_t,\boldsymbol{\sigma}_t,\mathrm{vech}(\mathbf{C}),p_{xy,t},p_{u,t}}$ is* ***differentiable in quadratic mean*** *at $t = 0$.*

Proof of Lemma 1.4.1

Proof. Compute the one-sided partial derivative of $\log q_t(\mathbf{x}, \mathbf{y})$ about $t = 0$ as $t \downarrow 0$.

$$
\begin{aligned}
\frac{\partial_+}{\partial t} \log q_t(\mathbf{x}, \mathbf{y}) \mid_{t=0} &= \frac{1}{q_t(\mathbf{x}, \mathbf{y})} \int \frac{\partial^+}{\partial \boldsymbol{\theta}'} p_{xy,t}(\mathbf{x}, \mathbf{y}; \boldsymbol{\theta}_t + \mathbf{U}\boldsymbol{\sigma}_t)\mathbf{c} p_{u,t}(\mathbf{u}) d\mathbf{u} \\
&+ \frac{1}{2q_t(\mathbf{x}, \mathbf{y})} \int \frac{\partial^+}{\partial \boldsymbol{\theta}'} p_{xy,t}(\mathbf{x}, \mathbf{y}; \boldsymbol{\theta}_t + \mathbf{U}\boldsymbol{\sigma}_t)\mathbf{U}\mathbf{s} t^{-1/2} p_{u,t}(\mathbf{u}) d\mathbf{u} + \eta \mid_{t=0} \\
&= \mathbf{c}' \frac{1}{p_{xy}(\mathbf{x}, \mathbf{y}; \boldsymbol{\theta}^*)} \frac{\partial}{\partial \boldsymbol{\theta}} p_{xy}(\mathbf{x}, \mathbf{y}; \boldsymbol{\theta}^*) + \frac{1}{2} \mathrm{vec}\Big(\frac{\partial^2 p_{xy}(\mathbf{x}, \mathbf{y}; \boldsymbol{\theta}^*)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}\Big)' \mathrm{vec}(\boldsymbol{\Lambda}) + \eta \\
&= \mathbf{c}' \frac{1}{p_{xy}(\mathbf{x}, \mathbf{y}; \boldsymbol{\theta}^*)} \frac{\partial}{\partial \boldsymbol{\theta}} p_{xy}(\mathbf{x}, \mathbf{y}; \boldsymbol{\theta}^*) + \frac{1}{2} \mathrm{vech}(\boldsymbol{\Lambda})'(\mathbf{D}'\mathbf{D}) \mathrm{vech}\Big(\frac{\partial^2 p_{xy}(\mathbf{x}, \mathbf{y}; \boldsymbol{\theta}^*)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}\Big) + \eta.
\end{aligned}
$$

The first equality follows from Condition i) of Assumption A.1.4. The second equality follows from applying L'Hôpital's rule to the second term. Note that the score of the density of $\mathbf{u}$ is zero since $\int \frac{\partial}{\partial t} p_{u,t}(\mathbf{u}) du = 0$.

□

Lemma A.1.5. *Impose Assumptions A.1.4 and suppose $P_{\boldsymbol{\theta}_t, \boldsymbol{\sigma}_t, \mathrm{vech}(\mathbf{C}), p_{xy,t}, p_{u,t}}$ is differentiable in quadratic mean at $t = 0$. The nuisance tangent space $\dot{\mathcal{P}}$ consists of all measurable mean-zero random functions $\eta(\mathbf{x}, \mathbf{y})$ with finite variance such that*

$$
\mathbf{0}_{m \times 1} = \iint \mathbf{g}(\mathbf{x}, \mathbf{y}; \boldsymbol{\theta}^*) \eta(\mathbf{x}, \mathbf{y}) p_{xy}(\mathbf{x}, \mathbf{y}; \boldsymbol{\theta}^*) d\mathbf{x} d\mathbf{y}.
$$

Proof. Let $\boldsymbol{\theta} = \boldsymbol{\theta}^*$. The moment condition implies that $p_{xy,t}(\mathbf{x}, \mathbf{y})$ is subject to the restriction

$$
\mathbf{0}_{m \times 1} = \iint \mathbf{g}(\mathbf{x}, \mathbf{y}; \boldsymbol{\theta}^*) p_{xy,t}(\mathbf{x}, \mathbf{y}; \boldsymbol{\theta}^*) d\mathbf{x} d\mathbf{y}.
$$

From Condition ii) of Assumption A.1.4, the partial derivative of the above display with respect to t implies that the score function for the nuisance parameter p_{xy} (noted as $\eta(\mathbf{x}, \mathbf{y})$) must satisfy

$$
\mathbf{0}_{m \times 1} = \iint \mathbf{g}(\mathbf{x}, \mathbf{y}; \boldsymbol{\theta}^*) \eta(\mathbf{x}, \mathbf{y}) p_{xy}(\mathbf{x}, \mathbf{y}; \boldsymbol{\theta}^*) d\mathbf{x} d\mathbf{y}.
$$

Therefore any member of the nuisance tangent space is necessarily a member of the conjectured tangent space.

Now instead suppose function $\eta(\mathbf{x},\mathbf{y})$ is a *bounded* element of the conjectured tangent space. Consider a parametric submodel with density

$$p_{xy,t}(\mathbf{x},\mathbf{y};\boldsymbol{\theta}^*) = (1+t\eta(\mathbf{x},\mathbf{y}))p_{xy}(\mathbf{x},\mathbf{y};\boldsymbol{\theta}^*)$$

for t sufficiently small so that $(1+t\eta(\mathbf{x},\mathbf{y}))$ is non-negative. The density is proper since $\eta(\mathbf{x},\mathbf{y})$ is mean zero. The density is compatible with the moment restrictions since

$$\iint \mathbf{g}(\mathbf{x},\mathbf{y};\boldsymbol{\theta}^*)p_{xy,t}(\mathbf{x},\mathbf{y};\boldsymbol{\theta}^*)d\mathbf{x}d\mathbf{y} = t\iint \mathbf{g}(\mathbf{x},\mathbf{y};\boldsymbol{\theta}^*)\eta(\mathbf{x},\mathbf{y})p_{xy}(\mathbf{x},\mathbf{y};\boldsymbol{\theta}^*)d\mathbf{x}d\mathbf{y} = \mathbf{0}_{m\times 1}.$$

Moreover, the score vector for this parametric submodel is $\frac{\partial}{\partial t}p_{xy,t}(\mathbf{x},\mathbf{y};\boldsymbol{\theta}^*)\mid_{t=0} = \eta(\mathbf{x},\mathbf{y})$. Thus $\eta(\mathbf{x},\mathbf{y})$ is a member of the nuisance tangent space, and the result follows since any element of the Hilbert space can be approximated by a sequence of bounded functions.

$\square$

Lemma A.1.6. *Impose Assumptions A.1.4 and A.1.1. Suppose $P_{\boldsymbol{\theta}_t,\boldsymbol{\sigma}_t,\text{vech}(\mathbf{C}),p_{xy,t},p_{u,t}}$ is differentiable in quadratic mean at $t=0$. The orthogonal complement of the nuisance tangent space is*

$$\dot{\mathcal{P}}^{\perp} = \left\{\mathbf{A}\in\mathbb{R}^{1\times m} : \mathbf{A}\mathbf{g}(\mathbf{x},\mathbf{y};\boldsymbol{\theta}^*)\right\}.$$

For any $\mathbf{h}\in\mathcal{H}$ for Hilbert space $\mathcal{H}$, the projection onto $\dot{\mathcal{P}}$ satisfies

$$\mathbf{h}-\Pi(\mathbf{h}|\dot{\mathcal{P}}) = \tag{A.1.3}$$

$$\iint \mathbf{h}\mathbf{g}(\mathbf{x},\mathbf{y};\boldsymbol{\theta}^*)'p_{xy}(\mathbf{x},\mathbf{y};\boldsymbol{\theta}^*)d\mathbf{x}d\mathbf{y}\left[\iint \mathbf{g}(\mathbf{x},\mathbf{y};\boldsymbol{\theta}^*)\mathbf{g}(\mathbf{x},\mathbf{y};\boldsymbol{\theta}^*)'p_{xy}(\mathbf{x},\mathbf{y};\boldsymbol{\theta}^*)d\mathbf{x}d\mathbf{y}\right]^{-1}\mathbf{g}(\mathbf{x},\mathbf{y};\boldsymbol{\theta}^*). \tag{A.1.4}$$

Proof. Take element $\eta(\mathbf{x},\mathbf{y}) \in \dot{\mathcal{P}}$ and element $\mathbf{A}\mathbf{g}(\mathbf{x},\mathbf{y};\boldsymbol{\theta}^*)$ for matrix $\mathbf{A}$. Then,

$$\iint \mathbf{A}\mathbf{g}(\mathbf{x},\mathbf{y};\boldsymbol{\theta}^*)\eta(\mathbf{x},\mathbf{y})p_{xy}(\mathbf{x},\mathbf{y};\boldsymbol{\theta}^*)d\mathbf{x}d\mathbf{y} = \mathbf{A}\iint \mathbf{g}(\mathbf{x},\mathbf{y};\boldsymbol{\theta}^*)\eta(\mathbf{x},\mathbf{y})p_{xy}(\mathbf{x},\mathbf{y};\boldsymbol{\theta}^*)d\mathbf{x}d\mathbf{y} = 0$$

so the proposed space is orthogonal to $\dot{\mathcal{P}}$.

Next, let $\mathbf{h}$ be a measurable function that is mean zero and with finite variance. From the Hilbert projection theorem, there exists $\mathbf{A}$ such that $\mathbf{h} - \mathbf{A}\mathbf{g}(\mathbf{x},\mathbf{y};\boldsymbol{\theta}^*) \in \dot{\mathcal{P}}$. Equivalently, $\iint[\mathbf{h} - \mathbf{A}\mathbf{g}(\mathbf{x},\mathbf{y};\boldsymbol{\theta}^*)]\mathbf{g}(\mathbf{x},\mathbf{y};\boldsymbol{\theta}^*)'p_{xy}(\mathbf{x},\mathbf{y};\boldsymbol{\theta}^*)d\mathbf{x}d\mathbf{y} = \mathbf{0}$. Condition (vii) of Assumption A.1.1 implies $\iint \mathbf{g}(\mathbf{x},\mathbf{y};\boldsymbol{\theta}^*)\mathbf{g}(\mathbf{x},\mathbf{y};\boldsymbol{\theta}^*)'p_{xy}(\mathbf{x},\mathbf{y};\boldsymbol{\theta}^*)d\mathbf{x}d\mathbf{y}$ is invertible. Rearranging gives Equation A.1.4.

□

Lemma A.1.7. *Impose Assumptions 1.3.3, 1.3.4, A.1.4, A.1.1, and A.1.2. Suppose $P_{\boldsymbol{\theta}_t,\boldsymbol{\sigma}_t,\mathrm{vech}(\mathbf{C}),p_{xy,t},p_{u,t}}$ is differentiable in quadratic mean at $t = 0$. The efficient information matrix for $\boldsymbol{\lambda}$ is $\widetilde{\mathbf{I}} = \mathbf{E}(\mathbf{D}'\mathbf{D})\mathbf{H}'\boldsymbol{\Sigma}^{-1}\mathbf{M}\mathbf{H}(\mathbf{D}'\mathbf{D})\mathbf{E}'/4$. The efficient influence function for $\boldsymbol{\lambda}$ is*

$$\widetilde{\boldsymbol{\psi}}(\mathbf{x},\mathbf{y}) = -\frac{1}{2}\widetilde{\mathbf{I}}^{-1}\mathbf{E}(\mathbf{D}'\mathbf{D})\mathbf{H}'\boldsymbol{\Sigma}^{-1}\mathbf{M}\mathbf{g}(\mathbf{x},\mathbf{y},\boldsymbol{\theta}^*).$$

Proof. Begin with computing $\iint \dot{\boldsymbol{\ell}}(\mathbf{x},\mathbf{y})\mathbf{g}(\mathbf{x},\mathbf{y})'p_{xy}(\mathbf{x},\mathbf{y};\boldsymbol{\theta}^*)d\mathbf{x}d\mathbf{y}$. The moment condition and Condition iii) of Assumption A.1.4 implies

$$\begin{aligned}\mathbf{0}_{m\times p} &= \iint \frac{\partial}{\partial\boldsymbol{\theta}'}\mathbf{g}(\mathbf{x},\mathbf{y};\boldsymbol{\theta}^*)p_{xy}(\mathbf{x},\mathbf{y};\boldsymbol{\theta}^*)d\mathbf{x}d\mathbf{y} + \iint \mathbf{g}(\mathbf{x},\mathbf{y};\boldsymbol{\theta}^*)\frac{\partial p_{xy}(\mathbf{x},\mathbf{y};\boldsymbol{\theta}^*)}{\partial\boldsymbol{\theta}'}d\mathbf{x}d\mathbf{y}\\ -\mathbf{G} &= \iint \mathbf{g}(\mathbf{x},\mathbf{y};\boldsymbol{\theta}^*)\frac{\partial p_{xy}(\mathbf{x},\mathbf{y};\boldsymbol{\theta}^*)}{\partial\boldsymbol{\theta}'}d\mathbf{x}d\mathbf{y}.\end{aligned}$$

Condition iv) of Assumption A.1.4 implies

$$\frac{\partial^2}{\partial\boldsymbol{\sigma}\partial\boldsymbol{\sigma}'}\iiint g_r(\mathbf{x},\mathbf{y};\boldsymbol{\theta}^*)p_{xy}(\mathbf{x},\mathbf{y};\boldsymbol{\theta}^* + \mathbf{U}\boldsymbol{\sigma})d\mathbf{x}d\mathbf{y}d\mathbf{u}\mid_{\boldsymbol{\sigma}=\mathbf{0}} = \iint g_r(\mathbf{x},\mathbf{y};\boldsymbol{\theta}^*)\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\Big[p_{xy}(\mathbf{x},\mathbf{y};\boldsymbol{\theta}^*)\Big]d\mathbf{x}d\mathbf{y}.$$

Equation 1.3.1, Lemma A.1.3, and Condition ii) of Assumption A.1.2 implies that the integral can equivalently be expressed as

$$\frac{\partial^2}{\partial\boldsymbol{\sigma}\partial\boldsymbol{\sigma}'}\iiint g_r(\mathbf{x},\mathbf{f}(\mathbf{x},\boldsymbol{\varepsilon};\boldsymbol{\theta}^*+\mathbf{U}\boldsymbol{\sigma});\boldsymbol{\theta}^*)p_{x\varepsilon}(\mathbf{x},\boldsymbol{\varepsilon})d\boldsymbol{\varepsilon}d\mathbf{x}d\mathbf{u}\mid_{\boldsymbol{\sigma}=\mathbf{0}}=-\iint\frac{\partial^2 g_r(\mathbf{x},\mathbf{y}^0;\boldsymbol{\theta}^*)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}p_{x\varepsilon}(\mathbf{x},\boldsymbol{\varepsilon})d\mathbf{x}d\boldsymbol{\varepsilon}.$$

Then, stacking moment conditions,

$$-\frac{\mathbf{H}(\mathbf{D}'\mathbf{D})}{2}=\frac{1}{2}\iint\mathbf{g}(\mathbf{x},\mathbf{y};\boldsymbol{\theta}^*)\frac{1}{p_{xy}(\mathbf{x},\mathbf{y})}\text{vech}\Big(\frac{\partial^2 p_{xy}(\mathbf{x},\mathbf{y};\boldsymbol{\theta}^*)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\Big)(\mathbf{D}'\mathbf{D})p_{xy}(\mathbf{x},\mathbf{y})d\mathbf{x}d\mathbf{y}.$$

Thus $\iint\dot{\boldsymbol{\ell}}(\mathbf{x},\mathbf{y})\mathbf{g}(\mathbf{x},\mathbf{y};\boldsymbol{\theta}^*)'p_{xy}(\mathbf{x},\mathbf{y};\boldsymbol{\theta}^*)d\mathbf{x}d\mathbf{y}=\begin{bmatrix}-\mathbf{G}'\\ -\frac{1}{2}\mathbf{E}(\mathbf{D}'\mathbf{D})\mathbf{H}'\end{bmatrix}$.

Next, verify the guess for the conjectured influence function for $\text{vech}(\boldsymbol{\Lambda})$:

$$\begin{aligned}&\iint\widetilde{\boldsymbol{\psi}}(\mathbf{x},\mathbf{y})(\mathbf{a}'\dot{\boldsymbol{\ell}}(\mathbf{x},\mathbf{y})+\eta(\mathbf{x},\mathbf{y}))p_{xy}(\mathbf{x},\mathbf{y};\boldsymbol{\theta}^*)d\mathbf{x}d\mathbf{y}\\&\quad=-\frac{1}{2}\widetilde{\mathbf{I}}^{-1}\mathbf{E}(\mathbf{D}'\mathbf{D})\mathbf{H}'\boldsymbol{\Sigma}^{-1}\mathbf{M}\Big[\iint\mathbf{g}(\mathbf{x},\mathbf{y};\boldsymbol{\theta}^*)\dot{\boldsymbol{\ell}}(\mathbf{x},\mathbf{y})'p_{xy}(\mathbf{x},\mathbf{y};\boldsymbol{\theta}^*)d\mathbf{x}d\mathbf{y}\Big]\mathbf{a}\\&\quad=\frac{1}{4}\widetilde{\mathbf{I}}^{-1}\mathbf{E}(\mathbf{D}'\mathbf{D})(\mathbf{M}\mathbf{H})'\boldsymbol{\Sigma}^{-1}\mathbf{H}(\mathbf{D}'\mathbf{D})\mathbf{E}'\boldsymbol{\lambda}=\boldsymbol{\lambda}.\end{aligned}$$

The final line uses $\mathbf{MG}=\mathbf{0}$. The efficient information matrix and score function for $\boldsymbol{\lambda}$ immediately follow.

□

Proof of Theorem 1.4.1

Proof. These proofs closely follow and generalize the arguments of Chapter 25.6 of van der Vaart (1998) to multivariate testing and are applicable to the score functions that arise in my specific settings.

Let $(\mathbf{c}',\boldsymbol{\lambda}')'\dot{\boldsymbol{\ell}}+\eta$ be an element of the tangent set. Then, the functional of interest is

$$\boldsymbol{\psi}(P_{t,(\mathbf{c}',\boldsymbol{\lambda}')'\dot{\boldsymbol{\ell}}+\eta})=t\mathbb{E}[\widetilde{\boldsymbol{\psi}}((\mathbf{c}',\boldsymbol{\lambda}')'\dot{\boldsymbol{\ell}}+\eta)]+o(t)=t\boldsymbol{\lambda}+o(t)$$

where the second equality follows from the form of the efficient influence function (see Lemma A.1.7). The existence of the efficient influence function implies that the functional $\boldsymbol{\psi}$ is differentiable at P relative to the tangent space. For what follows, I consider paths of the form $t = h/\sqrt{n}$.

1. Beginning with the first part of the theorem, I generalize Theorem 25.44 of van der Vaart (1998) to multivariate testing and follow the argument closely. Fix h_1 and arbitrary $(\mathbf{c}_1', \boldsymbol{\lambda}_1')'\dot{\boldsymbol{\ell}} + \eta_1$. Assume that $\mathbb{E}[((\mathbf{c}_1', \boldsymbol{\lambda}_1')'\dot{\boldsymbol{\ell}} + \eta_1)^2] = 1$ to ease notation.

   Define the orthonormal base $\mathcal{G} = ((\mathbf{c}_1', \boldsymbol{\lambda}_1')'\dot{\boldsymbol{\ell}} + \eta_1, \ldots, (\mathbf{c}_k', \boldsymbol{\lambda}_k')'\dot{\boldsymbol{\ell}} + \eta_k)'$ of an arbitrary finite-dimensional subspace of the tangent space. Note that the first element of the orthonormal base is $(\mathbf{c}_1', \boldsymbol{\lambda}_1')'\dot{\boldsymbol{\ell}} + \eta_1$. Lemma 25.14 of van der Vaart (1998) implies that for any $(\mathbf{c}', \boldsymbol{\lambda}')'\dot{\boldsymbol{\ell}} + \eta \in \operatorname{lin}\mathcal{G}$, $P_{t,(\mathbf{c}',\boldsymbol{\lambda}')'\dot{\boldsymbol{\ell}}+\eta}$ is locally asymptotically normal at $t = 0$. Let S^{k-1} be the unit sphere of $\mathbb{R}^k$. In the sense of the convergence of experiments,

$$(P^n_{h/\sqrt{n},\mathbf{b}'\mathcal{G}} \colon h > 0, \mathbf{b} \in S^{k-1}) \rightsquigarrow \Big(N_m(h\mathbf{b}, \mathbf{I}) \colon h > 0, \mathbf{b} \in S^{k-1}, \sum_{i=1}^{k} b_k \boldsymbol{\Lambda}_k \in \mathcal{C}\Big).$$

   The last restriction restricts $\mathbf{b}$ so that $\mathbf{b}'\mathcal{G}$ lies within the tangent set. For score function $(\mathbf{c}_1', \boldsymbol{\lambda}_1')'\dot{\boldsymbol{\ell}} + \eta_1$ and $h = h_1$, fix a subsequence for which $\limsup_{n\to\infty} \pi_n(P_{h_1 n^{-1/2},(\mathbf{c}_1',\boldsymbol{\lambda}_1')'\dot{\boldsymbol{\ell}}+\eta_1})$ is taken. For each $(h, \mathbf{b})$, contiguity implies that there exists another subsequence along which the functions $\pi_n(P_{h/\sqrt{n}}, \mathbf{b})$ converges pointwise to a limit $\pi(h, \mathbf{b})$. Theorem 15.1 of van der Vaart (1998) implies that the function $\pi(h, \mathbf{b})$ is the power function of a test in the normal limit experiment $\mathbf{Z}_k \sim \mathcal{N}(\mathbb{E}[\widetilde{\boldsymbol{\psi}}\mathcal{G}']\mathbf{b}, \mathbb{E}[\widetilde{\boldsymbol{\psi}}\mathcal{G}']\mathbb{E}[\mathcal{G}\widetilde{\boldsymbol{\psi}}])$. By choosing a sufficiently large base, the covariance of the limit experiment is arbitrarily close to $\mathbb{E}[\widetilde{\boldsymbol{\psi}}\widetilde{\boldsymbol{\psi}}']$. Now choose $(h, \mathbf{b}) = (1, \mathbf{e}_1)$, denoting $\pi(\boldsymbol{\lambda}_1)$ as the power function in the normal limit experiment

$$\mathbf{Z} \sim \mathcal{N}(\boldsymbol{\lambda}_1, 4\boldsymbol{\Omega}^{-1}) \tag{A.1.5}$$

where $\mathbb{E}[\widetilde{\boldsymbol{\psi}}\widetilde{\boldsymbol{\psi}}'] = 4\boldsymbol{\Omega}^{-1}$ from Lemma A.1.7. $\pi(\boldsymbol{\lambda}_1)$ is level-α since π_n is level α in the local experiment for each n. Now fix local parameter $\mathbf{c}$ and $\eta \in \dot{\mathcal{P}}$. Since $\pi^*(\boldsymbol{\lambda}; r)$ is the power function for the level-α test that maximizes a weighted average power criterion with weight function $w_r(\boldsymbol{\lambda})$

$$\int \pi(\boldsymbol{\lambda}) w_r(\boldsymbol{\lambda}) d\boldsymbol{\lambda} \leq \int \pi^*(\boldsymbol{\lambda}; r) w_r(\boldsymbol{\lambda}) d\boldsymbol{\lambda}.$$

Since $\limsup_{n\to\infty} \pi_n(P_{h/\sqrt{n},(\mathbf{c}',\boldsymbol{\lambda}')'\dot{\boldsymbol{\ell}}+\eta}) \leq \pi(\boldsymbol{\lambda})$, the desired result follows taking $r \to \infty$.

2. For the second part of the theorem, define $\mathbf{Z}_n = -2\widehat{\boldsymbol{\Omega}}^{-1}\mathbf{S}_n/\sqrt{n}$. Lemma 25.23 of van der Vaart (1998) implies $\mathbf{T}_n$ is regular at P. Like Lemma 25.45 of van der Vaart (1998), by the efficiency of $\mathbf{T}_n$ and differentiability of $\boldsymbol{\psi}$, $\sqrt{n}T_n$ converges under $P_{1/\sqrt{n},(\mathbf{c}',\boldsymbol{\lambda}')'\dot{\boldsymbol{\ell}}+\eta}$ to $\mathcal{N}(\boldsymbol{\lambda}, \mathbb{E}[\widetilde{\boldsymbol{\psi}}\widetilde{\boldsymbol{\psi}}'])$, which matches the limit experiment. Since $\mathbf{Z}_n'(\widehat{\boldsymbol{\Omega}}/4)\mathbf{Z}_n - \inf_{\boldsymbol{\lambda}:\, \mathrm{vech}^{-1}(\mathbf{E}'\boldsymbol{\lambda})\in\mathcal{C}} \Big[\mathbf{Z}_n - \boldsymbol{\lambda}\Big]'(\widehat{\boldsymbol{\Omega}}/4)\Big[\mathbf{Z}_n - \boldsymbol{\lambda}\Big] = \mathbf{S}_n'\widehat{\boldsymbol{\Omega}}^{-1}\mathbf{S}_n - \inf_{\boldsymbol{\lambda}:\, \mathrm{vech}^{-1}(\mathbf{E}'\boldsymbol{\lambda})\in\mathcal{C}} \Big[\mathbf{S}_n + \frac{1}{2}\widehat{\boldsymbol{\Omega}}\boldsymbol{\lambda}\Big]'\widehat{\boldsymbol{\Omega}}^{-1}\Big[\mathbf{S}_n + \frac{1}{2}\widehat{\boldsymbol{\Omega}}\boldsymbol{\lambda}\Big]$, the likelihood ratio test of $\mathbf{S}_n$ matches the corresponding likelihood ratio test of the limit experiment.

Now, I will show that the likelihood ratio test is equivalent to a test that maximizes weighted average power for distant alternatives. Observe that the multivariate Gaussian shift experiment of Equation A.1.5 for arbitrary $\boldsymbol{\lambda}$ can be rescaled and written as a Gaussian linear regression model. The dependent variable is $\overline{\mathbf{Z}} = \frac{1}{2}\boldsymbol{\Omega}^{1/2\prime}\mathbf{Z}$, independent variable is $\frac{1}{2}\boldsymbol{\Omega}^{1/2\prime}$, coefficients $\boldsymbol{\lambda}$, and errors are standard normal

$$\overline{\mathbf{Z}} = \frac{1}{2}\boldsymbol{\Omega}^{1/2\prime}\boldsymbol{\lambda} + \boldsymbol{\eta}, \quad \boldsymbol{\eta} \sim \mathcal{N}(\mathbf{0}_{\frac{p(p+1)}{2}\times 1}, \mathbf{I}_{\frac{p(p+1)}{2}\times\frac{p(p+1)}{2}}).$$

Since the alternative hypothesis parameter space is positively homogeneous (that $\boldsymbol{\Lambda} \in \mathcal{C}$ implies $a\boldsymbol{\Lambda} \in \mathcal{C}$ for positive scalar a), Theorem 1 of Andrews (1996) applies. A test based on the weighted average power criterion for radius $r > 0$ is equivalent to finding the best test that distinguishes between the following simple null and simple alternative hypotheses

$$H_0\colon f_z(\overline{\mathbf{Z}}|\mathbf{0}) \text{ versus } H_1\colon \int f_z(\overline{\mathbf{Z}}|\boldsymbol{\lambda}) w_r(\boldsymbol{\lambda}) d\boldsymbol{\lambda}.$$

Let LR_r be the corresponding likelihood ratio test statistic under a weighted average power criterion with radius r. Then (up to normalization) Andrews (1996) shows that the $\lim_{r\to\infty} LR_r$ is equivalent to the generalized likelihood ratio test statistic for $\overline{\mathbf{Z}}$ (and consequently for $\mathbf{Z}$)

$$\begin{aligned}\mathcal{LR} &= \overline{\mathbf{Z}}'\overline{\mathbf{Z}} - \inf_{\boldsymbol{\lambda}\colon \mathrm{vech}^{-1}(\mathbf{E}'\boldsymbol{\lambda})\in\mathcal{C}} (\overline{\mathbf{Z}} - \boldsymbol{\Omega}^{1/2\prime}\boldsymbol{\lambda})'(\overline{\mathbf{Z}} - \boldsymbol{\Omega}^{1/2\prime}\boldsymbol{\lambda})) \\ &= \mathbf{Z}'(\boldsymbol{\Omega}/4)\mathbf{Z} - \inf_{\boldsymbol{\lambda}\colon \mathrm{vech}^{-1}(\mathbf{E}'\boldsymbol{\lambda})\in\mathcal{C}} \Big[\mathbf{Z} - \boldsymbol{\lambda}\Big]'(\boldsymbol{\Omega}/4)\Big[\mathbf{Z} - \boldsymbol{\lambda}\Big].\end{aligned}$$

□

Corollary A.1.1. *Impose Assumptions 1.3.3, 1.3.4, A.1.4, A.1.1, and A.1.2. Suppose* $P_{\boldsymbol{\theta}_t,\boldsymbol{\sigma}_t,\mathrm{vech}(\mathbf{C}),p_{xy,t},p_{u,t}}$ *is differentiable in quadratic mean at* $t = 0$ *and that* θ *is a scalar.*

a) For every sequence of power functions $P \mapsto \pi_n(P)$ *of level-*α *tests of Equation 1.4.3 for* $\Lambda > 0$*,*

$$\limsup_{n\to\infty} \pi_n(P_{\Lambda,n^{-1/2}}) \leq 1 - \Phi\left(z_{1-\alpha} - \frac{\Lambda}{\sqrt{\widetilde{I}^{-1}}}\right).$$

b) Suppose $\phi_n^{\mathrm{het}} = 1$ *when the level* α *one-dimensional heterogeneity test rejects and 0 otherwise. Then,*

$$\lim_{n\to\infty} \mathbb{P}_{\Lambda,n^{-1/2}}(\phi_n = 1) = 1 - \Phi\left(z_{1-\alpha} - \frac{\Lambda}{\sqrt{\widetilde{I}^{-1}}}\right). \tag{A.1.6}$$

Proof. Part (i) follows from applying Lemma 25.44 of van der Vaart (1998), noting that the scalar limit experiment is $z \sim \mathcal{N}(\Lambda, \widetilde{I}^{-1})$. Part (ii) follows from applying Theorem 1.4.1 to the scalar case. $\square$

A.2 Additional empirical results

A.2.1 Production function estimation

ISIC4	Description	ϕ^m	SE(ϕ^m)	p_J	p_{HNS}	p_{het}	n
3111	Slaughtering, preparing and preserving meat	0.80	0.05	0.13	0.469	0.807	434
3112	Manufacture of dairy products	0.77	0.02	0.57	0.917	0.979	290
3113	Canning and preserving of fruits and vegetables	0.84	0.03	0.06	0.298	0.394	301
3114	Canning, preserving and processing of fish, crustaces and similar foods	0.82	0.02	0.05	0.257	0.154	465
3115	Manufacture of vegetable and animal oils and fats	0.86	0.03	0.23	0.637	0.068	260
3116	Grain mill products	0.94	0.01	0.00	0.004	$<.001$	381
3117	Manufacture of bakery products	0.76	0.03	0.00	$<.001$	$<.001$	3807
3119	Manufacture of cocoa, chocolate and sugar confectionery	0.73	0.15	0.67	0.957	0.206	114
3121	Manufacture of food products not elsewhere classified	0.68	0.09	0.00	0.033	0.080	289
3122	Manufacture of prepared animal feeds	0.89	0.04	0.21	0.608	0.157	78
3131	Distilling, rectifying and blending spirits	0.69	0.05	0.86	0.994	0.805	67
3132	Wine industries	0.68	0.03	0.29	0.708	0.003	210
3133	Malt liquors and malt	0.49	0.02	0.00	0.027	$<.001$	54
3134	Soft drinks and carbonated waters industries	0.84	0.01	0.03	0.190	$<.001$	129
3211	Spinning, weaving and finishing textiles	0.66	0.01	0.00	0.035	0.039	622
3212	Manufacture of made-up textile goods except wearing apparel	0.75	0.07	0.46	0.861	0.032	131
3213	Knitting mills	0.66	0.02	0.00	0.014	0.150	610
3214	Manufacture of carpets and rugs	0.76	0.04	0.36	0.778	$<.001$	69
3220	Manufacture of wearing apparel, except footwear	0.76	0.02	0.00	0.004	0.036	1332
3240	Manufacture of footwear, except vulcanized or moulded rubber or plastic footwear	0.68	0.02	0.00	0.027	0.009	709
3311	Sawmills, planing and other wood mills	0.76	0.03	0.16	0.527	0.008	1301
3312	Manufacture of wooden and cane containers and small cane ware	0.83	0.09	0.18	0.556	$<.001$	61
3319	Manufacture of wood and cork products not elsewhere classified	0.75	0.02	0.44	0.844	$<.001$	76
3320	Manufacture of furniture and fixtures, except primarily of metal	0.66	0.04	0.96	0.999	0.933	499
3411	Manufacture of pulp, paper and paperboard	0.73	0.02	0.05	0.263	$<.001$	78
3412	Manufacture of containers and boxes of paper and paperboard	0.81	0.03	0.47	0.864	$<.001$	118
3419	Manufacture of pulp, paper and paperboard articles not elsewhere classified	0.78	0.02	0.02	0.111	$<.001$	121
3420	Printing, publishing and allied industries	0.56	0.05	0.24	0.648	$<.001$	933

Table A.2.1: Estimation results by ISIC4 category (for ISIC2=31-34). ϕ^m and SE(ϕ^m) give the point estimate and standard errors respectively of the output elasticity of materials. p_J, p_{HNS}, and p_{het} give p-values for the J test, Hahn et al. (2014), and heterogeneity tests respectively. The final column lists the sample size.

ISIC4	Description	ϕ^m	SE(ϕ^m)	p_J	p_{HNS}	p_{het}	n
3521	Manufacture of paints, varnishes and laquers	0.55	0.01	0.01	0.046	<.001	154
3522	Manufacture of drugs and medicines	0.57	0.06	0.32	0.742	0.605	233
3523	Manufacture of soap and cleaning preparations, perfumes, cosmetics and other toilet preparations	0.84	0.02	0.46	0.859	<.001	202
3529	Manufacture of chemical products not elsewhere classified	0.76	0.03	0.06	0.284	0.136	256
3560	Manufacture of plastic products not elsewhere classified	0.70	0.02	0.00	<.001	<.001	969
3691	Manufacture of structural clay products	0.60	0.02	0.01	0.053	<.001	111
3692	Manufacture of cement, lime and plaster	0.63	0.06	0.01	0.082	<.001	40
3693	Manufacture of sports goods	0.75	0.10	0.02	0.141	<.001	288
3699	Manufacture of non-metallic mineral products not elsewhere classified	0.66	0.03	0.06	0.271	<.001	85
3811	Manufacture of cutlery, hand tools and general hardware	0.53	0.08	0.36	0.778	<.001	104
3812	Manufacture of furniture and fixtures primarily of metal	0.69	0.02	0.00	0.013	<.001	131
3813	Manufacture of structural metal products	0.73	0.02	0.03	0.195	<.001	467
3814	Organic composite solvents and thinners, not elsewhere specified or included; prepared paint or varnish removers	0.66	0.07	0.55	0.911	0.449	323
3815	Reaction initiators, reaction accelerators and catalytic preparations n.e.c. or included	0.67	0.03	0.17	0.544	<.001	137
3819	Manufacture of fabricated metal products except machinery and equipment not elsewhere classified	0.56	0.02	0.01	0.055	<.001	326
3822	Manufacture of agricultural machinery and equipment	0.71	0.04	0.13	0.457	<.001	75
3823	Manufacture of metal and wood working machinery	0.57	0.07	0.00	0.005	<.001	57
3824	Manufacture of special industrial machinery and equipment except metal and wood working machinery	0.45	0.12	0.12	0.451	0.819	107
3829	Machinery and equipment except electrical not elsewhere classified	0.74	0.02	0.00	0.008	<.001	331
3841	Ship building and repairing	0.56	0.12	0.00	<.001	<.001	56
3843	Manufacture of motor vehicles	0.74	0.02	0.02	0.129	<.001	270
3844	Manufacture of motorcycles and bicycles	0.85	0.12	0.06	0.278	<.001	40

Table A.2.2: Estimation results by ISIC4 category (for ISIC2=35-38). ϕ^m and SE(ϕ^m) give the point estimate and standard errors respectively of the output elasticity of materials. p_J, p_{HNS}, and p_{het} give p-values for the J test, Hahn et al. (2014), and heterogeneity tests respectively. The final column lists the sample size.

Appendix B

Appendix to Chapter 2

B.1 Proofs

B.1.1 Proof of Theorem 2.3.1

I will begin by showing that the hypotheses of Theorem 3.1 of Pakes and Pollard (1989) are satisfied by Assumption 2.3.1. Doing so gives consistency of a GMM estimator with moment function $g(\mathbf{r}_t, \boldsymbol{\phi})$ and the weight matrix taken to be the identity matrix. To do so, I proceed by verifying the (more stringent) conditions of Newey and McFadden (1994) Theorem 2.6, taking the weight matrix as the identity matrix. I close by showing that the constructed weight matrix $\widehat{W}(\boldsymbol{\phi})$ satisfies Lemma 3.4 of Pakes and Pollard (1989), implying consistency of the RGIV estimator.

Below, I verify the conditions of Theorem 2.6 of Newey and McFadden (1994) for an identity weight matrix $I_{n(n-1)/2}$ (preserving the same numbering as the original reference):

i. Since the weight matrix $\widehat{W} = I_{n(n-1)/2}$ is taken to be the identity matrix, trivially $\widehat{W} \xrightarrow{p} I_{n(n-1)/2}$. Lemma 2.3.1 shows $\mathbb{E}[g(\mathbf{r}_t, \boldsymbol{\phi})] = 0$ if and only if $\boldsymbol{\phi} = \boldsymbol{\phi}_0$.

ii. Parameter space $\boldsymbol{\Phi}$ is compact from Assumption 2.3.1(iii).

iii. By inspection, $g(\mathbf{r}_t, \boldsymbol{\phi})$ is continuous at each $\boldsymbol{\phi}$ with probability 1.

iv. Take $\widetilde{\boldsymbol{\phi}} \in \boldsymbol{\Phi}$. Then, consider the element of $g(\mathbf{r}_t, \widetilde{\boldsymbol{\phi}})$ corresponding to the orthogonality of shocks to units i and j. Then, Assumption 2.3.1(i) implies $(r_{it} - \widetilde{\phi}_i r_{St})(r_{jt} - \widetilde{\phi}_j r_{St}) = [u_{it} + (\phi_i - \widetilde{\phi}_i) r_{St}][u_{jt} + (\phi_j - \widetilde{\phi}_j) r_{St}] = [u_{it} + (\phi_i - \widetilde{\phi}_i)\frac{u_{St}}{1-\phi_S}][u_{jt} + (\phi_j - \widetilde{\phi}_j)\frac{u_{St}}{1-\phi_S}]$. Thus, the moment condition is quadratic in idiosyncratic shocks. Since $\mathbb{E}[\|\mathbf{u}_t\|^2] < \infty$, $\mathbb{E}[\sup_{\phi \in \boldsymbol{\Phi}} \|g(\mathbf{r}, \boldsymbol{\phi})\|] < \infty$.

Since the more stringent conditions of Newey and McFadden (1994) Theorem 2.6 are satisfied, the conditions of Theorem 3.1 of Pakes and Pollard (1989) are also satisfied.

Since $\widehat{W}(\boldsymbol{\phi})$ is positive definite, it can be decomposed as $\widehat{W}(\boldsymbol{\phi}) = \widehat{A}(\boldsymbol{\phi})'\widehat{A}(\boldsymbol{\phi})$ where $\widehat{A}(\boldsymbol{\phi}) = \sqrt{\widehat{W}(\boldsymbol{\phi})}$ where $\sqrt{\cdot}$ represents the element-wise square root operator. Next, proceed by verifying the conditions of Lemma 3.4 of Pakes and Pollard (1989) (preserving the same numbering as the original reference). Conditions (i) and (ii) follow from the law of large numbers, continuous mapping theorem, and $0 < \sigma_i^2 < \infty$. Hence, RGIV is consistent.

B.1.2 Proof of Theorem 2.3.2

I will first show that the hypotheses of Theorem 3.3 of Pakes and Pollard (1989) are satisfied by Assumption 2.3.1. Doing so gives asymptotic normality of a GMM estimator with moment function $g(\mathbf{r}_t, \boldsymbol{\phi})$ and the weight matrix taken to be the identity matrix. I begin by verifying the (more stringent) conditions of Newey and McFadden (1994) Theorem 3.4 where the weight matrix is taken to be the identity matrix. Then, I close by showing that the estimator with weight matrix $\widehat{W}(\boldsymbol{\phi})$ satisfies Lemma 3.5 of Pakes and Pollard (1989), implying that the RGIV estimator is asymptotically normal with asymptotic variance $(G'WG)^{-1}G'W\Sigma WG(G'WG)^{-1}$.

Below, I verify the conditions of Newey and McFadden (1994) Theorem 3.4. Before doing so, I pre-compute $\nabla_\theta g(\mathbf{r}_t, \widetilde{\boldsymbol{\phi}})$ for $\widetilde{\boldsymbol{\phi}} \in \boldsymbol{\Phi}$

$$\nabla_\theta g(\mathbf{r}_t, \widetilde{\boldsymbol{\phi}}) = \begin{bmatrix} M_2 \\ M_3 \\ \vdots \\ M_n \end{bmatrix} \quad M_i = \begin{bmatrix} 0_{(n-i+1)\times(i-2)} & \begin{matrix} -r_{St}(r_{it} - \widetilde{\phi}_i r_{St}) \\ -r_{St}(r_{(i+1)t} - \widetilde{\phi}_{i+1} r_{St}) \\ \vdots \\ -r_{St}(r_{nt} - \widetilde{\phi}_n r_{St}) \end{matrix} & -I_{n-i+1} \cdot r_{St}(r_{(i-1)t} - \widetilde{\phi}_{i-1} r_{St}) \end{bmatrix}.$$

Also, for more compact notation, let $K = \sum_{i=1}^n S_i^2 \sigma_i^2$ where σ_i^2 is given by Assumption 2.3.1(ii). Moreover, the hypotheses of Newey and McFadden (1994) Theorem 2.6 are satisfied in Theorem 2.3.1. I check the hypotheses of Theorem 3.4 of Newey and McFadden (1994) for identity weight matrix $\widehat{W} = I_{n(n-1)/2}$ (preserving the numbering of the original reference):

i. By Assumption 2.3.1(iii), $\boldsymbol{\phi}_0 \in \text{interior}(\boldsymbol{\Phi})$.

ii. Inspecting the functional form of $\nabla_\theta g(\mathbf{r}, \widetilde{\boldsymbol{\phi}})$, $g(\mathbf{r}, \widetilde{\boldsymbol{\phi}})$ is continuously differentiable in a neighborhood $\mathcal{N}$ of $\boldsymbol{\phi}_0$ with probability approaching 1.

iii. $\mathbb{E}[g(\mathbf{r}, \boldsymbol{\phi}_0)] = 0$ is shown in Lemma 2.3.1. The square of the shock orthogonality condition between countries i and j is bounded $\mathbb{E}[u_{it}^2 u_{jt}^2]^2 \leq \mathbb{E}[u_{it}^4]\mathbb{E}[u_{jt}^4] < \infty$. Thus $\mathbb{E}[\|g(\mathbf{r}, \boldsymbol{\phi}_0)\|^2]$ is finite.

iv. Fix $i \in 1, ..., n$ and $\widetilde{\phi} \in \boldsymbol{\Phi}$. Then

$$\begin{aligned} \| -r_{St}(r_{it} - \widetilde{\phi}_i r_{St})\| &= \| -r_{St}\Big(u_{it} + (\phi_i - \widetilde{\phi}_i) r_{St}\Big)\| && \text{(Assumption 2.3.1(i))} \\ &\leq \|r_{St} u_{it}\| + \|\phi_i - \widetilde{\phi}_i\| \cdot \|r_{St}^2\| = \|\frac{u_{St}}{1-\phi_S} u_{it}\| + \|\phi_i - \widetilde{\phi}_i\| \cdot \|\Big(\frac{u_{St}}{1-\phi_S} u_{it}\Big)^2\|. \end{aligned}$$

The above expression is quadratic in idiosyncratic shocks u_{it}. From Assumption 2.3.1(ii), $\mathbb{E}[\mathbf{u}_t \mathbf{u}_t']$ exists and $\mathbb{E}[\sup_{\boldsymbol{\phi} \in \boldsymbol{\Phi}} \| -r_{St}(r_{it} - \widetilde{\phi}_i r_{St})\|] < \infty$, so $\mathbb{E}[\sup_{\boldsymbol{\phi} \in \boldsymbol{\Phi}} \|\nabla_\phi g(r, \boldsymbol{\phi})\|] < \infty$.

v. To show $G'G$ is full rank, it suffices to show that $\text{rank}(G) = n$. Recall $G = \mathbb{E}[\nabla_\phi g(\mathbf{r}_t, \boldsymbol{\phi}_0)]$. Since $\mathbb{E}[-r_{St}(r_{it} - \phi_i r_{St})] = \mathbb{E}[r_{St} u_{it}] = -\mathbb{E}\Big[\frac{u_{St} u_{it}}{1-\phi_S}\Big] = -\frac{S_i \sigma_i^2}{1-\phi_S}$, G can be explicitly computed:

$$G = \begin{bmatrix} \mathbb{E}(M_2) \\ \mathbb{E}(M_3) \\ \vdots \\ \mathbb{E}(M_n) \end{bmatrix}, \quad \mathbb{E}(M_i) = -\frac{1}{1-\phi_S} \begin{bmatrix} & S_i\sigma_i^2 & \\ 0_{(n-i+1)\times(i-2)} & S_{i+1}\sigma_{i+1}^2 & I_{n-i+1}\cdot S_{i-1}\sigma_{i-1}^2 \\ & \vdots & \\ & S_n\sigma_n^2 & \end{bmatrix}.$$

To show G is full rank, consider the $n \times n$ submatrix $G_{1:n,1:n}$ formed by taking the first n rows of G. Then, the determinant of $G'_{1:n,1:n}$ can be computed by applying elementary row operations:

$$\det(G'_{1:n,1:n}) = \det\left(-\frac{1}{1-\phi_S} \begin{bmatrix} S_2\sigma_2^2 & S_3\sigma_3^2 & \dots & S_n\sigma_n^2 & 0 \\ & & & & S_3\sigma_3^2 \\ & & & & S_2\sigma_2^2 \\ & & S_1\sigma_1^2 I_{n-1} & & 0 \\ & & & & \vdots \\ & & & & 0 \end{bmatrix} \right)$$

$$= (-1)^{n-1}\det\left(-\frac{1}{1-\phi_S} \begin{bmatrix} & & & & & S_3\sigma_3^2 \\ & & & & & S_2\sigma_2^2 \\ & S_1\sigma_1^2 I_{n-1} & & & & 0 \\ & & & & & \vdots \\ 0 & 0 & 0 & \dots & 0 & -\frac{S_3\sigma_3^2}{S_1\sigma_1^2}S_2\sigma_2^2 - \frac{S_2\sigma_2^2}{S_1\sigma_1^2}S_3\sigma_3^2 \end{bmatrix} \right)$$

$$= (-1)^{n-1}\Big(-\frac{1}{1-\phi_S}\Big)^n \big[S_1\sigma_1^2\big]^{n-1}\Big[-\frac{S_3\sigma_3^2}{S_1\sigma_1^2}S_2\sigma_2^2 - \frac{S_2\sigma_2^2}{S_1\sigma_1^2}S_3\sigma_3^2\Big] = 2\Big(\frac{1}{1-\phi_S}\Big)^n (S_1\sigma_1^2)^{n-2}S_2S_3\sigma_2^2\sigma_3^2 \neq 0.$$

The determinant in line 2 is the product of the matrix's diagonal elements because the matrix is upper triangular. Since $1-\phi_S \neq 0$, $S_i > 0$, and $\sigma_i^2 > 0$, the determinant is not zero. Hence, submatrix $G'_{1:n,1:n}$ is full rank, which immediately implies G is also full rank $n = \text{rank}(G'_{1:n,1:n}) \leq \text{rank}(G) \leq \min(\frac{n(n-1)}{2}, n) = n$.

Since the more stringent conditions of Newey and McFadden (1994) Theorem 3.4(i)-(v) are satisfied, the conditions of Theorem 3.3 of Pakes and Pollard (1989) are also satisfied. Continue to verifying the conditions of Lemma 3.5 of Pakes and Pollard (1989). Recall from the proof of Theorem B.1.1 that the weight matrix can be decomposed as $\widehat{W}(\boldsymbol{\phi}) = \widehat{A}(\boldsymbol{\phi})'\widehat{A}(\boldsymbol{\phi})$ where $\widehat{A}(\boldsymbol{\phi}) = \operatorname{diag}\Big(\sqrt{\frac{1}{T}\sum_{t=1}^{T} g(\mathbf{r}_t, \boldsymbol{\phi})^{\circ 2}}\Big)^{-1}$ for element-wise square root operator $\sqrt{\cdot}$. From the continuous mapping theorem and law of large numbers, $\widehat{A}(\boldsymbol{\phi}_0) \xrightarrow{p} \operatorname{diag}(\sigma_1\sigma_2, \ldots, \sigma_1\sigma_n, \sigma_2\sigma_3, \ldots, \sigma_{n-1}\sigma_n) = A$. Let $\{\delta_n\}$ be a sequence of positive numbers that converges to zero. Then, $\sup_{\|\boldsymbol{\phi}-\boldsymbol{\phi}_0\|<\delta_n} \|\widehat{A}(\boldsymbol{\phi}) - A\| = o_p(1)$ from continuity of the map from $\boldsymbol{\phi}$ to $\widehat{A}(\boldsymbol{\phi})$ at $\boldsymbol{\phi} = \boldsymbol{\phi}_0$. Hence, the RGIV estimator is asymptotically normal with asymptotic variance $(G'WG)^{-1}G'W\Sigma WG(G'WG)^{-1}$ for $W = A(\boldsymbol{\phi}_0)'A(\boldsymbol{\phi}_0) = \operatorname{diag}(\sigma_1^2\sigma_2^2, \ldots, \sigma_1^2\sigma_n^2, \sigma_2^2\sigma_3^2, \ldots, \sigma_{n-1}^2\sigma_n^2)$.

B.1.3 Proof of Proposition 2.3.1

Before computing the GIV estimand, $\mathbb{E}[(u_{St} - u_{Et})u_{Et}]$, $\mathbb{E}[z_t u_{Et}]$, and $\mathbb{E}[(r_{St} - r_{Et})r_{St}]$ are precomputed below:

$$\mathbb{E}[(u_{St} - u_{Et})u_{Et}] = \frac{1}{n}\sum_{i=1}^{n} = \mathbb{E}\Big[(\sum_{i=1}^{n}(S_i - \frac{1}{n})u_{it})\frac{1}{n}\sum_{i=1}^{n} u_{it}\Big] = \frac{1}{n}\sum_{i=1}^{n}(S_i - \frac{1}{n})\sigma^2 = 0.$$

$$\begin{aligned}\mathbb{E}[z_t u_{Et}] &= \mathbb{E}[(r_{St} - r_{Et})u_{Et}] = \mathbb{E}[(\phi_S - \phi_E)r_{St}u_{Et} + (u_{St} - u_{Et})u_{Et}] \\ &= (\phi_S - \phi_E)\mathbb{E}[r_{St}u_{Et}] = \frac{\phi_S - \phi_E}{1 - \phi_S}\mathbb{E}[u_{St}u_{Et}] = \frac{\phi_S - \phi_E}{1 - \phi_S}\frac{1}{n}\sigma^2.\end{aligned}$$

$$\mathbb{E}[(r_{St} - r_{Et})r_{St}] = (\phi_S - \phi_E)\mathbb{E}[r_{St}^2] + \mathbb{E}[r_{St}(u_{St} - u_{Et})] = \frac{(\phi_S - \phi_E)\sigma^2\sum_{i=1}^{n} S_i^2}{(1 - \phi_S)^2} + \frac{\sigma^2}{1 - \phi_S}[\sum_{i=1}^{n} S_i^2 - \frac{1}{n}].$$

Combining, the GIV estimand is $\frac{\mathbb{E}[z_t r_{Et}]}{\mathbb{E}[z_t r_{St}]} = \frac{\mathbb{E}[z_t(\phi_E r_{St} + u_{Et})]}{\mathbb{E}[z_t r_{St}]} = \phi_E + \frac{\phi_S - \phi_E}{n}\frac{1}{\frac{\phi_S - \phi_E}{1 - \phi_S}\sum_{i=1}^{n} S_i^2 - \frac{1}{n} + \sum_{i=1}^{n} S_i^2}$.

B.1.4 Proof of Lemma 2.3.1

I prove Lemma 2.3.1 under the weaker condition that the covariance of shocks i and j are known. The below proof will use Assumption 2.3.1 after replacing Condition (ii) with Condition (ii'):

Condition (ii'). *For $\sigma_i^2 > 0$, shocks $\mathbf{u}_t$ are i.i.d. with moments $\mathbb{E}(\mathbf{u}_t) = 0$, $\mathbb{E}(u_{it}^2) = \sigma_i^2$, and $\mathbb{E}(\|\mathbf{u}_t\|^4) < \infty$. Moreover, $\mathbb{E}(u_{it}u_{jt}) = \mu_{ij}$ for known μ_{ij}.*

Take $\widetilde{\boldsymbol{\phi}} \in \boldsymbol{\Phi}$. The moment condition for the idiosyncratic shocks of units i and j

$$\mu_{ij} = \mathbb{E}[(r_{it} - \widetilde{\phi}_i r_{St})(r_{jt} - \widetilde{\phi}_j r_{St})] = \mathbb{E}[((\phi_i - \widetilde{\phi}_i) r_{St} + u_{it})((\phi_j - \widetilde{\phi}_j) r_{St} + u_{jt})]$$

(Assumption 2.3.1(i))

$$= (\phi_i - \widetilde{\phi}_i)\mathbb{E}[r_{St}u_{jt}] + (\phi_j - \widetilde{\phi}_j)\mathbb{E}[r_{St}u_{it}] + (\phi_i - \widetilde{\phi}_i)(\phi_j - \widetilde{\phi}_j)\mathbb{E}[r_{St}^2] + \mu_{ij}. \tag{B.1.1}$$

Rearranging, the above display implies

$$\phi_i - \widetilde{\phi}_i = -\frac{(\phi_j - \widetilde{\phi}_j)\mathbb{E}[r_{St}u_{it}]}{\mathbb{E}[r_{St}u_{jt}] + (\phi_j - \widetilde{\phi}_j)\mathbb{E}[r_{St}^2]} \text{ and } \phi_k - \widetilde{\phi}_k = -\frac{(\phi_1 - \widetilde{\phi}_1)\mathbb{E}[r_{St}u_{kt}]}{\mathbb{E}[r_{St}u_{1t}] + (\phi_1 - \widetilde{\phi}_1)\mathbb{E}[r_{St}^2]} \tag{B.1.2}$$

for $k > 1$. Substituting the above display into Equation B.1.1 and rearranging gives

$$0 = \frac{(\phi_1 - \widetilde{\phi}_1)\mathbb{E}[r_{St}u_{it}]\mathbb{E}[r_{St}u_{jt}]}{\mathbb{E}[r_{St}u_{1t}] + (\phi_1 - \widetilde{\phi}_1)\mathbb{E}[r_{St}^2]} \left\{-2 + \frac{(\phi_1 - \widetilde{\phi}_1)\mathbb{E}[r_{St}^2]}{\mathbb{E}[r_{St}u_{1t}] + (\phi_1 - \widetilde{\phi}_1)\mathbb{E}[r_{St}^2]}\right\}. \tag{B.1.3}$$

Equation B.1.3 equals zero when either the first or second terms equal zero. Beginning with the first case, the first multiplicative term equals zero when $\widetilde{\phi}_1 = \phi_1$. Applying Equation B.1.2, $\widetilde{\phi}_k = \phi_k$. Thus, the first case corresponds to the true solution $\widetilde{\boldsymbol{\phi}} = \boldsymbol{\phi}$.

Focusing on the second case, the second multiplicative term of Equation B.1.3 equals zero when $0 = -2 + \frac{(\phi_1 - \widetilde{\phi}_1)\mathbb{E}[r_{St}^2]}{\mathbb{E}[r_{St}u_{1t}] + (\phi_1 - \widetilde{\phi}_1)\mathbb{E}[r_{St}^2]}$. Rearranging terms and substituting into Equation B.1.2 implies $\widetilde{\phi}_S = \phi_1 - \widetilde{\phi}_1 = -2\frac{\mathbb{E}[r_{St}u_{1t}]}{\mathbb{E}[r_{St}^2]}$ and $\phi_k - \widetilde{\phi}_k = -2\frac{\mathbb{E}[r_{St}u_{kt}]}{\mathbb{E}[r_{St}^2]}$ for $k > 1$. After taking a size-weighted sum of $\widetilde{\phi}_k$, $\sum_{i=1}^n S_i\widetilde{\phi}_i = \phi_S + 2(1 - \phi_S) = 1 + (1 - \phi_S) > 1$. The final inequality follows from the $\phi_S < 1$ condition from Assumption 2.3.1(i). In words, the second multiplicative term of Equation B.1.3 equals zero for a value of $\widetilde{\boldsymbol{\phi}}$ that is outside of the parameter space. Hence, the only solution to $\mathbb{E}[g(\mathbf{r}_t, \widetilde{\boldsymbol{\phi}})] = 0$ is $\boldsymbol{\phi} = \boldsymbol{\phi}_0$.

B.1.5 Proof of Lemma 2.5.1

The size-weighted outcome variable can be written as idiosyncratic shocks and the common factor $r_{St} = \phi_S r_{St} + \lambda_S f_t + u_{St} = \frac{1}{1-\phi_S}[\lambda_S f_t + u_{St}]$. For true parameter values $\boldsymbol{\theta} = [\boldsymbol{\phi}', \boldsymbol{\lambda}']'$. The below results will show that the system of equations given by the moment conditions doesn't admit a unique solution; it needn't be true that $\widetilde{\boldsymbol{\theta}} = \boldsymbol{\theta}$.

For units $i \neq j$, the moment condition can be written as

$$\begin{aligned}
\widetilde{\lambda}_i\widetilde{\lambda}_j &= \mathbb{E}\{[r_{it} - \widetilde{\phi}_i r_{St}][r_{jt} - \widetilde{\phi}_j r_{St}]\} = \mathbb{E}\{[(\phi_i - \widetilde{\phi}_i)r_{St} + \lambda_i f_t + u_{it}][(\phi_i - \widetilde{\phi}_j)r_{St} + \lambda_j f_t + u_{jt}]\} \\
&= (\phi_i - \widetilde{\phi}_i)(\phi_j - \widetilde{\phi}_j)\Big[\frac{\lambda_S^2 + \sum_{k=1}^n S_k^2\sigma_k^2}{1-\phi_S}\Big]^2 + (\phi_i - \widetilde{\phi}_i)\lambda_j\frac{\lambda_S}{1-\phi_S} \\
&\quad + (\phi_j - \widetilde{\phi}_j)\lambda_i\frac{\lambda_S}{1-\phi_S} + (\phi_i - \widetilde{\phi}_i)\frac{S_j\sigma_j^2}{1-\phi_S} + (\phi_j - \widetilde{\phi}_j)\frac{S_i\sigma_i^2}{1-\phi_S} + \lambda_i\lambda_j
\end{aligned} \tag{B.1.4}$$

One solution to above is $\widetilde{\boldsymbol{\theta}} = \boldsymbol{\theta}$. Next, guess that another solution satisfies the restrictions $\widetilde{\phi}_1 = \phi_1$ and $\widetilde{\lambda}_1 = \lambda_1$. From this guess, ϕ_k for $k > 1$ can be expressed in terms of ϕ_1

$$\phi_k - \widetilde{\phi}_k = \frac{\widetilde{\lambda}_k\widetilde{\lambda}_1 - \lambda_k\lambda_1 - (\phi_1 - \widetilde{\phi}_1)\left[\frac{\lambda_k\lambda_S + S_k\sigma_k^2}{1-\phi_S}\right]}{\frac{\phi_1 - \widetilde{\phi}_1}{(1-\phi_S)^2}K + \frac{\lambda_1\lambda_S + S_1\sigma_1^2}{1-\phi_S}} = \lambda_1(1-\phi_S)\frac{\widetilde{\lambda}_k - \lambda_k}{\lambda_1\lambda_S + S_1\sigma_1^2}. \tag{B.1.5}$$

The final equality uses the restrictions $\widetilde{\phi}_1 = \phi_1$ and $\widetilde{\lambda}_1 = \lambda_1$. Storing constants as $K_j = \frac{\lambda_1(\lambda_j\lambda_S + S_j\sigma_j^2)}{\lambda_1\lambda_S + S_1\sigma_1^2}$ for $j > 1$ and $K_0 = \frac{\lambda_1^2 K}{(\lambda_1\lambda_S + S_1\sigma_1^2)^2}$, substituting Equation B.1.5 into Equation B.1.4 gives

$$\widetilde{\lambda}_i\widetilde{\lambda}_j = K_0(\widetilde{\lambda}_i - \lambda_i)(\widetilde{\lambda}_j - \lambda_j) + K_j(\widetilde{\lambda}_i - \lambda_i) + K_i(\widetilde{\lambda}_j - \lambda_j) + \lambda_i\lambda_j. \tag{B.1.6}$$

Consider the guess $\widetilde{\lambda}_j = \frac{\lambda_j - 2K_j + K_0\lambda_j}{K_0 - 1}$. Continue by verifying that the guess satisfies Equation B.1.6. Expanding, the left-hand side can be written as

$$LHS = \widetilde{\lambda}_i\widetilde{\lambda}_j = \frac{2K_0\lambda_i\lambda_j - 2K_0K_j\lambda_i - 2K_0K_i\lambda_j + 4K_iK_j - 2K_j\lambda_i - 2K_i\lambda_j + \lambda_i\lambda_j K_0^2 + \lambda_i\lambda_j}{(K_0-1)^2}.$$

Noting that $\widetilde{\lambda}_j - \lambda_j = 2\frac{\lambda_j - K_j}{K_0 - 1}$, the RHS of Equation B.1.6 can be written as

$$RHS = \frac{1}{(K_0-1)^2}\{4K_0(\lambda_i - K_i)(\lambda_j - K_j) + 2K_j(\lambda_i - K_i)(K_0 - 1)$$
$$+ 2K_i(\lambda_j - K_j)(K_0 - 1) + \lambda_i\lambda_j(K_0 - 1)^2\} = LHS.$$

B.1.6 RGIV with observed explanatory variables

This section contains formal results for the RGIV model with observed explanatory variables.

Assumption B.1.1 (below) establishes conditions necessary for identification, consistency, and asymptotic normality. Assumption B.1.1 extends Assumption 2.3.1 to include control variables $\mathbf{x}_t$ that are independent of idiosyncratic shocks u_{it}. In the derivations that follow, it will be convenient to reparametrize the coefficient on the control variables as $\boldsymbol{\psi}_i = \boldsymbol{\beta}_i + \frac{\phi_i}{1-\phi_S}\boldsymbol{\beta}_S$. Then, store the re-parametrized coefficients in vector $\boldsymbol{\psi} = [\boldsymbol{\psi}_1', \dots \boldsymbol{\psi}_n']'$ in vector $\boldsymbol{\theta} = [\boldsymbol{\phi}', \boldsymbol{\psi}']'$ and its corresponding parameter space as $\boldsymbol{\Theta}$.

Assumption B.1.1. *(Baseline model)*

(i) ***Model***: *For $n \geq 3$ units, let fixed sizes $S_i \in (0,1)$ sum to 1. Outcome $\mathbf{r}_t = [r_{1t}, ..., r_{nt}]'$ responds to the size-aggregated outcome r_{St} according to spillover coefficient ϕ_i, coefficients $\boldsymbol{\beta}_i$ $(k \times 1)$, observed control variables $\mathbf{x}_t$ $(k \times 1)$, and unobserved shocks $\mathbf{u}_t = [u_{1t}, u_{2t}, ..., u_{nt}]'$, $r_{it} = \phi_i r_{St} + \boldsymbol{\beta}_i'\mathbf{x}_t + u_{it}$ for all $i = 1, ..., n$ where $\phi_S < 1$.*

(ii) ***Moments***: *For $\sigma_i^2 > 0$, shocks $\mathbf{u}_t$ are i.i.d. with moments $\mathbb{E}(\mathbf{u}_t) = 0$, $\mathbb{E}(\mathbf{u}_t\mathbf{u}_t') = \mathrm{diag}(\sigma_1^2, ..., \sigma_n^2)$, $\mathbb{E}(\|\mathbf{u}_t\|^4) < \infty$. Moreover, $u_{it} \perp\!\!\!\perp u_{jt}$ for $i \neq j$. $\mathbb{E}(\mathbf{x}_t\mathbf{x}_t') = \Sigma_{XX}$ for positive definite Σ_{XX} and $\mathbf{x}_t \perp\!\!\!\perp u_{it}$.*

(iii) ***Parameter space***: *For spillover coefficient $\boldsymbol{\phi} = [\phi_1, \dots, \phi_n]'$, store parameters in $\check{\boldsymbol{\theta}} = [\boldsymbol{\phi}', \boldsymbol{\beta}_1', \dots, \boldsymbol{\beta}_n']'$. Then, the true parameter $\check{\boldsymbol{\theta}}_0$ is in the interior of parameter space $\check{\boldsymbol{\Theta}}$. $\check{\boldsymbol{\Theta}}$ is compact and for any $\check{\widetilde{\boldsymbol{\theta}}} \in \check{\boldsymbol{\Theta}}$, $\widetilde{\phi}_S < 1$.*

Definition B.1.1 (below) extends the RGIV estimator to include observed explanatory variables. Step 1 estimates the total effect of control variables $\mathbf{x}_t$ on the outcome r_{it}. Step 2 writes the moment uncorrelatedness condition after subtracting the variation induced by the observed explanatory variables. See Section 2.5.1 for intuition.

Definition B.1.1. *Store data* $\mathbf{z}_t = [\mathbf{r}_t', \mathbf{x}_t']'$ *for outcome variable* $\mathbf{r}_t = [r_{1t}, \ldots, r_{nt}]'$. *Let* $u_i(\mathbf{z}_t, \boldsymbol{\phi}; \boldsymbol{\psi}) = (r_{it} - \boldsymbol{\psi}_i'\mathbf{x}_t) - \phi_i(r_{St} - \boldsymbol{\psi}_S' x_t)$ *for* $i = 1, ..., n$. *The moment function* $g^c(\mathbf{z}_t, \boldsymbol{\phi}; \boldsymbol{\psi})$ *is*

$$g^c(\mathbf{z}_t, \boldsymbol{\phi}; \boldsymbol{\psi}) = [u_1(\mathbf{z}_t, \boldsymbol{\phi}; \boldsymbol{\psi})u_2(\mathbf{z}_t, \boldsymbol{\phi}; \boldsymbol{\psi}), \ldots,\ u_1(\mathbf{z}_t, \boldsymbol{\phi}; \boldsymbol{\psi})u_n(\mathbf{z}_t, \boldsymbol{\phi}; \boldsymbol{\psi}),$$
$$u_2(\mathbf{z}_t, \boldsymbol{\phi}; \boldsymbol{\psi})u_3(\mathbf{z}_t, \boldsymbol{\phi}; \boldsymbol{\psi}), \ldots, u_{n-1}(\mathbf{z}_t, \boldsymbol{\phi}; \boldsymbol{\psi})u_n(\mathbf{z}_t, \boldsymbol{\phi}; \boldsymbol{\psi})]'.$$

Let the sample weight matrix $\widehat{W}$ *be defined as* $\widehat{W} = \mathrm{diag}\left(\frac{1}{\widehat{\sigma_1^2}\widehat{\sigma_2^2}}, \ldots,\ \frac{1}{\widehat{\sigma_1^2}\widehat{\sigma_n^2}}, \frac{1}{\widehat{\sigma_2^2}\widehat{\sigma_3^2}}, \ldots,\ \frac{1}{\widehat{\sigma_{n-1}^2}\widehat{\sigma_n^2}}\right)$. *where* $\widehat{\sigma}_i^2$ *is a consistent estimator of idiosyncratic shock variances* σ_i^2 *for* $i = 1, \ldots, n$. *Then the* ***robust granular instrumental variables estimator with observed explanatory variables*** *is characterized as the following two step estimator:*

1. *For each* $i = 1, \ldots, n$, $\widehat{\boldsymbol{\psi}}_i^{\textit{Step 1}} = (\frac{1}{T}\sum_{t=1}^T \mathbf{x}_t\mathbf{x}_t')^{-1}(\frac{1}{T}\sum_{t=1}^T \mathbf{x}_t r_{it})$.
2. *For* $\widehat{Q}_T(\boldsymbol{\phi}; \widehat{\boldsymbol{\psi}}^{\textit{Step 1}}) = \left[\frac{1}{T}\sum_{t=1}^T g^c(\mathbf{z}_t, \boldsymbol{\phi}; \widehat{\boldsymbol{\psi}}^{\textit{Step 1}})\right]'\widehat{W}\left[\frac{1}{T}\sum_{t=1}^T g^c(\mathbf{z}_t, \boldsymbol{\phi}; \widehat{\boldsymbol{\psi}}^{\textit{Step 1}})\right]$, $\widehat{\boldsymbol{\phi}}^{\textit{RGIV,c}} = \arg\min_{\boldsymbol{\phi}\in\boldsymbol{\Phi}} \widehat{Q}_T(\boldsymbol{\phi}; \widehat{\boldsymbol{\psi}}^{\textit{Step 1}})$.

For brevity, the formal results rely on simple extensions of the Newey and McFadden (1994) GMM identification, consistency, and asymptotic normality results as detailed in Appendix B.2.1 to allow for the first step estimation of a nuisance parameter. Note that the two-step estimation results of Newey and McFadden (1994) (Section 6) cannot be directly applied since there are potentially more moments than estimated parameters. Alternatively, the estimator described in Definition B.1.1 could be defined using a continuously updating GMM objective function in the second step. The procedure could be formalized by modifying the more general results of Pakes and Pollard (1989) to accommodate a first-step estimator.

Lemma B.1.1 (below) establishes that the first step estimator $\widehat{\psi}^{\text{Step 1}}$ is consistent and $O_p(1)$. Lemma B.1.2 establishes identification of spillover coefficients $\boldsymbol{\phi}$. Theorems B.1.1 and B.1.2 give consistency and asymptotic normality of $\widehat{\boldsymbol{\phi}}^{\text{RGIV,c}}$.

Lemma B.1.1. *Impose Assumption B.1.1. As $T \to \infty$, $\widehat{\boldsymbol{\psi}}_i^{\text{Step 1}} \xrightarrow{p} \boldsymbol{\psi}_i$ and $\sqrt{T}(\widehat{\boldsymbol{\psi}}_i^{\text{Step 1}} - \boldsymbol{\psi}_i) = O_p(1)$.*

Proof. Since $0 = \mathbb{E}[\mathbf{x}_t(r_{it} - \boldsymbol{\psi}_i'\mathbf{x}_t)] = \mathbb{E}[\mathbf{x}_t u_{it}] = 0$ and $\mathbb{E}[\mathbf{x}_t\mathbf{x}_t']$ is full rank, the OLS estimator $\widehat{\boldsymbol{\psi}}_i^{\text{Step 1}}$ is consistent. Assumption B.1.1(ii) (that $\mathbb{E}(\mathbf{u}_t\mathbf{u}_t') = \text{diag}(\sigma_1^2, \ldots, \sigma_n^2)$ and $\mathbb{E}(\mathbf{x}_t\mathbf{x}_t')$ full rank) implies $\widehat{\boldsymbol{\psi}}_i^{\text{Step 1}}$ is asymptotically normal so $\sqrt{T}(\widehat{\boldsymbol{\psi}}_i^{\text{Step 1}} - \boldsymbol{\psi}_i) = O_p(1)$. □

Lemma B.1.2 (Identification with observed explanatory variables)**.** *Impose Assumption B.1.1. For $g_0^c(\boldsymbol{\phi}) = \mathbb{E}[g^c(\mathbf{z}_t, \boldsymbol{\phi}; \boldsymbol{\psi}_0)]$, $g_0^c(\boldsymbol{\phi}_0) = 0$ for the true parameter $\boldsymbol{\phi}_0$ and $g_0^c(\widetilde{\boldsymbol{\phi}}) \neq 0$ for $\widetilde{\boldsymbol{\phi}} \in \boldsymbol{\Phi}$ such that $\widetilde{\boldsymbol{\phi}} \neq \boldsymbol{\phi}_0$.*

Proof. The result immediately follows after applying Lemma 2.3.1 to $\dot{r}_{it} = r_{it} - \boldsymbol{\psi}_i'\mathbf{x}_t$. □

Theorem B.1.1 (Consistency of RGIV with observed explanatory variables)**.** *Impose Assumption B.1.1. The RGIV estimator with observed explanatory variables is consistent $\widehat{\boldsymbol{\phi}}^{RGIV,c} \xrightarrow{p} \boldsymbol{\phi}_0$ for the true parameter $\boldsymbol{\phi}_0$ as $T \to \infty$.*

Proof. Verify the conditions of Theorem B.2.1. Condition (i) follows from Lemma B.1.2 and Lemma B.2.1. Condition (ii) follows from Assumption B.1.1(iii). Condition (iii) follows by inspection of function $g^c(\mathbf{z}_t, \boldsymbol{\phi}, \boldsymbol{\gamma})$. Condition (iv) follows from Assumption B.1.1(ii). □

Theorem B.1.2. *Impose Assumption B.1.1. The RGIV estimator with observed explanatory variables is asymptotically normal*

$$\sqrt{T}(\widehat{\boldsymbol{\phi}}^{RGIV,c} - \boldsymbol{\phi}_0) \xrightarrow{d} \mathcal{N}(0, (G'WG)^{-1}G'W\Sigma WG(G'WG)^{-1})$$

for RGIV population weight matrix $W = \text{diag}(\frac{1}{\sigma_1^2\sigma_2^2}, \ldots, \frac{1}{\sigma_1^2\sigma_n^2}, \frac{1}{\sigma_2^2\sigma_3^2}, \ldots, \frac{1}{\sigma_{n-1}^2\sigma_n^2})$, moment covariance matrix $\Sigma = \mathbb{E}[g^c(\mathbf{r}_t, \boldsymbol{\phi}_0; \boldsymbol{\psi}_0)g^c(\mathbf{r}_t, \boldsymbol{\phi}_0; \boldsymbol{\psi}_0)']$ and $G = \mathbb{E}[\nabla_\phi g^c(\mathbf{r}_t, \boldsymbol{\phi}_0; \boldsymbol{\psi}_0)]$ as $T \to \infty$.

Proof. Verify the conditions of Theorem B.2.2. The hypotheses for Theorem B.2.1 are satisfied. Lemma B.1.1 implies $\sqrt{T}(\widehat{\boldsymbol{\psi}}^{\text{Step 1}} - \boldsymbol{\psi}_0) = O_p(1)$. Next, compute the elements of

$\mathbb{E}[\nabla_{\boldsymbol{\psi}} g^c(\mathbf{z}_t, \boldsymbol{\phi}_0; \boldsymbol{\psi}_0)]$. Let $i, j, k \in \{1, \ldots, n\}$ and $i \neq j \neq k$. Since $\mathbf{x}_t \perp\!\!\!\perp u_{it}$,

$$
\begin{aligned}
&\mathbb{E}\{\frac{\partial}{\partial \boldsymbol{\psi}_j}[r_{it} - \boldsymbol{\psi}_i'\mathbf{x}_t - \phi_i(r_{St} - \boldsymbol{\psi}_S'\mathbf{x}_t)][r_{jt} - \boldsymbol{\psi}_j'\mathbf{x}_t - \phi_j(r_{St} - \boldsymbol{\psi}_S'\mathbf{x}_t)]\} \\
&\quad = \mathbb{E}\{(-\mathbf{x}_t + \phi_i S_i \mathbf{x}_t) u_{jt}(\mathbf{z}_t, \boldsymbol{\psi}_0, \boldsymbol{\phi}_0)\} + \mathbb{E}[\phi_j S_i \mathbf{x}_t u_{it}(\mathbf{z}_t, \boldsymbol{\psi}_0, \boldsymbol{\phi}_0)] = 0 \\
&\mathbb{E}\{\frac{\partial}{\partial \boldsymbol{\psi}_k}[r_{it} - \boldsymbol{\psi}_i'\mathbf{x}_t - \phi_i(r_{St} - \boldsymbol{\psi}_S'\mathbf{x}_t)][r_{jt} - \boldsymbol{\psi}_j'\mathbf{x}_t - \phi_j(r_{St} - \boldsymbol{\psi}_S'\mathbf{x}_t)]\} \\
&\quad = \mathbb{E}\{\phi_i S_k \mathbf{x}_t u_{jt}(\mathbf{z}_t, \boldsymbol{\psi}_0, \boldsymbol{\phi}_0)\} + \mathbb{E}[\phi_j S_k \mathbf{x}_t u_{it}(\mathbf{z}_t, \boldsymbol{\psi}_0, \boldsymbol{\phi}_0)] = 0.
\end{aligned}
$$

Hence, $\mathbb{E}[\nabla_{\boldsymbol{\psi}} g^c(\mathbf{z}_t, \boldsymbol{\phi}_0; \boldsymbol{\psi}_0)] = \mathbf{0}$. Condition (i) holds from Assumption B.1.1(iii). Condition (ii) holds by inspection of moment function $g(\mathbf{z}_t, \boldsymbol{\theta}; \boldsymbol{\gamma})$. Conditions (iii) and (iv) hold from Assumption B.1.1(ii). Condition (v) holds from applying the proof of Theorem 2.3.2. $\square$

B.2 Online appendix

B.2.1 Observed explanatory variables: Second-step GMM estimation with a consistent first-step estimator of a nuisance parameter

For completeness, the below identification, consistency, and asymptotic normality results are straightforward adaptations of those found in Newey and McFadden (1994) (abbreviated as NM). These results give conditions for identification and inference for two-step estimators, given a consistent first-step estimator of a nuisance parameter. Following the notation of NM, the first step nuisance parameter estimator is consistent $\widehat{\boldsymbol{\gamma}} \xrightarrow{p} \boldsymbol{\gamma}_0$ where $\sqrt{n}(\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0) = O_p(1)$. In the second step, $\boldsymbol{\theta}$ is the parameter of interest and is estimated using GMM with moment function $g(\mathbf{z}, \boldsymbol{\theta}; \boldsymbol{\gamma})$ and weight matrix estimator $\widehat{W} \xrightarrow{p} W$, where $\widehat{\boldsymbol{\gamma}}$ is an estimator for $\boldsymbol{\gamma}$. Moreover, the condition $\mathbb{E}[\nabla_{\boldsymbol{\gamma}} g(\mathbf{z}, \boldsymbol{\theta}_0; \boldsymbol{\gamma}_0)] = 0$ (for true parameter values $\boldsymbol{\gamma}_0$ and $\boldsymbol{\theta}_0$) ensures that estimation uncertainty for the first-step estimator doesn't enter the expression for the asymptotic variance of the second-step estimator, as is the case for the RGIV estimator with observed explanatory variables.

The below identification lemma (Lemma B.2.1) extends NM Lemma 2.3, the consistency theorem (Theorem B.2.1) extends NM Lemma 2.6, and the asymptotic normality theorem (Theorem B.2.2) extends NM Theorem 3.4. The proofs for these extensions are, for the most part, identical to the original NM results. Notably, care is taken to ensure uniform convergence of the GMM objective function in Theorem B.2.1 and the expected Jacobian matrices in Theorem B.2.2.

These results can be applied to estimators where the second step GMM estimator is over-identified, as is the case for RGIV when $n \geq 4$. In contrast, an application of the results featured in Newey and McFadden (1994) Chapter 6 requires the number of moment conditions to equal the number of parameters of interest.

Lemma B.2.1 (Identification, two-step). *If W is positive semi-definite and, for $g_0(\boldsymbol{\theta}) = \mathbb{E}[g(\mathbf{z}, \boldsymbol{\theta}; \boldsymbol{\gamma}_0)]$, $g_0(\boldsymbol{\theta}_0) = 0$, and $W g_0(\boldsymbol{\theta}) \neq 0$ for $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$ then $Q_0(\boldsymbol{\theta}) = -g_0(\boldsymbol{\theta})' W g_0(\boldsymbol{\theta})$ has a unique maximum at $\boldsymbol{\theta}_0$.*

Proof. Apply an identical argument as the one used for the proof of NM Lemma 2.3.

Let R be such that $R'R = W$. If $\theta \neq \boldsymbol{\theta}_0$, then $0 \neq W g_0(\boldsymbol{\theta}) = R'Rg_0(\boldsymbol{\theta})$ implies $Rg_0(\boldsymbol{\theta}) \neq 0$ and hence $Q_0(\boldsymbol{\theta}) = -[Rg_0(\boldsymbol{\theta})]'[Rg_0(\boldsymbol{\theta})] < Q_0(\boldsymbol{\theta}_0) = 0$ for $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$. □

Theorem B.2.1 (Consistency, two-step). *Suppose that $\mathbf{z}_i, (i = 1, 2, \dots)$, are i.i.d., $\widehat{W} \xrightarrow{p} W$, $\widehat{\boldsymbol{\gamma}} \xrightarrow{p} \boldsymbol{\gamma}_0$, and (i) W is positive semidefinite and $W\mathbb{E}[g(\mathbf{z}, \boldsymbol{\theta}; \boldsymbol{\gamma}_0)] = 0$ only if $\boldsymbol{\theta} = \theta_0$; (ii) $\boldsymbol{\theta}_0 \times \boldsymbol{\gamma}_0 \in \boldsymbol{\Theta} \times \boldsymbol{\Gamma}$ compact; (iii) $g(\mathbf{z}, \boldsymbol{\theta}; \boldsymbol{\gamma})$ is continuous at each $\boldsymbol{\theta} \times \boldsymbol{\gamma} \in \boldsymbol{\Theta} \times \boldsymbol{\Gamma}$ with probability one; (iv) $\mathbb{E}[\sup_{\boldsymbol{\theta}, \boldsymbol{\gamma}} \|g(\mathbf{z}, \boldsymbol{\theta}; \boldsymbol{\gamma})\|] < \infty$. Then $\widehat{\boldsymbol{\theta}} \xrightarrow{p} \boldsymbol{\theta}_0$.*

Proof. Proceed by verifying the hypotheses of Theorem 2.1 of Newey and McFadden (1994)(abbreviated as NM2.1):

- NM2.1(i): Follows from Lemma B.2.1 and Condition (i).
- NM2.1(ii): Follows from Condition (ii).
- NM2.1(iii): Newey and McFadden Lemma 2.4, Condition (iii), and Condition (iv) imply that $\sup_{\boldsymbol{\theta}} \|\frac{1}{n}\sum_{i=1}^n g(\mathbf{z}_i, \boldsymbol{\theta}; \widehat{\boldsymbol{\gamma}}) - \mathbb{E}[g(\mathbf{z}, \boldsymbol{\theta}; \widehat{\boldsymbol{\gamma}})]\| \xrightarrow{p} 0$ and $\mathbb{E}[g(\mathbf{z}, \boldsymbol{\theta}; \boldsymbol{\gamma}_0)]$ is continuous in θ. Moreover, $\widehat{\boldsymbol{\gamma}} \xrightarrow{p} \boldsymbol{\gamma}_0$ implies $\sup_{\boldsymbol{\theta}} \|\frac{1}{n}\sum_{i=1}^n g(\mathbf{z}_i, \boldsymbol{\theta}; \widehat{\boldsymbol{\gamma}}) - \mathbb{E}[g(\mathbf{z}, \boldsymbol{\theta}; \boldsymbol{\gamma}_0)]\| \xrightarrow{p} 0$. Hence, 2.1(iii) holds because $Q_0(\theta) = -\mathbb{E}[g(\mathbf{z}, \boldsymbol{\theta}; \boldsymbol{\gamma}_0)]' W \mathbb{E}[g(\mathbf{z}, \boldsymbol{\theta}; \boldsymbol{\gamma}_0)]$ is continuous.
- NM2.1(iv): For uniform convergence of the objective function $\widehat{Q}_n(\boldsymbol{\theta})$ to $Q_0(\boldsymbol{\theta})$, follow similar computations to Newey and McFadden (1994) Theorem 2.6:

$$
\begin{aligned}
|\widehat{Q}_n(\boldsymbol{\theta}) - Q_0(\boldsymbol{\theta})| \leq \|\frac{1}{n}\sum_{i=1}^n g(\mathbf{z}_i, \boldsymbol{\theta}; \widehat{\boldsymbol{\gamma}}) - \mathbb{E}[g(\mathbf{z}, \boldsymbol{\theta}; \boldsymbol{\gamma}_0)]\|^2 \|\widehat{W}\| + \\
2\|\frac{1}{n}\sum_{i=1}^n g(\mathbf{z}_i, \boldsymbol{\theta}; \widehat{\boldsymbol{\gamma}}) - \mathbb{E}[g(\mathbf{z}, \boldsymbol{\theta}; \boldsymbol{\gamma}_0)]\| \|\frac{1}{n}\sum_{n=1}^N g(\mathbf{z}_t, \boldsymbol{\theta}; \widehat{\boldsymbol{\gamma}}) - \mathbb{E}[g(\mathbf{z}, \boldsymbol{\theta}; \boldsymbol{\gamma}_0)]\| \|\widehat{W}\| + \\
\|\mathbb{E}[g(\mathbf{z}, \boldsymbol{\theta}; \boldsymbol{\gamma}_0)]\|^2 \| |\widehat{W} - W\|^2.
\end{aligned}
$$

$\square$

Theorem B.2.2 (NM (1994) Theorem 3.4, modified)**.** *Suppose the hypotheses of Theorem B.2.1 are satisfied. Also assume* $\sqrt{n}(\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0) = O_p(1)$, $\mathbb{E}[\nabla_{\boldsymbol{\gamma}} g(\mathbf{z}, \boldsymbol{\theta}_{\mathbf{0}}; \boldsymbol{\gamma}_0)] = \mathbf{0}$, *(i)* $\boldsymbol{\theta}_0 \times \boldsymbol{\gamma}_0 \in$ interior$(\boldsymbol{\Theta} \times \boldsymbol{\Gamma})$*; (ii)* $g(\mathbf{z}, \boldsymbol{\theta}; \boldsymbol{\gamma})$ *is continuously differentiable in a neighborhood* $\mathcal{N}$ *of* $\boldsymbol{\theta}_0 \times \boldsymbol{\gamma}_0$ *with probability approaching 1; (iii)* $\mathbb{E}[g(\mathbf{z}, \boldsymbol{\theta}_0; \boldsymbol{\gamma}_0)] = 0$ *and* $\mathbb{E}[\|g(\mathbf{z}, \boldsymbol{\theta}_0; \boldsymbol{\gamma}_0)\|^2] < \infty$*, (iv)* $\mathbb{E}[\sup_{\boldsymbol{\theta} \times \boldsymbol{\gamma}} \|\nabla_{\boldsymbol{\theta} \times \boldsymbol{\gamma}} g(\mathbf{z}, \boldsymbol{\theta}, \boldsymbol{\gamma})\|] < \infty$*; (v)* $G'WG$ *is non-singular for* $G = \mathbb{E}[\nabla_{\boldsymbol{\theta}} g(\mathbf{z}, \boldsymbol{\theta}_0; \boldsymbol{\gamma}_0)]$*. Then, for* $\Sigma = \mathbb{E}[g(\mathbf{z}, \boldsymbol{\theta}_0; \boldsymbol{\gamma}_0) g(\mathbf{z}, \boldsymbol{\theta}_0; \boldsymbol{\gamma}_0)']$*,*

$$\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} \mathcal{N}(0, (G'WG)^{-1} G'W\Sigma WG(G'WG)^{-1}).$$

Proof. Conditions (i),(ii), and (iii) imply that the first order condition is satisfied with probability approaching one: $2\widehat{G}_n(\widehat{\boldsymbol{\theta}})'\widehat{W}\widehat{g}_n(\widehat{\boldsymbol{\theta}}) = 0$ for $\widehat{G}_n(\widehat{\boldsymbol{\theta}}) = \nabla_{\boldsymbol{\theta}} \frac{1}{n}\sum_{i=1}^n g(\mathbf{z}_i, \widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\gamma}})$. Expanding the FOC about $\boldsymbol{\theta}_0$ and $\boldsymbol{\gamma}_0$ and rearranging gives

$$\begin{aligned}
\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) &= -(\widehat{G}_n(\widehat{\boldsymbol{\theta}})'\widehat{W}\widehat{G}_n(\overline{\boldsymbol{\theta}}))^{-1}\widehat{G}_n(\widehat{\boldsymbol{\theta}})\widehat{W}\frac{\sqrt{n}}{n}\sum_{i=1}^n g(\mathbf{z}_i, \boldsymbol{\theta}_0; \widehat{\boldsymbol{\gamma}}) \\
&= -(\widehat{G}_n(\widehat{\boldsymbol{\theta}})'\widehat{W}\widehat{G}_n(\overline{\boldsymbol{\theta}}))^{-1}\widehat{G}_n(\widehat{\boldsymbol{\theta}})\widehat{W}\frac{\sqrt{n}}{n}\sum_{i=1}^n g(\mathbf{z}_i, \boldsymbol{\theta}_0; \boldsymbol{\gamma}_0) \\
&\quad - (\widehat{G}_n(\widehat{\boldsymbol{\theta}})'\widehat{W}\widehat{G}_n(\overline{\boldsymbol{\theta}}))^{-1}\widehat{G}_n(\widehat{\boldsymbol{\theta}})\widehat{W}\frac{1}{n}\sum_{i=1}^n \nabla_{\boldsymbol{\gamma}} g(\mathbf{z}_i, \boldsymbol{\theta}_0; \overline{\boldsymbol{\gamma}})\sqrt{n}(\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0) \\
&\xrightarrow{p} \mathcal{N}(0, (G'WG)^{-1}G'W\Sigma WG(G'WG)^{-1})
\end{aligned}$$

where $\overline{\boldsymbol{\theta}}$ and $\overline{\boldsymbol{\gamma}}$ are intermediate values. In the above display, $\widehat{\boldsymbol{\gamma}} \xrightarrow{p} \boldsymbol{\gamma}_0$, $\widehat{\boldsymbol{\theta}} \xrightarrow{p} \boldsymbol{\theta}_0$, and condition (iv) imply $\widehat{G}_n(\widehat{\theta}) \xrightarrow{p} G$, $\widehat{G}_n(\overline{\theta}) \xrightarrow{p} G$, and $\frac{1}{n}\sum_{i=1}^n \nabla_{\boldsymbol{\gamma}} g(\mathbf{z}_i, \boldsymbol{\theta}_0; \overline{\boldsymbol{\gamma}}) \xrightarrow{p} \mathbf{0}$. The final line follows from the Slutsky theorem and an i.i.d. central limit theorem.

$\square$

B.2.2 RGIV when the shock factor is determined by known observables

In this section, I show that differencing RGIV moments can accommodate the case where the shock factor structure is entirely determined by known observables. I also show that the identification reduces to an overdetermined system of nonlinear equation, which admits a unique solution (with the parameter space restriction $\phi_S < 1$) outside of knife-edge cases.

B.2.2.1 Setup

Take Assumption 2.3.1 and augment with latent factors $\boldsymbol{\eta}_t$ $(m \times 1)$ where $\mathbb{E}[\boldsymbol{\eta}_t \boldsymbol{\eta}_t'] = \mathbf{I}$ (and finite fourth moments) and loadings $\boldsymbol{\lambda}_i$ $(m \times 1)$, which are entirely determined by observables $\mathbf{x}_i$ $(m \times 1)$ through $\boldsymbol{\Pi}$ $(m \times m)$

$$r_{it} = \phi_i r_{St} + \underbrace{\boldsymbol{\lambda}_i' \boldsymbol{\eta}_t + u_{it}}_{v_{it}}, \quad \boldsymbol{\lambda}_i = \boldsymbol{\Pi} \mathbf{x}_i.$$

Also suppose there are more GMM moment conditions (less the square of the number of observables) than there are unknowns $(n(n-1)/2 - m^2 \geq n)$. $\boldsymbol{\Pi}$ $(m \times m)$ maps the unit-specific observables to the loadings and is unknown to the researcher. In matrix form, for $\mathbf{X} = \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \dots & \mathbf{x}_n \end{bmatrix}$ and $\boldsymbol{\Lambda} = \begin{bmatrix} \boldsymbol{\lambda}_1 & \boldsymbol{\lambda}_2 & \dots & \boldsymbol{\lambda}_n \end{bmatrix}$,

$$\boldsymbol{\Lambda} = \boldsymbol{\Pi}\mathbf{X}, \ \mathbf{r}_t = \boldsymbol{\phi} r_{St} + \underbrace{\boldsymbol{\Lambda}' \boldsymbol{\eta}_t + \mathbf{u}_t}_{\mathbf{v}_t}, \text{ and } \operatorname{rank}(\mathbf{X}) = m.$$

The shock covariance matrix is $\mathbb{E}[\mathbf{v}_t \mathbf{v}_t'] = \mathbf{X}' \boldsymbol{\Pi}' \boldsymbol{\Pi} \mathbf{X} + \mathbf{D}$ for diagonal matrix $\mathbf{D}$ with diagonal entries $\sigma_1^2, \sigma_2^2, \dots, \sigma_n^2$. Vectorizing and selecting the off-diagonal elements of the covariance matrix with selection matrix $\mathbf{S}$ $(n(n-1)/2 \times n^2)$,

$$\mathbf{S}\text{vec}\left(\mathbb{E}[\mathbf{v}_t \mathbf{v}_t']\right) = \mathbf{S}\left(\mathbf{X}' \otimes \mathbf{X}'\right) \text{vec}\left(\boldsymbol{\Pi}' \boldsymbol{\Pi}\right).$$

Premultiplying the above display by the annihilator matrix (for arbitrary matrix W, defined as $\mathbf{W}$ $M_{\mathbf{W}} \equiv \mathbf{I} - \mathbf{W}(\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}'$),

$$\mathbf{M}_{\mathbf{S}(\mathbf{X}'\otimes\mathbf{X}')}\mathbf{S}\text{vec}\left(\mathbb{E}[\mathbf{v}_t\mathbf{v}_t']\right) = \mathbf{0}. \tag{B.2.1}$$

Hence, a linear combination of the off-diagonal elements of $\mathbb{E}[\mathbf{v}_t\mathbf{v}_t']$ equals zero.

B.2.2.2 Identification

Equation B.2.1 sets a linear combination of RGIV moments to zero. Below, I show that the spillover coefficients are generally identified outside of a set of knife-edge DGPs. Equation B.2.1 implies a system of (univariate) quadratic equations in the errors of a particular spillover coefficient. For there to exist a solution other than the true solution, these quadratic equations must have identical zeros—which isn't true in general.

For true spillover coefficients ϕ_i and proposed solutions $\widetilde{\phi}_i$, let $\check{\phi}_i \equiv \phi_i - \widetilde{\phi}_i$ give the error in spillover coefficient ϕ_i. The moment condition corresponding to the idiosyncratic shocks of units i and j can be written as

$$\mathbb{E}\left[(r_{it} - \widetilde{\phi}_i r_{St})(r_{jt} - \widetilde{\phi}_j r_{St})\right] = \check{\phi}_i\check{\phi}_j k + \check{\phi}_i k_j + \check{\phi}_j k_i$$

for $k \equiv \mathbb{E}[r_{St}^2]$ and $k_i \equiv \mathbb{E}[r_{St}(u_{it} + \boldsymbol{\lambda}_i'\boldsymbol{\eta}_t)]$. For $i \neq j$, the above display's moments can be stored in an $n(n-1)/2 \times 1$ vector $\widetilde{\boldsymbol{g}}$ matching the ordering of $\mathbf{S}\text{vec}\left(\mathbb{E}[\mathbf{v}_t\mathbf{v}_t']\right)$ in Equation B.2.1. Hence, identification of the moments characterized by Equation B.2.1 reduces to studying the roots of $\mathbf{M}_{\mathbf{S}(\mathbf{X}'\otimes\mathbf{X}')}\widetilde{\mathbf{g}} = 0$.

Since $\text{rank}(\mathbf{M}_{\mathbf{S}(\mathbf{X}'\otimes\mathbf{X}')}) = n(n-1)/2 - m^2$, $\mathbf{M}_{\mathbf{S}(\mathbf{X}'\otimes\mathbf{X}')}$ can be expressed in reduced row echelon form (after elementary row operations)

$$\begin{bmatrix} \mathbf{I}_{n(n-1)/2-m^2} & -\mathbf{A} \\ \mathbf{0}_{m^2\times(n(n-1)/2-m^2)} & \mathbf{0}_{m^2\times m^2} \end{bmatrix}.$$

Letting $\widetilde{\boldsymbol{g}}^I$ be the first $n(n-1)/2-m^2$ moments and $\widetilde{\boldsymbol{g}}^{II}$ be the last m^2 moments, $\mathbf{M}_{\mathbf{S}(\mathbf{X}'\otimes\mathbf{X}')}\widetilde{\mathbf{g}} = 0$ can equivalently be expressed as

$$\widetilde{\mathbf{g}}^I = \mathbf{A}\widetilde{\mathbf{g}}^{II}. \tag{B.2.2}$$

The moments $\widetilde{\mathbf{g}}^I$ equal some (known) linear combination of $\widetilde{\mathbf{g}}^{II}$ determined by matrix $\mathbf{A}$.

To characterize the solutions, fix $\widetilde{\mathbf{g}}^{II}$ and let $\mathbf{a} \equiv \mathbf{A}\widetilde{\mathbf{g}}^{II}$ where a_{ij} is the entry of vector $\mathbf{a}$ that corresponds to the moment condition for shocks $i \neq j$. Then the row of Equation B.2.2 corresponding to the shocks of units i and j is $\check{\phi}_i(k_j + \check{\phi}_j k) + \check{\phi}_j k_i = a_{ij}$.

After rearranging and setting unit 1 as the reference unit, $\check{\phi}_i = \frac{a_{i1}-\check{\phi}_1 k_i}{k_1+\check{\phi}_1 k}$ and $\check{\phi}_j = \frac{a_{j1}-\check{\phi}_1 k_j}{k_1+\check{\phi}_1 k}$. Substituting $\check{\phi}_i$ and $\check{\phi}_j$,

$$a_{ij} = \check{\phi}_i(k_j + \check{\phi}_j k) + \check{\phi}_j k_i = \frac{a_{i1}-\check{\phi}_1 k_i}{k_1+\check{\phi}_1 k}k_j + \frac{a_{j1}-\check{\phi}_1 k_j}{k_1+\check{\phi}_1 k}k_i + \frac{(a_{i1}-\check{\phi}_1 k_i)(a_{j1}-\check{\phi}_1 k_j)}{(k_1+\check{\phi}_1 k)^2}k.$$

Rearranging yields a quadratic equation in $\check{\phi}_1$

$$0 = \alpha_1^{(ij)}\check{\phi}_1^2 + \alpha_2^{(ij)}\check{\phi}_1 + \alpha_3^{(ij)} \tag{B.2.3}$$

where $\alpha_1^{(ij)} \equiv a_{ij}k^2 + k_i k_j k$, $\alpha_2^{(ij)} \equiv 2a_{ij}k_1 k + 2k_i k_j k_1$, and $\alpha_3^{(ij)} \equiv a_{ij}k_1^2 - (a_{i1}k_j + a_{j1}k_i)k_1 - a_{i1}a_{j1}k$. Hence, the above display gives a system of quadratic equations (in $\check{\phi}_1$) defined by the pairs $i \neq j$ contained in moments $\widetilde{\mathbf{g}}^I$.

If $\widetilde{\mathbf{g}}^{II} = \mathbf{0}$, then $\mathbf{a} = \mathbf{0}$, $\alpha_1^{(ij)} = k_i k_j k$, $\alpha_2^{(ij)} = 2k_i k_j k$, and $\alpha_3^{(ij)} = 0$. Here, $0 = \check{\phi}_1(k\check{\phi}_1 + 2k_1)$. There are two solutions, which both do not depend on the index i, j: (1) the true solution $\check{\phi}_1 = 0$ and (2) a false solution $\check{\phi}_1 = -2\frac{k_1}{k}$ (which can be ruled out from the condition $\phi_S < 1$ as is done in Lemma 2.3.1).

Now instead suppose $\widetilde{\mathbf{g}}^{II} \neq \mathbf{0}$. From the quadratic formula, the roots of Equation B.2.3 (dividing by $\alpha_1^{(ij)}$) are determined by $\frac{\alpha_2^{(ij)}}{\alpha_1^{(ij)}}$ and $\frac{\alpha_3^{(ij)}}{\alpha_1^{(ij)}}$. $\frac{\alpha_3^{(ij)}}{\alpha_1^{(ij)}}$ in particular depends on the units

i, j through terms k_i and k_j. Hence, besides knife-edge cases, it is not generally true that the roots for Equation B.2.3 coincide when $\tilde{\mathbf{g}}^{II} \neq \mathbf{0}$.

B.2.3 Application: Robustness

Table B.2.1 reports coefficients and standard errors under alternative winsorization schemes. Columns 1-3 show that the results are qualitatively similar to the preferred specification after winsorizing at 5%, 0.5%, and 10% respectively. The fourth column, however, shows that estimation without winsorization is unreliable. These results show that the qualitative features of the main text's preferred specification aren't specific to the choice of 1% winsorization and that the linear form of the model described in Assumption 2.3.1(i) is sensitive to extreme outliers.

	5%	0.5%	10%	0%
RGIV results:				
ϕ_S	0.57	0.54	0.58	-16.57
	(0.06)	(0.08)	(0.05)	(10275.21)
$\phi_{\text{IRL,PRT,ESP}}$	0.82	0.82	0.83	0.98
	(0.05)	(0.06)	(0.05)	(2.38)
$\phi_{\text{GRC, ITA}}$	0.49	0.45	0.5	-28.08
	(0.12)	(0.16)	(0.11)	(17084.06)
ϕ_{Core}	0.42	0.38	0.42	0.47
	(0.03)	(0.03)	(0.03)	(2.48)
ϕ_{SVN}	0.33	0.33	0.34	0.34
	(0.05)	(0.06)	(0.05)	(0.39)
Tests (p-values):				
Specification	0.817	0.872	0.679	0.616
Homogeneity	0	0	0	0.275

Table B.2.1: Estimation results. Coefficient estimates for $\phi_S, \phi_{\text{IRL,PRT,ESP}}, \phi_{\text{GRC, ITA}}, \phi_{\text{Core}}$ and ϕ_{SVN} are listed above standard errors, which are provided in parentheses. p values are provided in the bottom section of the table for the specification test and parameter homogeneity tests. Columns 1-4 give estimation results for winsorization at the 2.5 and 97.5 percentiles, 0.25 and 99.75 percentiles, 5 and 95 percentiles, and no winsorization respectively.

Table B.2.2 summarizes RGIV results under alternative blocking schemes. Starting with the preferred specification, the first column shifts Ireland from the Portugal/Spain periphery block to the Greece/Italy periphery block (splitting the periphery countries into an "Iberia" block and "Other" block). The specification test's p-value of 0.95 suggests that "Iberia" specification's moments are consistent with the data. Moreover, the spillover coefficient for the Portugal/Spain block is 0.89 compared to 0.83 for the Portugal/Spain/Ireland block in the preferred specification, suggesting that Ireland's spillover coefficient is lower than the size-weighted combination of Portugal/Spain's. Columns 2 and 3 show that spillover coefficients are nearly unchanged after moving Slovenia to the core and periphery block respectively (no specification test is reported since the model is just-identified). The specification in Column 4 separates France from the core countries and moves it into its own group. The specification test is rejected, giving indirect evidence that French shocks are correlated with those of other Euro area core countries.

	Iberia	Core SVN	Periphery SVN	Separate FRA
RGIV results:				
ϕ_S	0.51	0.54	0.54	0.52
	(0.11)	(0.08)	(0.08)	(0.09)
ϕ_I	0.89	0.82	0.83	0.83
	(0.06)	(0.06)	(0.06)	(0.06)
ϕ_{II}	0.39	0.45	0.44	0.42
	(0.2)	(0.16)	(0.16)	(0.18)
ϕ_{III}	0.4	0.39	0.39	0.4
	(0.04)	(0.03)	(0.03)	(0.03)
ϕ_{IV}	0.33			0.38
	(0.06)			(0.04)
Tests (p-values):				
Specification	0.95	–	–	0.001
Homogeneity	<.001	<.001	<.001	<.001

Table B.2.2: Estimation results for alternative groupings. Coefficient estimates for $\phi_S, \phi_{\text{I}}, \phi_{\text{II}}, \phi_{\text{III}}$ and ϕ_{IV} are listed above standard errors, which are provided in parentheses. p values are provided in the bottom section of the table for the specification test and parameter homogeneity tests. Groups: "Iberia" (I: PRT, ESP; II: GRC, ITA, IRL; III: AUT, BEL, FRA, FIN, NLD; IV: SVN), "Core SVN" (I: PRT, ESP; II: GRC, ITA, IRL; III: AUT, BEL, FRA, FIN, NLD, SVN), "Periphery SVN" (I: PRT, ESP, IRL; II: GRC, ITA, SVN; III: AUT, BEL, FRA, FIN, NLD), "Separate FRA" (I: PRT, ESP, IRL; II: GRC, ITA; III: AUT, BEL, FIN, NLD, SVN; IV: FRA).

	Andrews rule	No Fama-French	0-factor GIV	2-factor GIV
RGIV results:				
ϕ_S	0.54	0.54		
	(0.07)	(0.08)		
$\phi_{\text{PRT,ESP,IRL}}$	0.83	0.83		
	(0.05)	(0.06)		
$\phi_{\text{GRC, ITA}}$	0.44	0.45		
	(0.14)	(0.17)		
ϕ_{Core}	0.39	0.4		
	(0.03)	(0.03)		
ϕ_{SVN}	0.33	0.34		
	(0.05)	(0.06)		
ϕ^{GIV}			0.42	0.31
			(0.03)	(0.06)
Tests (p-values):				
Specification	0.93	0.952		
Homogeneity	<.001	<.001		
First-stage F-statistic			411	112

Table B.2.3: Estimation results. Coefficient estimates for $\phi_S, \phi_{\text{IRL,PRT,ESP}}, \phi_{\text{GRC, ITA}}, \phi_{\text{Core}}$ and ϕ_{SVN} are listed above standard errors, which are provided in parentheses. p values are provided in the bottom section of the table for the specification test and parameter homogeneity tests.

Appendix C

Appendix to Chapter 3

C.1 Further theoretical results

C.1.1 Least favorable misspecification

Figure C.1.1 plots some numerical examples of the least favorable MA polynomial $\alpha^\dagger(L;h,M) = \sum_{\ell=1}^{\infty} \alpha^\dagger_{\ell,h,M} L^\ell$ discussed in Section 3.4.2. We focus here on a univariate local-to-AR(1) model $y_t = \rho y_{t-1} + [1 + T^{-1/2}\alpha(L)]\varepsilon_t$, though unreported numerical experiments suggest that the qualitative features mentioned below also apply to multivariate models. Recall that the least favorable MA coefficients depend on the horizon h of interest, while M only influences the overall scale of the coefficients, and not their shape as a function of ℓ. The figure shows that the shape of the coefficients either takes the form of a hump or of a single zig-zag pattern, with the largest absolute value of the coefficients generally occurring at $\ell = h$. Notice that we can flip the signs of all coefficients without changing the absolute value of the bias.

C.1.2 More efficient bias-aware confidence interval

Generalizing the bias-aware VAR confidence interval in Section 3.4.3, consider a bias-aware confidence interval that is centered at the model averaging estimator $\hat{\theta}_h(\omega) = \omega\hat{\beta}_h + (1-\omega)\hat{\delta}_h$

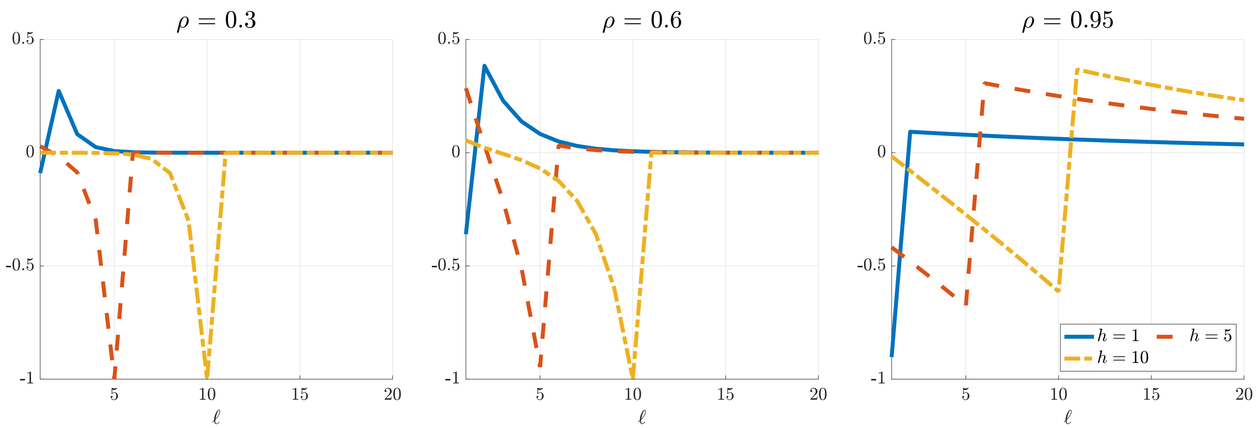

Figure C.1.1: Least favorable $\alpha^{\dagger}(L; h)$ for horizons $h \in \{1, 5, 10\}$ for local-to-AR(1) models with different persistence parameters ρ (left, middle, and right panel).

from Corollary 3.4.2:

$$\mathrm{CI}_B(\hat{\theta}_h(\omega); M) \equiv \left[\hat{\theta}_h(\omega) \pm \mathrm{cv}_{1-a}\left(\frac{(1-\omega)M\tau}{\sqrt{1+\omega^2\tau^2}}\right)\sqrt{(1+\omega^2\tau^2)\,\mathrm{aVar}(\hat{\delta}_h)/T}\right],$$

where $\tau \equiv \sqrt{\mathrm{aVar}(\hat{\beta}_h)/\,\mathrm{aVar}(\hat{\delta}_h) - 1}$. This interval equals the conventional LP interval when $\omega = 1$ and the bias-aware VAR interval when $\omega = 0$.

Corollary C.1.1. *Impose Assumption 3.2.1, $\zeta = 1/2$, and $\mathrm{aVar}(\hat{\delta}_h) > 0$. Then, for any $\omega \in [0, 1]$,*

$$\inf_{\alpha(L)\colon \|\alpha(L)\| \le M} \lim_{T\to\infty} P(\theta_{h,T} \in \mathrm{CI}_B(\hat{\theta}_h(\omega); M)) = 1 - a.$$

Proof. The result follows from Propositions 3.3.1, 3.3.2 and 3.4.1, Corollary C.1.2, and the same calculations as in the proof of Corollary 3.4.2. □

Even if we choose the weight ω to minimize confidence interval length, the resulting bias-aware interval tends to be nearly as long as the LP interval. The length-optimal weight $\omega = \omega^*$ is given by

$$\omega^* \equiv \operatorname*{argmin}_{\omega\in[0,1]} \mathrm{cv}_{1-a}\left(\frac{(1-\omega)M\tau}{\sqrt{1+\omega^2\tau^2}}\right)\sqrt{1+\omega^2\tau^2}.$$

Figure C.1.2 shows this optimal weight as a function of M and the relative asymptotic standard deviation of the VAR and LP estimators, while Figure C.1.3 shows the length of the resulting optimal bias-aware confidence interval relative to the length of the conventional LP interval. We see that, for $M \geq 2$, there is little gain from reporting the optimal bias-aware interval rather than the LP interval, regardless of the relative precision of VAR and LP. An additional observation is that, for $M \geq 1.5$, the length-optimal ω^* is numerically close to the MSE-optimal weight $M^2/(1+M^2)$ derived in Corollary 3.4.2.

C.1.3 Covariance structure of LP and VAR estimators

The following result provides the asymptotic variance-covariance matrix of the LP and VAR estimators in the general multi-dimensional set-up of Section 3.4.4. Define $\Psi_{i^*,j^*,h}$ as in Proposition 3.3.2, but making the dependence on (i^*, j^*) explicit in the notation.

Corollary C.1.2. *Impose Assumption 3.2.1, with part (iii)) holding for all shock indices $j_1^*, \dots, j_k^*$. Then for any $a, b \in \{1, \dots, k\}$,*

$$
\begin{aligned}
\mathrm{aCov}(\hat{\beta}_{i_a^*,j_a^*,h_a}, \hat{\beta}_{i_b^*,j_b^*,h_b}) &= \mathbb{1}(j_a^* = j_b^*)\sigma_{j_a^*}^{-2}\left(\psi_{a,b} + \sum_{\ell=1}^{\min\{h_a,h_b\}} e'_{i_a^*,n} A^{h_a-\ell}\Sigma(A')^{h_b-\ell} e_{i_b^*,n}\right), \\
\mathrm{aCov}(\hat{\delta}_{i_a^*,j_a^*,h_a}, \hat{\delta}_{i_b^*,j_b^*,h_b}) &= \mathbb{1}(j_a^* = j_b^*)\sigma_{j_a^*}^{-2}\psi_{a,b} + \mathrm{tr}\left(\Psi_{i_a^*,j_a^*,h_a}\Sigma\Psi'_{i_b^*,j_b^*,h_b}S^{-1}\right), \\
\mathrm{aCov}(\hat{\beta}_{i_a^*,j_a^*,h_a}, \hat{\delta}_{i_b^*,j_b^*,h_b}) &= \mathrm{aCov}(\hat{\delta}_{i_a^*,j_a^*,h_a}, \hat{\delta}_{i_b^*,j_b^*,h_b}),
\end{aligned}
$$

where

$$
\psi_{a,b} \equiv e'_{i_a^*,n} A^{h_a} \overline{H}_{j_a^*} \overline{D}_{j_a^*} \overline{H}'_{j_a^*} (A')^{h_b} e_{i_b^*,n},
$$

and the "aCov" notation refers to elements of the asymptotic variance-covariance matrix in Equation (C.3.2) in Appendix C.3.4. In particular, $\mathrm{aCov}(\hat{\beta}_{i_a^*,j_a^*,h_a} - \hat{\delta}_{i_a^*,j_a^*,h_a}, \hat{\delta}_{i_b^*,j_b^*,h_b}) = 0$.

Proof. See Section C.2.7. □

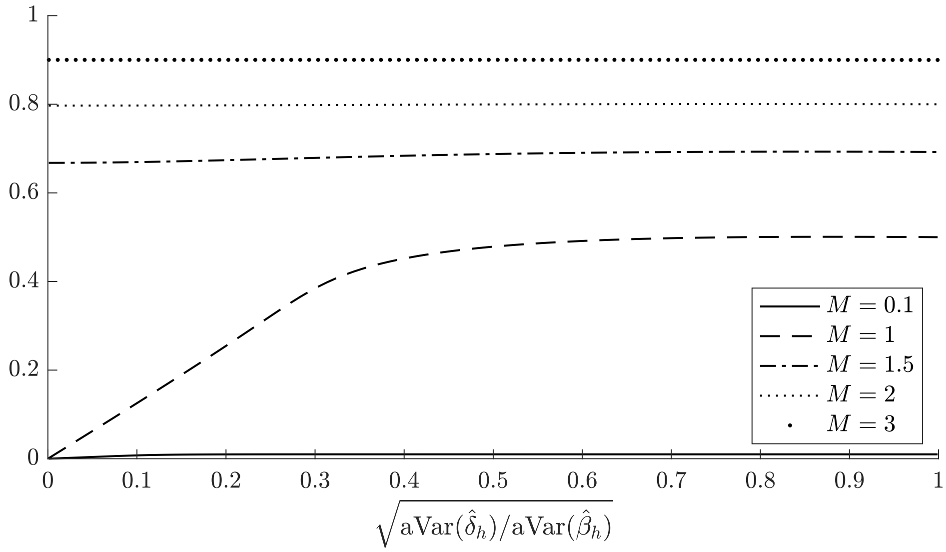

Figure C.1.2: Length-optimal weight on LP in bias-aware confidence interval. Significance level $a = 10\%$. Horizontal axis: relative asymptotic standard deviation of VAR vs. LP. Different lines: different bounds M on $\|\alpha(L)\|$.

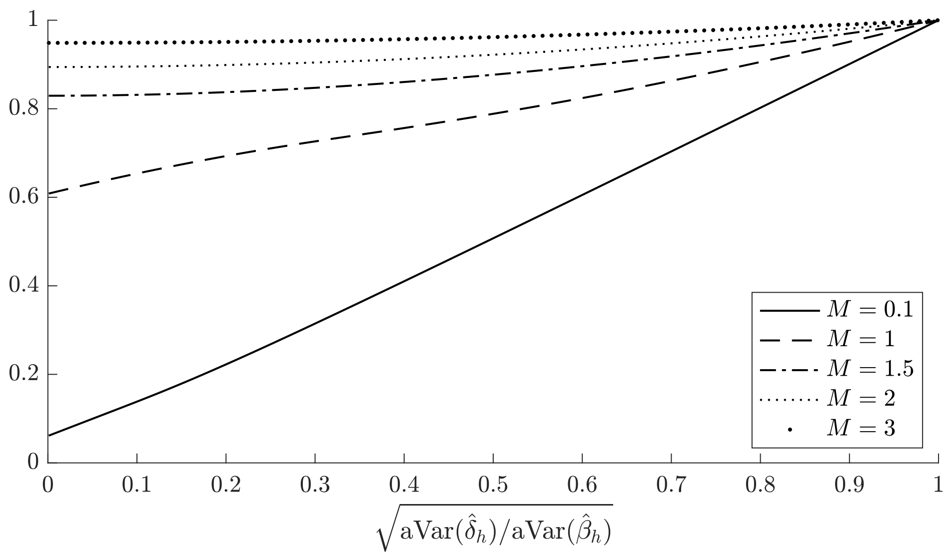

Figure C.1.3: Relative length of optimal bias-aware confidence interval vs. conventional LP interval. Significance level $a = 10\%$. Horizontal axis: relative asymptotic standard deviation of VAR vs. LP. Different lines: different bounds M on $\|\alpha(L)\|$.

C.2 Proofs

C.2.1 Proof of Proposition 3.3.1

Lemma C.5.1 shows that we can represent

$$y_{i^*,t+h} = \theta_{h,T}\varepsilon_{j^*,t} + \underline{B}'_{h,y}\underline{y}_{j^*,t} + B'_{h,y}y_{t-1} + \xi_{i^*,h,t} + T^{-\zeta}\Theta_h(L)\varepsilon_t, \tag{C.2.1}$$

where the expressions for the coefficient matrices and the $1 \times n$ two-sided lag polynomial $\Theta_h(L) = \sum_{\ell=-\infty}^{\infty} \Theta_{h,\ell}L^\ell$ are given in Lemma C.5.1.

Let $\hat{x}_{h,t}$ be the residual in a regression of $y_{j^*,t}$ on $\underline{y}_{j^*,t}$ and y_{t-1}, using data points $1, 2, \ldots, T-h$. By definition, $\hat{x}_{h,t}$ is in-sample orthogonal to $\underline{y}_{j^*,t}$ and y_{t-1}. Hence,

$$\begin{aligned}
\hat{\beta}_h &= \frac{\sum_{t=1}^{T-h} y_{i^*,t+h}\hat{x}_{h,t}}{\sum_{t=1}^{T-h} \hat{x}_{h,t}^2} \\
&= \theta_{h,T} + \frac{\sum_{t=1}^{T-h}(y_{i^*,t+h} - \theta_{h,T}\hat{x}_{h,t} - \underline{B}'_{h,y}\underline{y}_{j^*,t} - B'_{h,y}y_{t-1})\hat{x}_{h,t}}{\sum_{t=1}^{T-h} \hat{x}_{h,t}^2} \quad \text{by orthogonality} \\
&= \theta_{h,T} + \frac{T^{-1}\sum_{t=1}^{T-h}(y_{i^*,t+h} - \theta_{h,T}\hat{x}_{h,t} - \underline{B}'_{h,y}\underline{y}_{j^*,t} - B'_{h,y}y_{t-1})\hat{x}_{h,t}}{\sigma_{j^*}^2 + o_p(1)} \quad \text{by Lemma C.5.4(v)} \\
&= \theta_{h,T} + \frac{T^{-1}\sum_{t=1}^{T-h}(y_{i^*,t+h} - \theta_{h,T}\varepsilon_{j^*,t} - \underline{B}'_{h,y}\underline{y}_{j^*,t} - B'_{h,y}y_{t-1})\hat{x}_{h,t} + O_p(T^{-2\zeta}) + o_p(T^{-1/2})}{\sigma_{j^*}^2 + o_p(1)} \\
&\quad \text{by Lemma C.5.4(iv)} \\
&= \theta_{h,T} + \frac{T^{-1}\sum_{t=1}^{T-h}(\xi_{i^*,h,t} + T^{-\zeta}\Theta_h(L)\varepsilon_t)\hat{x}_{h,t} + O_p(T^{-2\zeta}) + o_p(T^{-1/2})}{\sigma_{j^*}^2 + o_p(1)} \quad \text{by (C.2.1)} \\
&= \theta_{h,T} + \frac{T^{-1}\sum_{t=1}^{T-h}(\xi_{i^*,h,t} + T^{-\zeta}\Theta_h(L)\varepsilon_t)\varepsilon_{j^*,t} + O_p(T^{-2\zeta}) + o_p(T^{-1/2})}{\sigma_{j^*}^2 + o_p(1)} \\
&\quad \text{by Lemma C.5.4(iii) and (vi)} \\
&= \theta_{h,T} + \frac{T^{-1}\sum_{t=1}^{T-h}\xi_{i^*,h,t}\varepsilon_{j^*,t} + O_p(T^{-2\zeta}) + o_p(T^{-1/2})}{\sigma_{j^*}^2 + o_p(1)} \quad \text{by Lemma C.5.1,}
\end{aligned}$$

and the result follows. $\square$

C.2.2 Proof of Proposition 3.3.2

Note first that

$$\begin{aligned}\hat{\delta}_h - e'_{i^*,n}A^h H_{\bullet,j^*} &= e'_{i^*,n}\hat{A}^h\hat{\nu} - e'_{i^*,n}A^h H_{\bullet,j^*} \\ &= e'_{i^*,n}\hat{A}^h H_{\bullet,j^*} - e'_{i^*,n}A^h H_{\bullet,j^*} + e'_{i^*,n}\hat{A}^h(\hat{\nu} - H_{\bullet,j^*}).\end{aligned}$$

Lemma C.5.2 shows that $\hat{A} - A = O_p(T^{-\zeta} + T^{-1/2})$. By Magnus and Neudecker (2007, Table 7, p. 208),

$$\left(\frac{\partial(e'_{i^*,n}A^h H_{\bullet,j^*})}{\partial \boldsymbol{e}(A)}\right)' = (H'_{\bullet,j^*} \otimes e'_{i^*,n})\left(\sum_{\ell=1}^{h}(A')^{h-\ell} \otimes A^{\ell-1}\right) = \sum_{\ell=1}^{h} H'_{\bullet,j^*}(A')^{h-\ell} \otimes e'_{i^*,n}A^{\ell-1},$$

so

$$\begin{aligned}\hat{\delta}_h - e'_{i^*,n}A^h H_{\bullet,j^*} &= \left(\sum_{\ell=1}^{h} H'_{\bullet,j^*}(A')^{h-\ell} \otimes e'_{i^*,n}A^{\ell-1}\right)\boldsymbol{e}(\hat{A} - A) + e'_{i^*,n}A^h(\hat{\nu} - H_{\bullet,j^*}) + o_p(T^{-\zeta} + T^{-1/2}) \\ &= \sum_{\ell=1}^{h} e'_{i^*,n}A^{\ell-1}(\hat{A} - A)A^{h-\ell}H_{\bullet,j^*} + e'_{i^*,n}A^h(\hat{\nu} - H_{\bullet,j^*}) + o_p(T^{-\zeta} + T^{-1/2}) \\ &= \operatorname{tr}\left\{\Psi_h(\hat{A} - A)\right\} + e'_{i^*,n}A^h(\hat{\nu} - H_{\bullet,j^*}) + o_p(T^{-\zeta} + T^{-1/2}),\end{aligned}$$

where $\Psi_h \equiv \sum_{\ell=1}^{h} A^{h-\ell}H_{\bullet,j^*}e'_{i^*,n}A^{\ell-1}$. Lemma C.5.2 further implies that

$$\begin{aligned}\operatorname{tr}\left\{\Psi_h(\hat{A} - A)\right\} &= T^{-\zeta}\operatorname{tr}\left\{S^{-1}\Psi_h H\sum_{\ell=1}^{\infty}\alpha_\ell D H'(A')^{\ell-1}\right\} \\ &\quad + \operatorname{tr}\left\{S^{-1}\Psi_h H T^{-1}\sum_{t=1}^{T}\varepsilon_t\tilde{y}'_{t-1}\right\} + o_p(T^{-\zeta}),\end{aligned}$$

where S was defined in Assumption 3.2.1. Lemma C.5.3 shows that

$$\hat{\nu} - H_{\bullet,j^*} = \frac{1}{\sigma^2_{j^*}}T^{-1}\sum_{t=1}^{T}\xi_{0,t}\varepsilon_{j^*,t} + o_p(T^{-\zeta} + T^{-1/2}).$$

Using the definition of $\theta_{h,T}$ and re-arranging terms gives the desired result. □

C.2.3 Proof of Corollary 3.3.2

Use the notation $E^*(z \mid w) = \text{cov}(z, w)\,\text{var}(w)^{-1} w$ for mean-square projection. Then

$$\begin{aligned}
\sigma_{j^*}^2 \Psi_h' S^{-1} \tilde{y}_{t-1} &= \left(\sum_{\ell=1}^{h} (A')^{h-\ell} e_{i^*,n} \underbrace{\sigma_{j^*}^2 H'_{\bullet,j^*} (A')^{\ell-1}}_{=\text{cov}(\varepsilon_{j^*,t-\ell}, \tilde{y}_{t-1})} \right) S^{-1} \tilde{y}_{t-1} \\
&= \sum_{\ell=1}^{h} (A')^{h-\ell} e_{i^*,n} E^*(\varepsilon_{j^*,t-\ell} \mid \tilde{y}_{t-1}) \\
&= \sum_{\ell=1}^{h} (A')^{h-\ell} e_{i^*,n} \varepsilon_{j^*,t-\ell},
\end{aligned}$$

where the last equality uses $\varepsilon_{j^*,t-\ell} \in \text{span}(\tilde{\tilde{y}}_{t-1}, \dots, \tilde{\tilde{y}}_{t-p})$ for $\ell = 1, \dots, h$. Thus,

$$\begin{aligned}
\text{var}(\varepsilon_t' H' \Psi_h' S^{-1} \tilde{y}_{t-1}) &= \text{var}\left(\frac{1}{\sigma_{j^*}^2} \sum_{\ell=1}^{h} \varepsilon_{j^*,t-\ell} \varepsilon_t' H' (A')^{h-\ell} e_{i^*,n} \right) \\
&= \frac{1}{\sigma_{j^*}^4} \sum_{\ell=1}^{h} \text{var}\left(\varepsilon_{j^*,t-\ell} \varepsilon_t' H' (A')^{h-\ell} e_{i^*,n} \right) \\
&= \frac{1}{\sigma_{j^*}^4} \sum_{\ell=1}^{h} E(\varepsilon_{j^*,t-\ell}^2)\, \text{var}(\varepsilon_t' H' (A')^{h-\ell} e_{i^*,n}) \\
&= \frac{1}{\sigma_{j^*}^2} \text{var}\left(e_{i^*,n}' \sum_{\ell=1}^{h} A^{h-\ell} H \varepsilon_{t+\ell} \right).
\end{aligned}$$

It now follows as in the proof of Corollary C.1.2 that $\text{aVar}(\hat{\beta}_h) = \text{aVar}(\hat{\delta}_h)$. Then Proposition 3.4.1 implies that $\text{aBias}(\hat{\delta}_h) = 0$. □

C.2.4 Proof of Corollary 3.4.1

By Proposition 3.4.1, $\sup_{\alpha(L) \colon \|\alpha(L)\| \le M} \text{aBias}(\hat{\delta}_h; \alpha(L))^2 = M^2 \{\text{aVar}(\hat{\beta}_h) - \text{aVar}(\hat{\delta}_h)\}$. The result follows. □

C.2.5 Proof of Corollary 3.4.2

Write $\hat{\theta}_h(\omega) = \hat{\delta}_h + \omega(\hat{\beta}_h - \hat{\delta}_h)$. By Corollary C.1.2, the two terms are asymptotically independent of each other, and the second term has asymptotic variance $\omega^2\{\text{aVar}(\hat{\beta}_h) -$

$\mathrm{aVar}(\hat{\delta}_h)\}$. Hence,

$$\mathrm{aMSE}(\hat{\theta}_h(\omega)) = \{(1-\omega)\,\mathrm{aBias}(\hat{\delta}_h)\}^2 + \mathrm{aVar}(\hat{\delta}_h) + \omega^2\{\mathrm{aVar}(\hat{\beta}_h) - \mathrm{aVar}(\hat{\delta}_h)\}.$$

By Proposition 3.4.1, the supremum of the above expression over $\alpha(L)$ satisfying $\|\alpha(L)\| \leq M$ equals

$$(1-\omega)^2 M^2\{\mathrm{aVar}(\hat{\beta}_h) - \mathrm{aVar}(\hat{\delta}_h)\} + \mathrm{aVar}(\hat{\delta}_h) + \omega^2\{\mathrm{aVar}(\hat{\beta}_h) - \mathrm{aVar}(\hat{\delta}_h)\}.$$

To find the ω that minimizes the above expression, we can equivalently minimize the function $(1-\omega)^2 M^2 + \omega^2$. The result follows. □

C.2.6 Proof of Corollary 3.4.4

Proposition 3.4.1 implies that the absolute relative VAR bias $|b_h|$ can be made to take any value in $[0, \infty)$ as $\alpha(L)$ varies over the set of all absolutely summable lag polynomials. The corollary then follows from Corollaries C.1.2 and 3.3.1 and Proposition 3.3.3. □

C.2.7 Proof of Corollary C.1.2

We first use Proposition 3.3.1 to compute $\mathrm{aCov}(\hat{\beta}_{i_a^*,j_a^*,h_a}, \hat{\beta}_{i_b^*,j_b^*,h_b})$. Define $\xi_{j^*,h,t} = (\xi_{1,j^*,h,t}, \dots, \xi_{n,j^*,h,t})'$ as in Proposition 3.3.1, but making the dependence on both i^* and j^* explicit in the notation. Observe that

$$E[\xi_{i_a^*,j_a^*,h_a,t}\varepsilon_{j_a^*,t}\xi_{i_b^*,j_b^*,h_b,s}\varepsilon_{j_b^*,s}] = 0 \quad \text{for all } s \neq t.$$

Hence,

$$\mathrm{aCov}(\hat{\beta}_{i_a^*,j_a^*,h_a}, \hat{\beta}_{i_b^*,j_b^*,h_b}) = \frac{1}{\sigma_{j_a^*}^2\sigma_{j_b^*}^2} E[\xi_{i_a^*,j_a^*,h_a,t}\varepsilon_{j_a^*,t}\xi_{i_b^*,j_b^*,h_b,t}\varepsilon_{j_b^*,t}].$$

If $j_a^* < j_b^*$, then $\varepsilon_{j_a^*,t}$ is independent of all the other terms in the above expectation, so the expectation equals zero; similarly if $j_a^* > j_b^*$. Now consider the case $j_a^* = j_b^*$:

$$\begin{aligned}
\text{aCov}(\hat{\beta}_{i_a^*,j_a^*,h_a}, \hat{\beta}_{i_b^*,j_a^*,h_b}) &= \frac{1}{\sigma_{j_a^*}^4} E[\xi_{i_a^*,j_a^*,h_a,t}\xi_{i_b^*,j_a^*,h_b,t}\varepsilon_{j_a^*,t}^2] \\
&= \frac{1}{\sigma_{j_a^*}^4} E[\xi_{i_a^*,j_a^*,h_a,t}\xi_{i_b^*,j_a^*,h_b,t}]E[\varepsilon_{j_a^*,t}^2] \\
&= \frac{1}{\sigma_{j_a^*}^2} E[\xi_{i_a^*,j_a^*,h_a,t}\xi_{i_b^*,j_a^*,h_b,t}] \\
&= \frac{1}{\sigma_{j_a^*}^2} \Bigg(E[e_{i_a^*,n}' A^{h_a} \overline{H}_{j_a^*} \overline{\varepsilon}_{j_a^*,t} \overline{\varepsilon}_{j_a^*,t}' \overline{H}_{j_a^*}' (A')^{h_b} e_{i_b^*,n}] \\
&\qquad + E\left[e_{i_a^*,n}' \sum_{\ell_1=1}^{h_a} \sum_{\ell_2=1}^{h_b} A^{h_a-\ell_1} H \varepsilon_{t+\ell_1} \varepsilon_{t+\ell_2}' H' (A')^{h_b-\ell_2} e_{i_b^*,n} \right] \Bigg) \\
&= \frac{1}{\sigma_{j_a^*}^2} \left(\psi_{a,b} + \sum_{\ell=1}^{\min\{h_a,h_b\}} e_{i_a^*,n}' A^{h_a-\ell} \Sigma (A')^{h_b-\ell} e_{i_b^*,n} \right),
\end{aligned}$$

as claimed.

We now derive $\text{aCov}(\hat{\delta}_{i_a^*,j_a^*,h_a}, \hat{\delta}_{i_b^*,j_b^*,h_b})$ using Proposition 3.3.2. Observe that the vector process $(\varepsilon_t' \otimes \tilde{y}_{t-1}', \xi_{j_a^*,0,t}' \varepsilon_{j_a^*,t}, \xi_{j_b^*,0,t}' \varepsilon_{j_b^*,t})'$ is a martingale difference sequence with respect to the filtration generated by $\{\varepsilon_t\}$. Moreover, $E[(\varepsilon_t \otimes \tilde{y}_{t-1})\xi_{j^*,0,t}' \varepsilon_{j^*,t}] = 0$ for any j^*. Hence,

$$\begin{aligned}
\text{aCov}(\hat{\delta}_{i_a^*,j_a^*,h_a}, \hat{\delta}_{i_b^*,j_b^*,h_b}) &= E\left[\text{tr}\left(S^{-1} \Psi_{i_a^*,j_a^*,h_a} H \varepsilon_t \tilde{y}_{t-1}' \right) \text{tr}\left(S^{-1} \Psi_{i_b^*,j_b^*,h_b} H \varepsilon_t \tilde{y}_{t-1}' \right) \right] \\
&\qquad + \frac{1}{\sigma_{j_a^*}^2 \sigma_{j_b^*}^2} E\left[e_{i_a^*,n}' A^{h_a} \xi_{j_a^*,0,t} \xi_{j_b^*,0,t}' (A')^{h_b} e_{i_b^*,n} \varepsilon_{j_a^*,t} \varepsilon_{j_b^*,t} \right].
\end{aligned}$$

The second term on the right-hand side above equals $\mathbb{1}(j_a^* = j_b^*)\sigma_{j_a^*}^{-2}\psi_{a,b}$, by similar arguments as in the earlier LP calculation. The first term on the right-hand side above equals

$$
\begin{aligned}
&E\left[\tilde{y}_{t-1}' S^{-1} \Psi_{i_a^*, j_a^*, h_a} H \varepsilon_t \varepsilon_t' H' \Psi_{i_b^*, j_b^*, h_b}' S^{-1} \tilde{y}_{t-1}\right] \\
&= \operatorname{tr}\left(E\left[\tilde{y}_{t-1} \tilde{y}_{t-1}' S^{-1} \Psi_{i_a^*, j_a^*, h_a} H \varepsilon_t \varepsilon_t' H' \Psi_{i_b^*, j_b^*, h_b}' S^{-1}\right]\right) \\
&= \operatorname{tr}\left(E\left[\tilde{y}_{t-1} \tilde{y}_{t-1}'\right] S^{-1} \Psi_{i_a^*, j_a^*, h_a} H E\left[\varepsilon_t \varepsilon_t'\right] H' \Psi_{i_b^*, j_b^*, h_b}' S^{-1}\right) \\
&= \operatorname{tr}\left(S S^{-1} \Psi_{i_a^*, j_a^*, h_a} H D H' \Psi_{i_b^*, j_b^*, h_b}' S^{-1}\right) \\
&= \operatorname{tr}\left(\Psi_{i_a^*, j_a^*, h_a} \Sigma \Psi_{i_b^*, j_b^*, h_b}' S^{-1}\right),
\end{aligned}
$$

as claimed.

Finally, we compute $\operatorname{aCov}(\hat{\beta}_{i_a^*, j_a^*, h_a}, \hat{\delta}_{i_b^*, j_b^*, h_b})$ using Propositions 3.3.1 and 3.3.2. Using arguments similar to above, we obtain

$$
\begin{aligned}
&\operatorname{aCov}(\hat{\beta}_{i_a^*, j_a^*, h_a}, \hat{\delta}_{i_b^*, j_b^*, h_b}) \\
&= \frac{1}{\sigma_{j_a^*}^2} \sum_{s=-\infty}^{\infty} E\left[e_{i_a^*, n}' \sum_{\ell=1}^{h_a} A^{h_a-\ell} H \varepsilon_{t+\ell} \varepsilon_{j_a^*, t} \operatorname{tr}\left(S^{-1} \Psi_{i_b^*, j_b^*, h_b} H \varepsilon_{t+s} \tilde{y}_{t+s-1}'\right)\right] + \mathbb{1}(j_a^* = j_b^*)\sigma_{j_a^*}^{-2}\psi_{a,b}.
\end{aligned}
$$

The first term on the left-hand side above equals

$$
\begin{aligned}
&\frac{1}{\sigma_{j_a^*}^2} \sum_{\ell=1}^{h_a} \sum_{s=-\infty}^{\infty} E\left[e_{i_a^*, n}' A^{h_a-\ell} H \varepsilon_{t+\ell} \varepsilon_{t+s}' H' \Psi_{i_b^*, j_b^*, h_b}' S^{-1} \tilde{y}_{t+s-1} \varepsilon_{j_a^*, t}\right] \\
&= \frac{1}{\sigma_{j_a^*}^2} \sum_{\ell=1}^{h_a} E\left[e_{i_a^*, n}' A^{h_a-\ell} H \varepsilon_{t+\ell} \varepsilon_{t+\ell}' H' \Psi_{i_b^*, j_b^*, h_b}' S^{-1} \tilde{y}_{t+\ell-1} \varepsilon_{j_a^*, t}\right] \\
&= \frac{1}{\sigma_{j_a^*}^2} \sum_{\ell=1}^{h_a} e_{i_a^*, n}' A^{h_a-\ell} H E[\varepsilon_{t+\ell} \varepsilon_{t+\ell}'] H' \Psi_{i_b^*, j_b^*, h_b}' S^{-1} E[\tilde{y}_{t+\ell-1} \varepsilon_{j_a^*, t}] \\
&= \frac{1}{\sigma_{j_a^*}^2} \sum_{\ell=1}^{h_a} e_{i_a^*, n}' A^{h_a-\ell} H D H' \Psi_{i_b^*, j_b^*, h_b}' S^{-1} A^{\ell-1} H_{\bullet, j_a^*} \sigma_{j_a^*}^2 \\
&= \operatorname{tr}\Bigg(\underbrace{\sum_{\ell=1}^{h_a} A^{\ell-1} H_{\bullet, j_a^*} e_{i_a^*, n}' A^{h_a-\ell}}_{=\sum_{\ell=1}^{h_a} A^{h_a-\ell} H_{\bullet, j_a^*} e_{i_a^*, n}' A^{\ell-1} = \Psi_{i_a^*, j_a^*, h_a}} \Sigma \Psi_{i_b^*, j_b^*, h_b}' S^{-1}\Bigg).
\end{aligned}
$$

It follows that $\text{aCov}(\hat{\beta}_{i_a^*,j_a^*,h_a}, \hat{\delta}_{i_b^*,j_b^*,h_b}) = \text{aCov}(\hat{\delta}_{i_a^*,j_a^*,h_a}, \hat{\delta}_{i_b^*,j_b^*,h_b})$, as claimed. □

C.3 Further theoretical results

C.3.1 Heteroskedasticity

The conclusions of Propositions 3.3.1 and 3.3.2 do not require shock independence, either cross-sectionally or across time. In particular, our proofs of these propositions and the auxiliary lemmas in Appendix C.5 replace Assumption 3.2.1(i) by the following:[1]

Assumption C.3.1. *ε_t is a strictly stationary martingale difference sequence with respect to its natural filtration;* $\text{var}(\varepsilon_t) = D \equiv \text{diag}(\sigma_1^2, \dots, \sigma_m^2)$*;* $\sigma_j > 0$ *for all* $j = 1, \dots, m$*;* $E[\|\varepsilon_t\|^4] < \infty$*; and* $\sum_{\ell=1}^{\infty} \sum_{\tau=1}^{\infty} \|\text{cov}(\varepsilon_t \otimes \varepsilon_t, \varepsilon_{t-\ell} \otimes \varepsilon_{t-\tau})\| < \infty$.

Assumption C.3.1 strictly weakens Assumption 3.2.1(i) by allowing the shocks to be conditionally heteroskedastic and by weakening mutual shock independence to orthogonality. The last part of Assumption C.3.1 is a vector version of the fourth-order cumulant summability condition of Kuersteiner (2001, Assumption A1). It restricts the higher-order dependence of the shocks, consistent with many stationary models of conditional heteroskedasticity.[2]

C.3.2 External instruments and proxies

Our framework can accommodate identification via an external instrument (also known as a proxy) by a simple reparametrization. To see this, let the proxy be ordered first in y_t. Set $j^* = 1$, and replace Assumption 3.2.1(iii) with the following:

[1] It is an interesting topic for future research to investigate whether the other results in Sections 3.3 and 3.4 also hold under the weaker condition.

[2] One sufficient condition is finite dependence, i.e., there exists an integer K such that $\{\varepsilon_s\}_{s \geq t+K}$ is independent of $\{\varepsilon_s\}_{s \leq t}$. Another set of sufficient conditions is that $E(\varepsilon_t \mid \{\varepsilon_s\}_{s \neq t}) = 0$ and $\{\varepsilon_t \otimes \varepsilon_t\}$ has absolutely summable autocovariance function. See also Kuersteiner (2001, Remark 2, p. 362).

Assumption C.3.2. *The first column of A consists of zeros, except possibly the first element; the first row of H equals $(1, 0, 0, \dots, 0, 1)$; the last column of H consists of zeros, except the first element; and the last column of $\alpha(L)$ consists of zeros.*

This assumption imposes the following restrictions:

- The proxy $y_{1,t}$ equals

$$y_{1,t} = A_{1,\bullet} y_{t-1} + \varepsilon_{1,t} + \varepsilon_{m,t} + T^{-\zeta}[\alpha_{1,\bullet}(L) + \alpha_{m,\bullet}(L)]\varepsilon_t,$$

  which generalizes Assumption 4 in Plagborg-Møller and Wolf (2021) to allow for local contamination by lagged shocks. The last shock $\varepsilon_{m,t}$ is viewed as measurement error or noise (v_t in the notation of Plagborg-Møller and Wolf, 2021).

- The dynamics of $(y_{2,t}, \dots, y_{n,t})$ (i.e., with the proxy excluded) follow a VAR(1), up to the local misspecification in the form of lags of $(\varepsilon_{1,t}, \dots, \varepsilon_{m-1,t})$.

- The proxy measurement error $\varepsilon_{m,t}$ is orthogonal to all leads and lags of $(y_{2,t}, \dots, y_{n,t})$.

Now transform the shocks from ε_t to $\tilde{\varepsilon}_t = (\tilde{\varepsilon}_{1,t}, \dots, \tilde{\varepsilon}_{m,t})'$ as follows:

$$\tilde{\varepsilon}_{1,t} \equiv \varepsilon_{1,t} + \varepsilon_{m,t}; \quad \tilde{\varepsilon}_{j,t} \equiv \varepsilon_{j,t} \text{ for } j = 2, \dots, m-1; \quad \tilde{\varepsilon}_{m,t} \equiv \varepsilon_{m,t} - \frac{\sigma_m^2}{\sigma_1^2 + \sigma_m^2}(\varepsilon_{1,t} + \varepsilon_{m,t}).$$

By construction, the elements of $\tilde{\varepsilon}_t$ are mutually orthogonal. Note that

$$\varepsilon_{1,t} = \frac{\sigma_1^2}{\sigma_1^2 + \sigma_m^2}\tilde{\varepsilon}_{1,t} - \tilde{\varepsilon}_{m,t}; \quad \varepsilon_{j,t} = \tilde{\varepsilon}_{j,t} \text{ for } j = 2, \dots, m-1; \quad \varepsilon_{m,t} = \frac{\sigma_m^2}{\sigma_1^2 + \sigma_m^2}\tilde{\varepsilon}_{1,t} + \tilde{\varepsilon}_{m,t};$$

write Q for the $m \times m$ matrix such that $\varepsilon_t = Q\tilde{\varepsilon}_t$. We can then re-express the VARMA$(1, \infty)$ model for y_t in Equation (3.2.1) as

$$y_t = A y_{t-1} + \tilde{H}[I + T^{-\zeta}\tilde{\alpha}(L)]\tilde{\varepsilon}_t, \tag{C.3.1}$$

where

$$\tilde{H} \equiv HQ = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ \frac{\sigma_1^2}{\sigma_1^2+\sigma_m^2} H_{2,1} & H_{2,2} & H_{2,3} & \cdots & H_{2,m-1} & -H_{2,1} \\ \vdots & & & & & \vdots \\ \frac{\sigma_1^2}{\sigma_1^2+\sigma_m^2} H_{n,1} & H_{n,2} & H_{n,3} & \cdots & H_{n,m-1} & -H_{n,1} \end{pmatrix},$$

and $\tilde{\alpha}(L) \equiv Q^{-1}\alpha(L)Q$. Under Assumption C.3.2 above, the impulse responses of $(y_{2,t}, \dots, y_{n,t})$ with respect to $\tilde{\varepsilon}_{1,t}$ in the reparametrized system (C.3.1) equal $\sigma_1^2/(\sigma_1^2 + \sigma_m^2)$ times the impulse responses of $(y_{2,t}, \dots, y_{n,t})$ with respect to $\varepsilon_{1,t}$ in the original parametrization in Equation (3.2.1). This follows by inspection of the elements $\tilde{H}_{i,1}$ for $i \geq 2$, the assumption that $A_{i,1} = 0$ for $i \geq 2$, and the fact that the first column of $\tilde{H}\tilde{\alpha}(L)$ equals $H\alpha(L)Q_{\bullet,1} = \frac{\sigma_1^2}{\sigma_1^2+\sigma_m^2} H\alpha_{\bullet,1}(L)$ (here we use the assumption that the last column of $\alpha(L)$ is zero).

If the original shocks (and measurement error) ε_t satisfy Assumption C.3.1, then the transformed shocks $\tilde{\varepsilon}_t$ do also, provided that the fourth-order cumulant condition holds for the transformed shocks (recall that Assumption C.3.1 only requires the shocks to be mutually orthogonal, not independent). The transformed system (C.3.1) therefore satisfies Assumption C.3.1 and Assumption 3.2.1(ii)–(v), with $\tilde{H}$ and $\tilde{\alpha}(L)$ in place of H and $\alpha(L)$. Hence, Propositions 3.3.1 and 3.3.2 apply. We conclude that LPs on the proxy $y_{1,t}$—and recursive VARs with the proxy ordered first—consistently estimate the true impulse responses of $(y_{2,t}, \dots, y_{n,t})$ with respect to $\varepsilon_{1,t}$, up to the scale factor $\frac{\sigma_1^2}{\sigma_1^2+\sigma_m^2}$. This scale factor, which is the same across all response variables i^* and horizons h, reflects attenuation bias caused by the measurement error in the proxy. It gets canceled out if one reports relative (i.e., unit-effect-normalized) impulse responses, as explained by Plagborg-Møller and Wolf (2021, Section 3.3). Of course, though the VAR estimator is consistent, it suffers from asymptotic bias of order $T^{-\zeta}$, while the LP estimator has bias of the smaller order $T^{-2\zeta}$.

In summary, the main results in the paper carry over to identification via an external instrument/proxy. A caveat is that the instrument/proxy must be strong, in the sense that

$\sigma_1^2 = \text{var}(\varepsilon_{1,t})$ is not close to zero; otherwise, we will end up dividing by a number close to zero when computing relative impulse responses.

C.3.3 Data-dependent lag selection

Here we argue that local projection inference is more robust to data-dependent lag selection errors than conventional VAR inference. The following proposition establishes the properties of conventional information criteria in our class of DGPs.

Proposition C.3.1. *Assume that the $\check{n}$-dimensional process $\{\check{y}_t\}$ is a stationary solution of the local-to-SVAR(p_0) model in Equation (3.2.2). Assume also that $\check{A}_{p_0} \neq 0$ so that p_0 is the minimal true autoregressive lag order, $\check{H}D\check{H}'$ is positive definite, and the process (written in companion form as on p. 7 of the main paper) satisfies Assumption 3.2.1.*

Let $\hat{\check{\Sigma}}(p)$ denote the $\check{n} \times \check{n}$ sample residual variance-covariance matrix from a least-squares VAR(p) regression on the data $(\check{y}_1, \ldots, \check{y}_T)$. Let $\{g_T\}_T$ be a deterministic scalar sequence. Fix a maximal lag length $K \geq p_0$. Suppose we select the lag length by minimizing an information criterion:

$$\hat{p} \equiv \underset{0 \leq p \leq K}{\operatorname{argmin}} \left\{ \log \det \left(\hat{\check{\Sigma}}(p) \right) + p \times g_T \right\}.$$

Then the following statements hold:

i) If $g_T \to 0$, then $P(\hat{p} < p_0) \to 0$ as $T \to \infty$.

ii) If $T g_T \to \infty$ and $\zeta \geq 1/2$, then $P(\hat{p} > p_0) \to 0$ as $T \to \infty$.

The proposition implies that, when applied to a VAR in the data $\{\check{y}_t\}$, both the BIC ($g_T = \check{n}^2(\log T)/T$) and AIC ($g_T = 2\check{n}^2/T$) select p_0 or more lags with high probability in large samples, and in fact the BIC selects exactly p_0 lags asymptotically when $\zeta \geq 1/2$.[3]

[3]For additional intuition, consider the critical case $\zeta = 1/2$ where the moving average coefficients are of order $T^{-1/2}$. Under smoothness assumptions on the density of the shocks ε_t, the local-to-SVAR(p_0) DGP is then contiguous to an exact SVAR(p_0) DGP (for formal results, see Hallin and Puri, 1988; Hallin et al., 1989). Since the BIC is consistent for p_0 in latter DGP, contiguity implies that the BIC also selects p_0 lags with probability approaching 1 in the former DGP.

Hence, in the latter case, it follows from Corollary 3.3.1 that it is pointwise asymptotically valid to report a local projection confidence interval that controls for $\hat{p}$ lags of the data, where $\hat{p}$ is selected by applying the BIC to auxiliary VAR regressions as defined above.

Unlike LP inference, VAR inference is sensitive to minor model selection errors. Leeb and Pötscher (2005) show that VAR confidence intervals with lag length selected by BIC or AIC fail to control coverage uniformly over the VAR parameter space. The breakdown in performance happens for (a sequence of) DGPs that satisfy a VAR($\tilde{p}_0$) model where the first p_0 lags have large coefficients and the remaining $\tilde{p}_0 - p_0$ lags have small coefficients of order $T^{-1/2}$; this implies a VARMA(p_0, ∞) representation where the moving average coefficients are of order $T^{-1/2}$. Because the BIC or AIC cannot reliably detect the small coefficients, they will tend to select fewer than $\tilde{p}_0$ lags (cf. Proposition C.3.1). The small coefficients on the omitted lags impart an asymptotic bias in the VAR estimator that can cause large coverage distortions for the associated confidence interval, consistent with Corollary 3.3.1. However, if in this context the BIC-selected lag length is instead used for *local projection* inference, the bias imparted by the omitted lags is much smaller than for the VAR estimator and is in fact asymptotically negligible: this is precisely the message of Proposition 3.3.1. Though our arguments fall short of proving that local projection inference with data-dependent lag length is *uniformly* valid over some parameter space, they nevertheless show that LP inference remains valid under the types of drifting parameter sequences that Leeb and Pötscher (2005) show are responsible for the non-uniformity of VAR inference.

C.3.3.1 Proof of Proposition C.3.1

The proof follows standard arguments for exact VAR models, see Lütkepohl (2005, Chapter 4.3.2) and references therein. We merely show that extra terms induced by the vanishing moving average process are asymptotically negligible under our assumptions. Let $\|\cdot\|$ denote the Frobenius norm.

Lemma C.5.6 implies that $T^{-1}\sum_{t=1}^{T}\check{y}_t\check{y}_{t-\ell}' = T^{-1}\sum_{t=1}^{T}\tilde{\check{y}}_t\tilde{\check{y}}_{t-\ell}' + O_p(T^{-\zeta})$ for any $\ell \geq 0$, where $\{\tilde{\check{y}}_t\}$ is a process that satisfies the VAR(p_0) model with no moving average term ($\alpha(L) = 0$), see also Corollary 3.3.2. It follows from least-squares algebra that the probability limit of $\check{\hat{\Sigma}}(p)$ for any fixed p is the same as it would be in an exact VAR(p_0) DGP with no moving average term. Statement (i)) of the proposition then follows immediately from standard arguments for lag length estimation in exact VAR models (Lütkepohl, 2005, Proposition 4.2).

To prove statement (ii)), assume $p \geq p_0$. It suffices to show that $\check{\hat{\Sigma}}(p) = \check{\hat{\Sigma}}_0 + O_p(T^{-1})$ for some data-dependent matrix $\check{\hat{\Sigma}}_0$ independent of p, since this implies[4]

$$P\left(\log\det\left(\check{\hat{\Sigma}}(p)\right) + pg_T > \log\det\left(\check{\hat{\Sigma}}(p_0)\right) + p_0 g_T\right) = P\Big((p-p_0)Tg_T > O_p(1)\Big) \to 1$$

when $p > p_0$. Let y_t denote the $(\check{n}p)$-dimensional companion form vector obtained by stacking p lags of $\check{y}_t$, as in Equation (3.2.2). Then $\check{\hat{\Sigma}}(p)$ is the upper left $\check{n} \times \check{n}$ block of the $(\check{n}p) \times (\check{n}p)$ matrix $\hat{\Sigma} = T^{-1}\sum_{t=1}^{T}\hat{u}_t\hat{u}_t'$ defined in Section 3.2.2. To finish the proof, we show that $T^{-1}\sum_{t=1}^{T}\|\hat{u}_t - H\varepsilon_t\|^2 = O_p(T^{-1})$, which by Cauchy-Schwarz implies $\hat{\Sigma} = T^{-1}\sum_{t=1}^{T}H\varepsilon_t\varepsilon_t'H' + O_p(T^{-1})$, as needed.

Note that when the estimation lag length p weakly exceeds the true autoregressive order p_0, all results in our paper apply, as noted in the discussion surrounding Equation (3.2.2). Hence, using the definition of u_t in Lemma C.5.6,

$$\begin{aligned}
\frac{1}{T}\sum_{t=1}^{T}\|\hat{u}_t - H\varepsilon_t\|^2 &\leq \frac{2}{T}\sum_{t=1}^{T}\|\underbrace{\hat{u}_t - u_t}_{(A-\hat{A})y_{t-1}}\|^2 + \frac{2}{T}\sum_{t=1}^{T}\|\underbrace{u_t - H\varepsilon_t}_{T^{-\zeta}H\alpha(L)\varepsilon_t}\|^2 \\
&\leq 2\underbrace{\|\hat{A} - A\|^2}_{O_p((T^{-1/2}+T^{-\zeta})^2)}\underbrace{\frac{1}{T}\sum_{t=1}^{T}\|y_{t-1}\|^2}_{O_p(1)} + 2T^{-2\zeta}\|H\|^2\underbrace{\frac{1}{T}\sum_{t=1}^{T}\|\alpha(L)\varepsilon_t\|^2}_{O_p(1)} \\
&= O_p(T^{-1} + T^{-2\zeta}),
\end{aligned}$$

[4] Actually, in order to apply the delta method to the log determinant, we also use that $\operatorname{plim}\check{\hat{\Sigma}}_0 = \check{H}D\check{H}'$, a matrix that is non-singular by assumption.

where the three $O_p(\cdot)$ statements in the penultimate line rely on Lemmas C.5.2 and C.5.8 and Assumption 3.2.1(v), respectively. When $\zeta \geq 1/2$, the right-hand side above is $O_p(T^{-1})$. $\square$

C.3.4 Inference on multiple impulse responses

This subsection generalizes the worst-case bias formula in Proposition 3.4.1 to the multi-dimensional case and derives the worst-case coverage of the Wald confidence ellipsoid.

C.3.4.1 Set-up

We consider inference on any combination of impulse responses for various horizons h, response variables i^*, and shocks j^*. When referring to impulse responses and estimators of these, we need to make the response variable and shock explicit in the notation. Thus, we write $\theta_{i^*,j^*,h,T}$, $\hat{\beta}_{i^*,j^*,h}$, and $\hat{\delta}_{i^*,j^*,h}$, with the definitions being the same as in Section 3.2. Let k denote the total number of impulse responses of interest. We refer to the list of impulse responses by the collection of triples $\{(i_a^*, j_a^*, h_a)\}_{a=1}^k$ indexing the response variable, shock variable, and horizon, respectively. Define the k-dimensional vectors of true impulse responses and LP and VAR estimators:

$$\boldsymbol{\theta}_T \equiv \begin{pmatrix} \theta_{i_1^*,j_1^*,h_1,T} \\ \vdots \\ \theta_{i_k^*,j_k^*,h_k,T} \end{pmatrix}, \quad \hat{\boldsymbol{\beta}} \equiv \begin{pmatrix} \hat{\beta}_{i_1^*,j_1^*,h_1} \\ \vdots \\ \hat{\beta}_{i_k^*,j_k^*,h_k} \end{pmatrix}, \quad \hat{\boldsymbol{\delta}} \equiv \begin{pmatrix} \hat{\delta}_{i_1^*,j_1^*,h_1} \\ \vdots \\ \hat{\delta}_{i_k^*,j_k^*,h_k} \end{pmatrix}.$$

It follows from Propositions 3.3.1 and 3.3.2 that, when $\zeta = 1/2$,

$$\sqrt{T}\begin{pmatrix} \hat{\boldsymbol{\beta}} - \boldsymbol{\theta}_T \\ \hat{\boldsymbol{\delta}} - \boldsymbol{\theta}_T \end{pmatrix} \xrightarrow{d} N\left(\begin{pmatrix} 0_{k\times 1} \\ \mathrm{aBias}(\hat{\boldsymbol{\delta}}) \end{pmatrix}, \begin{pmatrix} \mathrm{aVar}(\hat{\boldsymbol{\beta}}) & \mathrm{aCov}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\delta}}) \\ \mathrm{aCov}(\hat{\boldsymbol{\delta}}, \hat{\boldsymbol{\beta}}) & \mathrm{aVar}(\hat{\boldsymbol{\delta}}) \end{pmatrix} \right), \tag{C.3.2}$$

for a k-dimensional vector $\mathrm{aBias}(\hat{\boldsymbol{\delta}})$ (defined in the proof of Proposition C.3.2 below) and $k \times k$ matrices $\mathrm{aVar}(\hat{\boldsymbol{\beta}})$, $\mathrm{aVar}(\hat{\boldsymbol{\delta}})$, and $\mathrm{aCov}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\delta}})$ given in Corollary C.1.2 in Appendix C.1.3.

This corollary also implies that the difference $\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\delta}}$ is asymptotically independent of $\hat{\boldsymbol{\delta}}$, which is not surprising given the general arguments of Hausman (1978) and the facts that (i) the asymptotic variances of the estimators are the same as in the model with $\alpha(L) = 0$ and (ii) the VAR estimator is the quasi-MLE in such a model. It follows that $\text{aVar}(\hat{\boldsymbol{\beta}}) \geq \text{aVar}(\hat{\boldsymbol{\delta}})$ in the positive semidefinite sense.

C.3.4.2 Worst-case bias

The following result generalizes the univariate worst-case bias formula in Proposition 3.4.1.

Proposition C.3.2. *Impose Assumption 3.2.1, with part (iii) holding for all shock indices $j_1^*, \ldots, j_k^*$, and let $\zeta = 1/2$. Let R be a constant matrix with k columns. Then*

$$\max_{\alpha(L)\colon\, \|\alpha(L)\| \leq M} \|R\, \text{aBias}(\hat{\boldsymbol{\delta}})\|^2 = M^2 \lambda_{\max}\Big(R[\text{aVar}(\hat{\boldsymbol{\beta}}) - \text{aVar}(\hat{\boldsymbol{\delta}})]R'\Big),$$

where $\lambda_{\max}(B)$ denotes the largest eigenvalue of the matrix B.

The proposition shows that the worst-case squared norm of the bias of the VAR estimator $R\hat{\boldsymbol{\delta}}$ of $R\boldsymbol{\theta}_T$ is a function of two simple quantities: the bound M on misspecification, and the largest eigenvalue of the difference $\text{aVar}(R\hat{\boldsymbol{\beta}}) - \text{aVar}(R\hat{\boldsymbol{\delta}})$ between the variance-covariance matrices for the LP and VAR estimators. The latter eigenvalue equals $\max_{\|\varsigma\|=1}\{\text{aVar}(\varsigma' R\hat{\boldsymbol{\beta}}) - \text{aVar}(\varsigma' R\hat{\boldsymbol{\delta}})\}$, i.e., the *largest* efficiency gain for VAR over LP across all linear combinations (with norm 1) of the estimated parameters. Consequently, the worst-case bias is non-negligible if the VAR offers efficiency gains for *any* linear combination of the parameters of interest, echoing our univariate results. When R is a row vector, then the proposition implies that our conclusions from Section 3.4.2 extend to inference on any *linear combination* of impulse responses.

C.3.4.3 Worst-case coverage of confidence ellipsoid

We next derive the coverage of the conventional Wald confidence ellipsoid based on the VAR estimator. The level-$(1-a)$ confidence ellipsoid is given by

$$\mathrm{CE}(\hat{\boldsymbol{\delta}}) \equiv \left\{\tilde{\boldsymbol{\theta}} \in \mathbb{R}^k \colon T(\hat{\boldsymbol{\delta}} - \tilde{\boldsymbol{\theta}})' \operatorname{aVar}(\hat{\boldsymbol{\delta}})^{-1}(\hat{\boldsymbol{\delta}} - \tilde{\boldsymbol{\theta}}) \leq \chi^2_{1-a,k}\right\},$$

where $\chi^2_{1-a,k}$ is the $1-a$ quantile of the χ^2 distribution with k degrees of freedom.

Corollary C.3.1. *Impose Assumption 3.2.1, with part (iii) holding for all shock indices $j_1^*, \dots, j_k^*$, and let $\zeta = 1/2$. Assume also that $\operatorname{aVar}(\hat{\boldsymbol{\delta}})$ is non-singular. Then*

$$\min_{\alpha(L)\colon \|\alpha(L)\| \leq M} \lim_{T\to\infty} P(\boldsymbol{\theta}_T \in \mathrm{CE}(\hat{\boldsymbol{\delta}})) = F_k\left(\chi^2_{1-a,k}; M^2\left[\lambda_{\max}(\operatorname{aVar}(\hat{\boldsymbol{\beta}}) \operatorname{aVar}(\hat{\boldsymbol{\delta}})^{-1}) - 1\right]\right),$$

where $F_k(x;c)$ is the cumulative distribution function, evaluated at point $x \geq 0$, of a non-central χ^2 distribution with k degrees of freedom and non-centrality parameter $c \geq 0$.

The worst-case coverage probability of the VAR confidence ellipsoid depends on three scalars: the bound M on misspecification, the dimension k of the ellipsoid, and the "multivariate relative standard error"

$$\sqrt{\lambda_{\min}(\operatorname{aVar}(\hat{\boldsymbol{\delta}}) \operatorname{aVar}(\hat{\boldsymbol{\beta}})^{-1})} = [\lambda_{\max}(\operatorname{aVar}(\hat{\boldsymbol{\beta}}) \operatorname{aVar}(\hat{\boldsymbol{\delta}})^{-1})]^{-1/2} = \min_{\varsigma \in \mathbb{R}^k} \sqrt{\operatorname{aVar}(\varsigma'\hat{\boldsymbol{\delta}}) / \operatorname{aVar}(\varsigma'\hat{\boldsymbol{\beta}})}.$$

Again, the worst-case coverage distortion is an increasing function of the *largest* efficiency gain for VAR over LP across *all* linear combinations of the impulse responses. Since VAR impulse response estimates are often highly correlated across horizons, this suggests that the VAR undercoverage can in fact be particularly severe in the multivariate case. Numerical calculations (available upon request from the authors) show that the coverage distortions can be severe even when $M = 1$, regardless of the dimension k.

C.3.4.4 Proof of Proposition C.3.2

Define $\tilde{\alpha}_\ell = D^{-1/2}\alpha_\ell D^{1/2}$ for all $\ell \geq 1$. Notice that $\|\alpha(L)\|^2 = \sum_{\ell=1}^{\infty} \|\tilde{\alpha}_\ell\|^2$.

By Proposition 3.3.2, we have $\text{aBias}(\hat{\delta}_{i^*,j^*,h}) = \sum_{\ell=1}^{\infty} \text{tr}(\Xi_{i^*,j^*,h,\ell}\tilde{\alpha}_\ell)$, where

$$\Xi_{i^*,j^*,h,\ell} \equiv D^{1/2}H'(A')^{\ell-1}S^{-1}\Psi_{i^*,j^*,h}HD^{1/2} - \mathbb{1}(\ell \leq h)D^{-1/2}e_{j^*,m}e_{i^*,n}'A^{h-\ell}HD^{1/2}.$$

Since $\text{tr}(\Xi_{i^*,j^*,h,\ell}\tilde{\alpha}_\ell) = \boldsymbol{e}(\Xi_{i^*,j^*,h,\ell})'\boldsymbol{e}(\tilde{\alpha}_\ell')$, we can write

$$\text{aBias}(\hat{\boldsymbol{\delta}}) = \sum_{\ell=1}^{\infty} \Upsilon_\ell \boldsymbol{e}(\tilde{\alpha}_\ell'),$$

where

$$\Upsilon_\ell \equiv \Big(\boldsymbol{e}(\Xi_{i_1^*,j_1^*,h_1,\ell}), \dots, \boldsymbol{e}(\Xi_{i_k^*,j_k^*,h_k,\ell})\Big)' \in \mathbb{R}^{k\times m^2}.$$

Hence,

$$\max_{\alpha(L)\colon \|\alpha(L)\|\leq M} \|R\,\text{aBias}(\hat{\boldsymbol{\delta}})\|^2 = \max_{\{\tilde{\alpha}_\ell\}_{\ell=1}^{\infty}\colon \sum_{\ell=1}^{\infty}\|\tilde{\alpha}_\ell'\|^2\leq M^2} \left\|\sum_{\ell=1}^{\infty} R\Upsilon_\ell \boldsymbol{e}(\tilde{\alpha}_\ell')\right\|^2.$$

Lemma C.3.1 below shows that the final expression above equals $M^2\lambda_{\max}(\sum_{\ell=1}^{\infty} R\Upsilon_\ell\Upsilon_\ell' R')$ (the lemma only explicitly considers the case $M = 1$, but the general case then follows from the homogeneity of degree 1 of the norm). Finally, Lemma C.3.2 below shows that $\sum_{\ell=1}^{\infty} \Upsilon_\ell\Upsilon_\ell' = \text{aVar}(\hat{\boldsymbol{\beta}}) - \text{aVar}(\hat{\boldsymbol{\delta}})$. This completes the proof of the proposition. The proof of Lemma C.3.1 shows that the maximum above is achieved when $\boldsymbol{e}(\tilde{\alpha}_\ell') \propto \Upsilon_\ell' v$ (with the constant of proportionality being independent of ℓ and chosen to satisfy the norm constraint), where v is the eigenvector corresponding to the largest eigenvalue of $R[\text{aVar}(\hat{\boldsymbol{\beta}}) - \text{aVar}(\hat{\boldsymbol{\delta}})]R'$. In the univariate case $k = 1$, this reduces to expression (3.4.1) in Section 3.4.2. $\square$

Lemma C.3.1. *Let $\mathcal{X}$ denote the set of sequences $\{x_\ell\}_{\ell=1}^{\infty}$ of $m \times m$ matrices x_ℓ satisfying $\sum_{\ell=1}^{\infty} \|x_\ell\|^2 \leq 1$. Let $\{L_\ell\}_{\ell=1}^{\infty}$ be a sequence of $r \times m^2$ matrices L_ℓ satisfying $\sum_{\ell=1}^{\infty} \|L_\ell\|^2 < \infty$. Then*

$$\max_{\{x_\ell\}_{\ell=1}^{\infty}\in\mathcal{X}} \left\|\sum_{\ell=1}^{\infty} L_\ell \boldsymbol{e}(x_\ell)\right\|^2 = \lambda_{\max}\left(\sum_{\ell=1}^{\infty} L_\ell L_\ell'\right). \tag{C.3.3}$$

Proof. A short proof using abstract functional analysis is available upon request from the authors. Below we provide a more elementary proof.

The statement of the lemma is obvious if $\sum_{\ell=1}^{\infty} \|L_\ell\|^2 = 0$, in which case both sides of the above display equal 0. Hence, we may assume that the series $V \equiv \sum_{\ell=1}^{\infty} L_\ell L_\ell'$ converges to a non-zero matrix. Let v be the unit-length eigenvector corresponding to the largest eigenvalue $\lambda \equiv \lambda_{\max}(V) \in (0, \infty)$ of V.

The purported maximum (C.3.3) is achieved by the sequence $\{x_\ell^*\}$ given by $\boldsymbol{e}(x_\ell^*) = \lambda^{-1/2} L_\ell' v$:

$$\left\| \sum_{\ell=1}^{\infty} L_\ell \boldsymbol{e}(x_\ell^*) \right\|^2 = \left\| \lambda^{-1/2} \sum_{\ell=1}^{\infty} L_\ell L_\ell' v \right\|^2 = \lambda^{-1} \|Vv\|^2 = \lambda^{-1} \|\lambda v\|^2 = \lambda \|v\|^2 = \lambda,$$

and

$$\sum_{\ell=1}^{\infty} \|x_\ell^*\|^2 = \sum_{\ell=1}^{\infty} \boldsymbol{e}(x_\ell^*)' \boldsymbol{e}(x_\ell^*) = \lambda^{-1} v' \sum_{\ell=1}^{\infty} L_\ell L_\ell' v = \lambda^{-1} v' V v = \lambda^{-1} \lambda = 1.$$

We complete the proof by showing that the left-hand side of (C.3.3) is bounded above by the right-hand side. Let K be an arbitrary positive integer. Then

$$\max_{\{x_\ell\}_{\ell=1}^{\infty} \in \mathcal{X}} \left\| \sum_{\ell=1}^{\infty} L_\ell \boldsymbol{e}(x_\ell) \right\| \leq \max_{\{x_\ell\}_{\ell=1}^{\infty} \in \mathcal{X}} \left\| \sum_{\ell=1}^{K} L_\ell \boldsymbol{e}(x_\ell) \right\| + \max_{\{x_\ell\}_{\ell=1}^{\infty} \in \mathcal{X}} \left\| \sum_{\ell=K+1}^{\infty} L_\ell \boldsymbol{e}(x_\ell) \right\|.$$

The second term on the right-hand side is bounded above by $(\sum_{\ell=K+1}^{\infty} \|L_\ell\|^2)^{1/2}$ by Cauchy-Schwarz. As for the first term, standard results for the eigenvalues of finite-dimensional matrices yield

$$\max_{\{x_\ell\}_{\ell=1}^{\infty} \in \mathcal{X}} \left\| \sum_{\ell=1}^{K} L_\ell \boldsymbol{e}(x_\ell) \right\|^2 = \max_{x \in \mathbb{R}^{Km^2} \colon \|x\| \leq 1} \left\| \begin{pmatrix} L_1 & L_2 & \cdots & L_K \end{pmatrix} x \right\|^2$$
$$= \lambda_{\max}\left(\begin{pmatrix} L_1 & \cdots & L_K \end{pmatrix}' \begin{pmatrix} L_1 & \cdots & L_K \end{pmatrix} \right) = \lambda_{\max}\left(\begin{pmatrix} L_1 & \cdots & L_K \end{pmatrix} \begin{pmatrix} L_1 & \cdots & L_K \end{pmatrix}' \right) = \lambda_{\max}\left(\sum_{\ell=1}^{K} L_\ell L_\ell' \right).$$

We have shown

$$\max_{\{x_\ell\}_{\ell=1}^\infty \in \mathcal{X}} \left\| \sum_{\ell=1}^{\infty} L_\ell \boldsymbol{e}(x_\ell) \right\| \leq \left(\lambda_{\max} \left(\sum_{\ell=1}^{K} L_\ell L_\ell' \right) \right)^{1/2} + \left(\sum_{\ell=K+1}^{\infty} \|L_\ell\|^2 \right)^{1/2}.$$

Now let $K \to \infty$. Since $\sum_{\ell=1}^\infty L_\ell L_\ell'$ is a convergent series, the first term on the right-hand side above converges to $\lambda^{1/2}$ by continuity of eigenvalues, while the second term converges to 0. This establishes the required bound. $\square$

Lemma C.3.2. *Under the assumptions of Proposition C.3.2, and using the notation in the proof of that proposition, we have*

$$\sum_{\ell=1}^{\infty} \Upsilon_\ell \Upsilon_\ell' = \mathrm{aVar}(\hat{\boldsymbol{\beta}}) - \mathrm{aVar}(\hat{\boldsymbol{\delta}}).$$

Proof. By definition of Υ_ℓ, it suffices to show that, for any indices $a, b \in \{1, \dots, k\}$,

$$\sum_{\ell=1}^{\infty} \boldsymbol{e}(\Xi_{i_a^*, j_a^*, h_a, \ell})' \boldsymbol{e}(\Xi_{i_b^*, j_b^*, h_b, \ell}) = \mathrm{aCov}(\hat{\beta}_{i_a^*, j_a^*, h_a}, \hat{\beta}_{i_b^*, j_b^*, h_b}) - \mathrm{aCov}(\hat{\delta}_{i_a^*, j_a^*, h_a}, \hat{\delta}_{i_b^*, j_b^*, h_b}). \tag{C.3.4}$$

Multiplying out terms, we find that the left-hand side above equals

$$\begin{aligned}\sum_{\ell=1}^{\infty} \mathrm{tr}(\Xi_{i_a^*, j_a^*, h_a, \ell}' \Xi_{i_b^*, j_b^*, h_b, \ell}) = &\sum_{\ell=1}^{\infty} \mathrm{tr}\left(A^{\ell-1} \Sigma (A')^{\ell-1} S^{-1} \Psi_{i_b^*, j_b^*, h_b} \Sigma \Psi_{i_a^*, j_a^*, h_a}' S^{-1} \right) \\ &- \sum_{\ell=1}^{h_b} \mathrm{tr}\left(A^{\ell-1} H_{\bullet, j_b^*} e_{i_b^*, n}' A^{h_b - \ell} \Sigma \Psi_{i_a^*, j_a^*, h_a}' S^{-1} \right) \\ &- \sum_{\ell=1}^{h_a} \mathrm{tr}\left(A^{\ell-1} H_{\bullet, j_a^*} e_{i_a^*, n}' A^{h_a - \ell} \Sigma \Psi_{i_b^*, j_b^*, h_b}' S^{-1} \right) \\ &+ \mathbb{1}(j_a^* = j_b^*) \sigma_{j_a^*}^{-2} \sum_{\ell=1}^{\min\{h_a, h_b\}} e_{i_b^*, n}' A^{h_b - \ell} \Sigma (A')^{h_a - \ell} e_{i_a^*, n}.\end{aligned}$$

We now evaluate each of the four terms on the right-hand side above. The *first* term equals

$$\mathrm{tr}\left(\underbrace{\sum_{\ell=1}^{\infty} A^{\ell-1} \Sigma (A')^{\ell-1}}_{=S} S^{-1} \Psi_{i_b^*, j_b^*, h_b} \Sigma \Psi_{i_a^*, j_a^*, h_a}' S^{-1} \right) = \mathrm{tr}\left(\Psi_{i_b^*, j_b^*, h_b} \Sigma \Psi_{i_a^*, j_a^*, h_a}' S^{-1} \right).$$

The *second* term (in the earlier display) equals

$$-\operatorname{tr}\left(\underbrace{\sum_{\ell=1}^{h_b} A^{\ell-1} H_{\bullet,j_b^*} e'_{i_b^*,n} A^{h_b-\ell}}_{=\sum_{\ell=1}^{h_b} A^{h_b-\ell} H_{\bullet,j_b^*} e'_{i_b^*,n} A^{\ell-1}=\Psi_{i_b^*,j_b^*,h_b}} \Sigma\Psi'_{i_a^*,j_a^*,h_a} S^{-1}\right) = -\operatorname{tr}\left(\Psi_{i_b^*,j_b^*,h_b}\Sigma\Psi'_{i_a^*,j_a^*,h_a} S^{-1}\right),$$

and the *third* term (in the earlier display) also equals this quantity by a symmetric calculation. In conclusion, we have shown

$$\begin{aligned}
&\sum_{\ell=1}^{\infty} \operatorname{tr}(\Xi'_{i_a^*,j_a^*,h_a,\ell}\Xi_{i_b^*,j_b^*,h_b,\ell}) \\
&= \mathbb{1}(j_a^* = j_b^*)\sigma_{j_a^*}^{-2} \sum_{\ell=1}^{\min\{h_a,h_b\}} e'_{i_b^*,n} A^{h_b-\ell}\Sigma(A')^{h_a-\ell} e_{i_a^*,n} - \operatorname{tr}\left(\Psi_{i_b^*,j_b^*,h_b}\Sigma\Psi'_{i_a^*,j_a^*,h_a} S^{-1}\right).
\end{aligned}$$

The desired result (C.3.4) now follows from Corollary C.1.2. □

C.3.4.5 Proof of Corollary C.3.1

The result follows straightforwardly from (C.3.2) if we can show that the maximal non-centrality parameter equals

$$\max_{\alpha(L)\colon \|\alpha(L)\|\le M} \operatorname{aBias}(\hat{\boldsymbol{\delta}})' \operatorname{aVar}(\hat{\boldsymbol{\delta}})^{-1} \operatorname{aBias}(\hat{\boldsymbol{\delta}}) = M^2\left[\lambda_{\max}(\operatorname{aVar}(\hat{\boldsymbol{\beta}}) \operatorname{aVar}(\hat{\boldsymbol{\delta}})^{-1}) - 1\right].$$

But this follows from applying Proposition C.3.2 with $R = \operatorname{aVar}(\hat{\boldsymbol{\delta}})^{-1/2}$, since

$$\begin{aligned}
&\lambda_{\max}\left(\operatorname{aVar}(\hat{\boldsymbol{\delta}})^{-1/2}[\operatorname{aVar}(\hat{\boldsymbol{\beta}}) - \operatorname{aVar}(\hat{\boldsymbol{\delta}})] \operatorname{aVar}(\hat{\boldsymbol{\delta}})^{-1/2\prime}\right) \\
&= \lambda_{\max}\left(\operatorname{aVar}(\hat{\boldsymbol{\delta}})^{-1/2} \operatorname{aVar}(\hat{\boldsymbol{\beta}}) \operatorname{aVar}(\hat{\boldsymbol{\delta}})^{-1/2\prime} - I_k\right) \\
&= \lambda_{\max}\left(\operatorname{aVar}(\hat{\boldsymbol{\beta}}) \operatorname{aVar}(\hat{\boldsymbol{\delta}})^{-1}\right) - 1. \quad \square
\end{aligned}$$

C.4 Simulation details and further results

We here provide supplementary details for the simulation study reported in Section 3.5.2.

C.4.1 Implementation details

All inference procedures (correctly) assume homoskedastic shocks. The VAR is estimated with an intercept, and confidence intervals are constructed either using the delta method or the recursive residual bootstrap; following the recommendation of Inoue and Kilian (2020), we report the Efron bootstrap confidence interval. For LPs, we include the shock measure as the only contemporaneous regressor, control for an intercept and the same p lags of all observables as in the VAR, and report the OLS coefficient on the shock. Confidence intervals are constructed either using homoskedastic OLS standard errors or by bootstrapping an auxiliary VAR as in Montiel Olea and Plagborg-Møller (2021), but using a recursive residual bootstrap instead of a wild bootstrap; we follow the latter paper and report the percentile-t bootstrap confidence interval. We use 1,000 bootstrap draws, and the maximal lag length considered for the AIC is 24.

C.4.2 Further results

Figure C.4.1 shows coverage probabilities and median confidence interval lengths for longer estimation lag lengths $p \in \{15, 18\}$. The results are consistent with the asymptotic theory: the longer the estimation lag length, the less severe VAR undercoverage, with correct coverage ensured through equivalence with LP. Of course, since the true DGP is a VAR(18), the $p = 18$ VAR estimator is more efficient than LP at longer horizons (middle panel). The bottom panel of the figure shows the root MSE for VAR and LP, with lag length selected by AIC. By this measure, VAR outperforms LP, indicating that while the VAR bias is large enough to seriously compromise inference (as shown earlier), it is not so large as to threaten the VAR estimator's status as a useful point estimator.

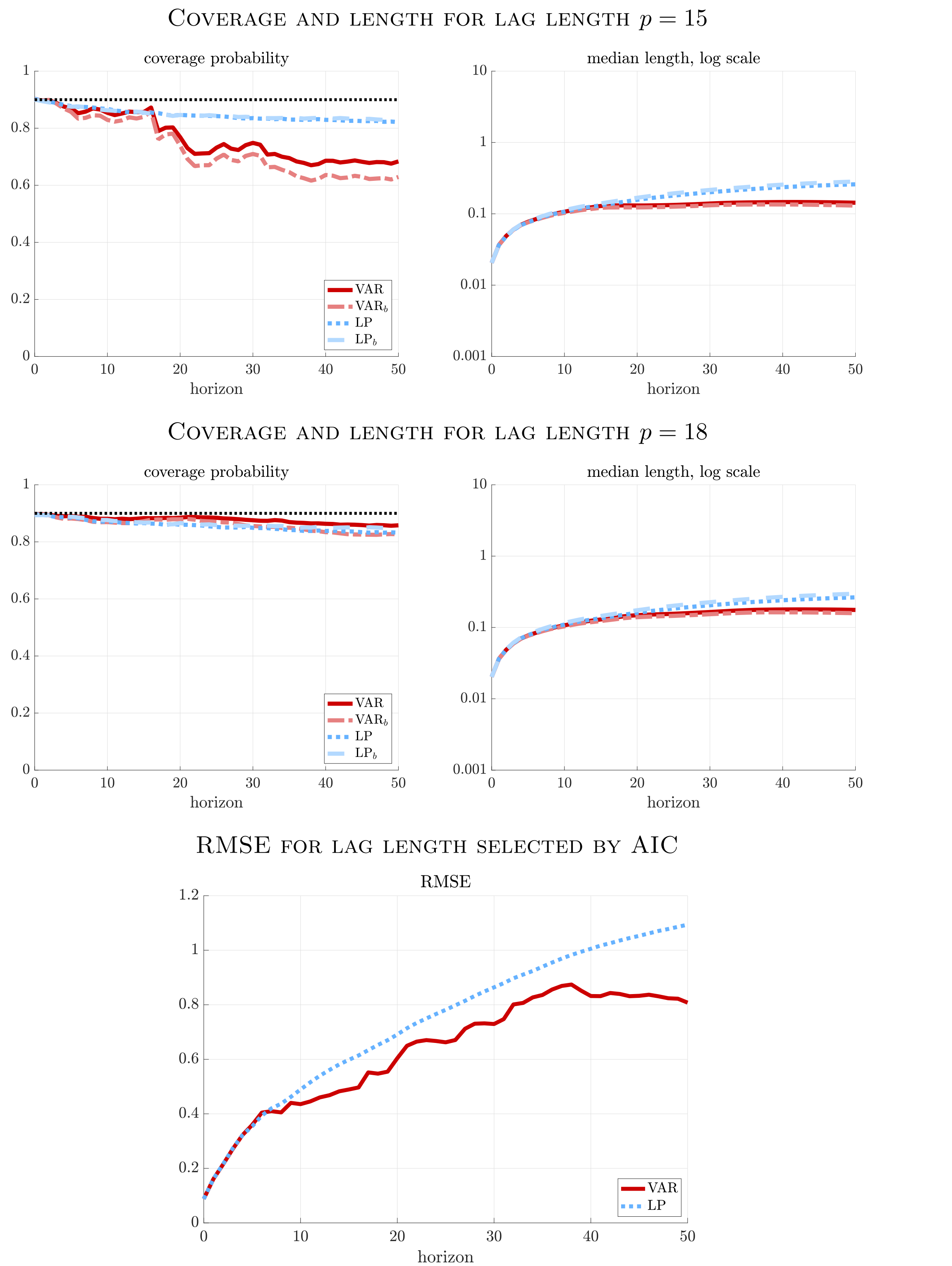

Figure C.4.1: Top two panels: coverage probability (left) and median length (right) for VAR (red) and LP (blue) 90% confidence intervals computed via the delta method or bootstrap (the latter are indicated with subscript "b" in the legends). Bottom panel: root MSE of estimators. Lag length: $p = 15$ in the top panel, $p = 18$ in the middle panel, and p selected by AIC in the bottom panel.

C.5 Further proofs

We impose Assumption C.3.1 and Assumption 3.2.1(ii)–(v) throughout; as discussed in Section C.3.1, none of the proofs below require Assumption 3.2.1(i). Let $\|B\|$ denote the Frobenius norm of any matrix B. It is well known that this norm is sub-multiplicative: $\|BC\| \leq \|B\| \cdot \|C\|$. Let I_n denote the $n \times n$ identity matrix, $0_{m\times n}$ the $m \times n$ matrix of zeros, and $e_{i,n}$ the n-dimensional unit vector with a 1 in the i-th position. Recall from Assumption 3.2.1 the definitions $D \equiv \operatorname{var}(\varepsilon_t) = \operatorname{diag}(\sigma_1^2, \dots, \sigma_m^2)$, $\tilde{y}_t \equiv (I_n - AL)^{-1} H\varepsilon_t = \sum_{s=0}^{\infty} A^s H \varepsilon_{t-s}$, and $S \equiv \operatorname{var}(\tilde{y}_t)$.

C.5.1 Main lemmas

Lemma C.5.1. *For any $i^* \in \{1, \dots, n\}$ and $j^* \in \{1, \dots, m\}$, we have*

$$y_{i^*,t+h} = \theta_{h,T}\varepsilon_{j^*,t} + \underline{B}'_{h,i^*,j^*}\underline{y}_{j^*,t} + B'_{h,i^*,j^*} y_{t-1} + \xi_{i^*,h,t} + T^{-\zeta}\Theta_h(L)\varepsilon_t,$$

where

$$\begin{aligned}
\theta_{h,T} &\equiv e'_{i^*,n}(A^h H + T^{-\zeta}\sum_{\ell=1}^{h} A^{h-\ell} H \alpha_\ell) e_{j^*,m}, \\
\underline{B}'_{h,i^*,j^*} &\equiv e'_{i^*,n} A^h \underline{H}_{j^*} H_{11}^{-1}, \\
B'_{h,i^*,j^*} &\equiv e'_{i^*,n}\left[A^{h+1} - A^h \underline{H}_{j^*} H_{11}^{-1} \underline{I}_{j^*} A\right], \\
\xi_{i^*,h,t} &\equiv e'_{i^*,n} A^h \overline{H}_{j^*} \overline{\varepsilon}_{j^*,t} + \sum_{\ell=1}^{h} e'_{i^*,n} A^{h-\ell} H \varepsilon_{t+\ell},
\end{aligned}$$

and $\Theta_h(L) = \sum_{\ell=-\infty}^{\infty} \Theta_{h,\ell} L^\ell$ is an absolutely summable, $1 \times n$ two-sided lag polynomial with the j^*-th element of $\Theta_{h,0}$ equal to zero. Moreover,

$$T^{-1}\sum_{t=1}^{T-h} (\Theta_h(L)\varepsilon_t)\varepsilon_{j^*,t} = O_p(T^{-1/2}).$$

Proof. Iteration on the model in Equation (3.2.1) yields

$$y_{t+h} = A^{h+1} y_{t-1} + \sum_{\ell=0}^{h} A^{h-\ell}(H\varepsilon_{t+\ell} + T^{-\zeta} H\alpha(L)\varepsilon_{t+\ell}). \tag{C.5.1}$$

As in Section 3.2.2, let $\underline{y}_{j^*,t} \equiv (y_{1,t}, \dots, y_{j^*-1,t})'$ denote the variables ordered before $y_{j^*,t}$ (if any). Analogously, let $\overline{y}_{j^*,t} \equiv (y_{j^*+1,t}, \dots y_{n,t})'$ denote the variables ordered after $y_{j^*,t}$.

Using Assumption 3.2.1(iii), partition

$$H = (\underline{H}_{j^*}, H_{\bullet,j^*}, \overline{H}_{j^*}) = \begin{pmatrix} H_{11} & 0 & 0 \\ H_{21} & H_{22} & 0 \\ H_{31} & H_{32} & H_{33} \end{pmatrix}$$

conformably with the vector $y_t = (\underline{y}'_{j^*,t}, y_{j^*,t}, \overline{y}'_{j^*,t})'$. Let $\underline{I}_{j^*}$ denote the first $j^* - 1$ rows of the $n \times n$ identity matrix. Using the definition of y_t in Equation (3.2.1),

$$\underline{y}_{j^*,t} = \underline{I}_{j^*} A y_{t-1} + H_{11} \underline{\varepsilon}_{j^*,t} + T^{-\zeta} H_{11} \underline{I}_{j^*} \alpha(L) \varepsilon_t,$$

where $\underline{\varepsilon}_{j^*,t} = \underline{I}_{j^*} \varepsilon_t$. Using the previous equation to solve for $\underline{\varepsilon}_{j^*,t}$ we get

$$\underline{\varepsilon}_{j^*,t} = H_{11}^{-1}(\underline{y}_{j^*,t} - \underline{I}_{j^*} A y_{t-1} - T^{-\zeta} H_{11} \underline{I}_{j^*} \alpha(L) \varepsilon_t). \tag{C.5.2}$$

Expanding the terms in (C.5.1) we get:

$$\begin{aligned}
y_{t+h} &= A^{h+1}y_{t-1} + A^hH\varepsilon_t + T^{-\zeta}A^hH\alpha(L)\varepsilon_t + \sum_{\ell=1}^{h} A^{h-\ell}(H\varepsilon_{t+\ell} + T^{-\zeta}H\alpha(L)\varepsilon_{t+\ell}) \\
&= A^{h+1}y_{t-1} + \left(A^h\underline{H}_{j^*}\underline{\varepsilon}_{j^*,t} + A^hH_{\bullet,j^*}\varepsilon_{j^*,t} + A^h\overline{H}_{j^*}\overline{\varepsilon}_{j^*,t}\right) \\
&\quad + T^{-\zeta}A^hH\alpha(L)\varepsilon_t + \sum_{\ell=1}^{h} A^{h-\ell}(H\varepsilon_{t+\ell} + T^{-\zeta}H\alpha(L)\varepsilon_{t+\ell}) \\
&= A^{h+1}y_{t-1} + A^h\underline{H}_{j^*}H_{11}^{-1}(\underline{y}_{j^*,t} - \underline{I}_{j^*}Ay_{t-1} - T^{-\zeta}H_{11}\underline{I}_{j^*}\alpha(L)\varepsilon_t) + A^hH_{\bullet,j^*}\varepsilon_{j^*,t} + A^h\overline{H}_{j^*}\overline{\varepsilon}_{j^*,t} \\
&\quad + T^{-\zeta}A^hH\alpha(L)\varepsilon_t + \sum_{\ell=1}^{h} A^{h-\ell}(H\varepsilon_{t+\ell} + T^{-\zeta}H\alpha(L)\varepsilon_{t+\ell}),
\end{aligned}$$

where the last equality follows from substituting (C.5.2). Re-arranging terms we get

$$\begin{aligned}
y_{i^*,t+h} &= \left(e'_{i^*,n}A^hH_{\bullet,j^*}\right)\varepsilon_{j^*,t} + \underbrace{\left(e'_{i^*,n}A^h\underline{H}_{j^*}H_{11}^{-1}\right)}_{\equiv\underline{B}'_{h,i^*,j^*}}\underline{y}_{j^*,t} + \underbrace{\left(e'_{i^*,n}\left[A^{h+1} - A^h\underline{H}_{j^*}H_{11}^{-1}\underline{I}_{j^*}A\right]\right)}_{\equiv B'_{h,i^*,j^*}}y_{t-1} \\
&\quad + \underbrace{e'_{i^*,n}\left(A^h\overline{H}_{j^*}\overline{\varepsilon}_{j^*,t} + \sum_{\ell=1}^{h} A^{h-\ell}H\varepsilon_{t+\ell}\right)}_{=\xi_{i^*,h,t}} \\
&\quad + T^{-\zeta}e'_{i^*,n}\left(-A^h\underline{H}_{j^*}H_{11}^{-1}H_{11}\underline{I}_{j^*}\alpha(L)\varepsilon_t + \sum_{\ell=0}^{h} A^{h-\ell}H\alpha(L)\varepsilon_{t+\ell}\right).
\end{aligned}$$

Using the definition of $\theta_{h,T} \equiv e'_{i^*,n}(A^hH+T^{-\zeta}\sum_{\ell=1}^{h} A^{h-\ell}H\alpha_\ell)e_{j^*,m}$ and adding and subtracting $e'_{i^*,n}\left(T^{-\zeta}\sum_{\ell=1}^{h} A^{h-\ell}H\alpha_\ell\right)e_{j^*,m}\varepsilon_{j^*,t}$ in the display above, we obtain a representation of the form

$$y_{i^*,t+h} = \theta_{h,T}\varepsilon_{j^*,t} + \underline{B}'_{h,i^*,j^*}\underline{y}_{j^*,t} + B'_{h,i^*,j^*}y_{t-1} + \xi_{i^*,h,t} + T^{-\zeta}\tilde{u}_t, \tag{C.5.3}$$

where

$$\tilde{u}_t \equiv e'_{i^*,n}\left(-A^h\underline{H}_{j^*}\underline{I}_{j^*}\alpha(L)\varepsilon_t + \sum_{\ell=0}^{h} A^{h-\ell}H\alpha(L)\varepsilon_{t+\ell} - \left(\sum_{\ell=1}^{h} A^{h-\ell}H\alpha_\ell e_{j^*,m}e'_{j^*,m}\right)\varepsilon_t\right). \tag{C.5.4}$$

Algebra shows that $\tilde{u}_t$ can be written as a two-sided lag polynomial, $\Theta_h(L) = \sum_{\ell=-\infty}^{\infty} \Theta_{h,\ell} L^\ell$, with coefficients of dimension $1 \times n$ given by the following formulae:

$$\Theta_{h,\ell} = \begin{cases} -e'_{i^*,n} A^h \underline{H}_{j^*} \underline{I}_{j^*} \alpha_\ell + \sum_{s=0}^{h} e'_{i^*,n} A^{h-s} H \alpha_{\ell+s} & \text{for } \ell \geq 1, \\ \sum_{s=1}^{h} e'_{i^*,n} A^{h-s} H \alpha_s - \sum_{s=1}^{h} e'_{i^*,n} A^{h-s} H \alpha_s e_{j^*,m} e'_{j^*,m} & \text{for } \ell = 0, \\ \sum_{s=1}^{h+\ell} e'_{i^*,n} A^{h-s+\ell} H \alpha_s & \text{for } \ell \in \{-(h-1), \ldots, -1\}, \\ 0_{1\times n} & \text{for } \ell \leq -h. \end{cases}$$

In particular, $\Theta_{h,0,j^*} \equiv \Theta_{h,0} e_{j^*,m} = 0$.

We next show that $\Theta_h(L)$ is absolutely summable, that is $\sum_{\ell=-\infty}^{\infty} \|\Theta_{h,l}\| < \infty$. To do this, it suffices to show that $\sum_{\ell=1}^{\infty} \|\Theta_{h,l}\| < \infty$, since all the coefficients with index $\ell \leq -h$ are 0. Note that, by definition,

$$\sum_{\ell=1}^{\infty} \|\Theta_{h,\ell}\| \leq \|A^h\| \|\underline{H}_{j^*} \underline{I}_{j^*}\| \sum_{\ell=1}^{\infty} \|\alpha_\ell\| + \|H\| \sum_{\ell=1}^{\infty} \sum_{s=0}^{h} \|A^{h-s}\| \|\alpha_{\ell+s}\|.$$

Let $\lambda \in [0, 1)$ and $C > 0$ be chosen such that $\|A^\ell\| \leq C\lambda^\ell$ for all $\ell \geq 0$ (such constants exists by Assumption 3.2.1(ii)). Then

$$\sum_{\ell=1}^{\infty} \sum_{s=0}^{h} \|A^{h-s}\| \|\alpha_{\ell+s}\| \leq C \sum_{\ell=1}^{\infty} \sum_{s=1}^{h} \lambda^{h-s} \|\alpha_{\ell+s}\| \leq C \sum_{\ell=1}^{\infty} \sum_{s=1}^{h} \|\alpha_{\ell+s}\| \leq Ch \sum_{\ell=1}^{\infty} \|\alpha_\ell\| < \infty,$$

where the last inequality holds because the coefficients of $\alpha(L)$ are summable. We thus conclude that

$$y_{i^*,t+h} = \theta_{h,T} \varepsilon_{j^*,t} + \underline{B}'_{h,y} \underline{y}_{j^*,t} + B'_{h,y} y_{t-1} + \xi_{i^*,h,t} + T^{-\zeta} \Theta_h(L) \varepsilon_t,$$

where $\Theta_h(L)$ is a two-sided lag-polynomial with summable coefficients.

Finally, we show that

$$T^{-1}\sum_{t=1}^{T-h}(\Theta_h(L)\varepsilon_t)\varepsilon_{j^*,t} = O_p(T^{-1/2}). \tag{C.5.5}$$

We can decompose $(\Theta_h(L)\varepsilon_t)\varepsilon_{j^*,t}$ into finitely many terms of the form $(\psi_j(L)\varepsilon_{j,t})\varepsilon_{j^*,t}$ for some two-sided, absolutely summable lag polynomial $\psi_j(L)$ and integer j. Since $\Theta_{h,0,j^*}=0$, each of these terms has mean zero by shock orthogonality. By Lemma C.5.11, each term also has absolutely summable autocovariances. Hence, Brockwell and Davis (1991, Thm. 7.1.1) and Chebyshev's inequality imply that the sample average of each term is $O_p(T^{-1/2})$. □

Lemma C.5.2.

$$\hat{A}-A = T^{-\zeta}H\sum_{\ell=1}^{\infty}\alpha_\ell DH'(A')^{\ell-1}S^{-1} + T^{-1}\sum_{t=1}^{T}H\varepsilon_t\tilde{y}_{t-1}'S^{-1} + o_p(T^{-\zeta}).$$

In particular, $\hat{A}-A = O_p(T^{-\zeta}+T^{-1/2})$.

Proof. Since,

$$\hat{A}-A = \left(T^{-1}\sum_{t=1}^{T-h}u_t y_{t-1}'\right)\left(T^{-1}\sum_{t=1}^{T-h}y_{t-1}y_{t-1}'\right)^{-1},$$

the result follows from Lemmas C.5.7 and C.5.8. □

Lemma C.5.3.

$$\hat{\nu} - H_{\bullet,j^*} = \frac{1}{\sigma_{j^*}^2}T^{-1}\sum_{t=1}^{T}\xi_{0,t}\varepsilon_{j^*,t} + O_p(T^{-2\zeta}) + o_p(T^{-1/2}).$$

Proof. By Lemma C.5.5, $\hat{\nu} = (0_{1\times(j^*-1)}, 1, \hat{\bar{\nu}}')$, where the j-th element of $\hat{\bar{\nu}}$ equals the on-impact local projection of $y_{i^*+j,t}$ on $y_{j^*,t}$, controlling for $\underline{y}_{j^*,t}$ and y_{t-1}. The statement of the lemma is therefore a direct consequence of Proposition 3.3.1 and the fact that (by definition) $\xi_{0,i,t}=0$ for $i\le j^*$. □

Lemma C.5.4. *Fix $h \geq 0$. Consider the regression of $y_{j^*,t}$ on $q_{j^*,t} \equiv (\underline{y}'_{j^*,t}, y'_{t-1})'$, using the observations $t = 1, 2, \dots, T - h$:*

$$y_{j^*,t} = \hat{\vartheta}'_h q_{j^*,t} + \hat{x}_{h,t}.$$

Note that the residuals $\hat{x}_{h,t}$ are consistent with the earlier definition in the proof of Proposition 3.3.1. Let $\underline{\lambda}'_{j^*}$ be the row vector containing the first $j^* - 1$ elements of the last row of $-\tilde{H}^{-1}$ (where $\tilde{H}$ is defined in Assumption 3.2.1(iii)). Let $\lambda'_{j^*} \equiv (-\underline{\lambda}'_{j^*}, 1, 0_{1\times(n-j^*)})$ and $\vartheta \equiv (\underline{\lambda}'_{j^*}, (\lambda'_{j^*}A))'$. Then:

i) $\hat{\vartheta}_h - \vartheta = O_p(T^{-\zeta} + T^{-1/2})$.

ii) *For $j \geq j^*$,* $T^{-1}\sum_{t=1}^{T-h}(\hat{x}_{h,t} - \varepsilon_{j^*,t})\varepsilon_{j,t} = O_p(T^{-2\zeta}) + o_p(T^{-1/2})$.

iii) *For $\ell \geq 1$,* $T^{-1}\sum_{t=1}^{T-h}(\hat{x}_{h,t} - \varepsilon_{j^*,t})\varepsilon_{t+\ell} = O_p(T^{-2\zeta}) + o_p(T^{-1/2})$.

iv) $T^{-1}\sum_{t=1}^{T-h}(\hat{x}_{h,t} - \varepsilon_{j^*,t})\hat{x}_{h,t} = O_p(T^{-2\zeta}) + o_p(T^{-1/2})$.

v) $T^{-1}\sum_{t=1}^{T-h}\hat{x}_{h,t}^2 \xrightarrow{p} \sigma_{j^*}^2$.

vi) *For any absolutely summable two-sided lag polynomial $B(L)$,* $T^{-1}\sum_{t=1}^{T-h}(\hat{x}_{h,t} - \varepsilon_{j^*,t})B(L)\varepsilon_t = O_p(T^{-\zeta} + T^{-1/2})$.

Proof. By Equation (3.2.1), the outcome variables in the model satisfy

$$y_t = Ay_{t-1} + H[I_m + T^{-\zeta}\alpha(L)]\varepsilon_t,\ t = 1, 2, \dots, T.$$

By Assumption 3.2.1(iii), the first j^* rows of the matrix H above are of the form $(\tilde{H}, 0_{j^*\times(j^*-m)})$, where m is the number of shocks and $\tilde{H}$ is a $j^* \times j^*$ lower triangular matrix with 1's on the diagonal.

We can premultiply the first j^* equations of (2.1) by $\tilde{H}^{-1}$ to obtain:

$$[\tilde{H}^{-1}, 0_{j^*\times(n-j^*)}]y_t = [\tilde{H}^{-1}, 0_{j^*\times(n-j^*)}]Ay_{t-1} + [I_{j^*}, 0_{j^*\times(m-j^*)}][I_m + T^{-\zeta}\alpha(L)]\varepsilon_t.$$

By definition, $-\underline{\lambda}'_{j^*}$ is the row vector containing the first j^*-1 elements of the last row of $\tilde{H}^{-1}$ and $\lambda'_{j^*} \equiv (-\underline{\lambda}'_{j^*}, 1, 0_{1\times(n-j^*)})$. Thus, we can re-write the j^*-th equation above as

$$[-\underline{\lambda}'_{j^*}, 1, 0_{j^*\times(n-j^*)}]y_t = \lambda'_{j^*}Ay_{t-1} + \varepsilon_{j^*,t} + T^{-\zeta}\alpha_{j^*}(L)\varepsilon_t,$$

where $\alpha_{j^*}(L)$ is the j^*-th row of $\alpha(L)$. Re-arranging terms we get

$$y_{j^*,t} = \vartheta' q_{j^*,t} + \varepsilon_{j^*,t} + T^{-\zeta}\alpha_{j^*}(L)\varepsilon_t,$$

where $\vartheta \equiv (\underline{\lambda}'_{j^*}, (\lambda'_{j^*}A))'$ and $q_{j^*,t} \equiv (\underline{y}'_{j^*,t}, y'_{t-1})'$. In a slight abuse of notation, and for notational simplicity, we henceforth replace $q_{j^*,t}$ by q_t.

Statement (i)) follows from standard OLS algebra if we can show that a) $T^{-1}\sum_{t=1}^{T-h} q_t\varepsilon_{j^*,t} = O_p(T^{-\zeta}+T^{-1/2})$, b) $(T^{-1}\sum_{t=1}^{T-h} q_t q_t')^{-1} = O_p(1)$, and c) $T^{-\zeta-1}\sum_{t=1}^{T-h} q_t(\alpha_{j^*}(L)\varepsilon_t) = O_p(T^{-\zeta})$. Lemma C.5.9 establishes these results.

The proofs of statements (ii)) and (iii)) are similar, so we focus on the latter. By definition of $\hat{x}_{h,t}$, we have $\hat{x}_{h,t} - \varepsilon_{j^*,t} = (\vartheta - \hat{\vartheta}_h)'q_t + T^{-\zeta}\alpha_{j^*}(L)\varepsilon_t$. Let $\tilde{q}_t \equiv (\underline{\tilde{y}}'_{j^*,t}, \tilde{y}'_{t-1})'$ and $\Delta_t \equiv q_t - \tilde{q}_t$. Then

$$T^{-1}\sum_{t=1}^{T-h}(\hat{x}_{h,t} - \varepsilon_{j^*,t})\varepsilon_{t+\ell} = (\vartheta - \hat{\vartheta}_h)'\left(\frac{1}{T}\sum_{t=1}^{T-h}\Delta_t\varepsilon_{t+\ell}\right) \tag{C.5.6}$$

$$+ (\vartheta - \hat{\vartheta}_h)'\left(\frac{1}{T}\sum_{t=1}^{T-h}\tilde{q}_t\varepsilon_{t+\ell}\right) \tag{C.5.7}$$

$$+ \frac{1}{T^{\zeta}}\left(\frac{1}{T}\sum_{t=1}^{T-h}\left(\alpha_{j^*}(L)\varepsilon_t\right)\varepsilon_{t+\ell}\right). \tag{C.5.8}$$

By (i)), $(\vartheta - \hat{\vartheta}_h) = O_p(T^{-\zeta}+T^{-1/2})$. Lemma C.5.6, Assumption C.3.1, and Cauchy-Schwarz imply that the sample average in parenthesis in (C.5.6) is $O_p\left(T^{-\zeta}\right)$. The two sample averages in parentheses in lines (C.5.7)–(C.5.8) have mean zero, since the shocks are white noise and mutually orthogonal, so Lemma C.5.11 and Brockwell and Davis (1991, Thm. 7.1.1) imply that they are each $O_p(T^{-1/2})$. It follows that

$$T^{-1}\sum_{t=1}^{T-h}(\hat{x}_{h,t}-\varepsilon_{j^*,t})\varepsilon_{t+\ell}=O_p(T^{-2\zeta})+o_p(T^{-1/2}).$$

For statement (iv)), note that

$$T^{-1}\sum_{t=1}^{T-h}(\hat{x}_{h,t}-\varepsilon_{j^*,t})\hat{x}_{h,t}=T^{-1}\sum_{t=1}^{T-h}(\hat{x}_{h,t}-\varepsilon_{j^*,t})^2+T^{-1}\sum_{t=1}^{T-h}(\hat{x}_{h,t}-\varepsilon_{j^*,t})\varepsilon_{j^*,t}.$$

Lemma C.5.10 shows that $T^{-1}\sum_{t=1}^{T-h}(\hat{x}_{h,t}-\varepsilon_{j^*,t})^2=O_p(T^{-2\zeta})+o_p(T^{-1/2})$. This result, combined with (ii)), implies that statement (iv)) holds.

For statement (v)), note that

$$\begin{aligned}T^{-1}\sum_{t=1}^{T-h}(\hat{x}_{h,t})^2&=T^{-1}\sum_{t=1}^{T-h}(\hat{x}_{h,t}-\varepsilon_{j^*,t}+\varepsilon_{j^*,t})^2\\&=T^{-1}\sum_{t=1}^{T-h}(\hat{x}_{h,t}-\varepsilon_{j^*,t})^2-2T^{-1}\sum_{t=1}^{T-h}(\hat{x}_{h,t}-\varepsilon_{j^*,t})\varepsilon_{j^*,t}+T^{-1}\sum_{t=1}^{T-h}\varepsilon_{j^*,t}^2.\end{aligned}$$

Lemma C.5.10 and statement (ii)) imply that the first two terms converge in probability to zero. Since $T^{-1}\sum_{t=1}^{T-h}\varepsilon_{j^*,t}^2\xrightarrow{p}\sigma_{j^*}^2$ by Lemma C.5.11 and Brockwell and Davis (1991, Thm. 7.1.1), statement (v)) holds.

Finally, statement (vi)) obtains by decomposing

$$\begin{aligned}T^{-1}\sum_{t=1}^{T-h}B(L)\varepsilon_t(\hat{x}_{h,t}-\varepsilon_{j^*,t})&=T^{-1}\sum_{t=1}^{T-h}B(L)\varepsilon_t q_t'(\vartheta-\hat{\vartheta}_h)+T^{-\zeta}T^{-1}\sum_{t=1}^{T-h}B(L)\varepsilon_t[\alpha_{j^*}(L)\varepsilon_t]'\\&=O_p(1)\times O_p(T^{-\zeta}+T^{-1/2})+T^{-\zeta}\times O_p(1),\end{aligned}$$

where the last line follows from statement (i)), Lemma C.5.6, and Lemma C.5.11. □

C.5.2 Auxiliary numerical lemma

Lemma C.5.5. *Define* $\overline{y}_{i,t} \equiv (y_{i+1,t}, y_{i+2,t}, \dots, y_{nt})'$ *to be the (possibly empty) vector of variables that are ordered after* $y_{i,t}$ *in* y_t. *Partition*

$$\hat{\Sigma} = \begin{pmatrix} \hat{\Sigma}_{11} & \hat{\Sigma}_{12} & \hat{\Sigma}_{13} \\ \hat{\Sigma}_{21} & \hat{\Sigma}_{22} & \hat{\Sigma}_{23} \\ \hat{\Sigma}_{31} & \hat{\Sigma}_{32} & \hat{\Sigma}_{33} \end{pmatrix}, \quad \hat{C} = \begin{pmatrix} \hat{C}_{11} & 0 & 0 \\ \hat{C}_{21} & \hat{C}_{22} & 0 \\ \hat{C}_{31} & \hat{C}_{32} & \hat{C}_{33} \end{pmatrix},$$

conformably with $y_t = (\underline{y}'_{j^*,t}, y_{j^*,t}, \overline{y}'_{j^*,t})'$, *where* $\hat{\Sigma} = \hat{C}\hat{C}'$ *(in particular,* $\hat{C}_{22} = \hat{C}_{j^*,j^*}$*). Then*

$$(\hat{\Sigma}_{31}, \hat{\Sigma}_{32}) \begin{pmatrix} \hat{\Sigma}_{11} & \hat{\Sigma}_{12} \\ \hat{\Sigma}_{21} & \hat{\Sigma}_{22} \end{pmatrix}^{-1} e_{j^*,j^*} = \hat{C}_{22}^{-1}\hat{C}_{32}. \tag{C.5.9}$$

Note that the lemma implies $\hat{\beta}_0 = \hat{\delta}_0$: If $i^* < j^*$ or $i^* = j^*$, then both estimators equal 0 or 1 (by definition), respectively; if $i^* > j^*$, then $\hat{\beta}_0$ is defined as the $i^* - j^*$ element of the left-hand side of (C.5.9) (by Frisch-Waugh), while $\hat{\delta}_0$ is defined as the $i^* - j^*$ element of the right-hand side of (C.5.9).

Proof. From the relationship $\hat{\Sigma} = \hat{C}\hat{C}'$, we get

$$\begin{pmatrix} \hat{\Sigma}_{11} & \hat{\Sigma}_{12} \\ \hat{\Sigma}_{21} & \hat{\Sigma}_{22} \\ \hat{\Sigma}_{31} & \hat{\Sigma}_{32} \end{pmatrix} = \begin{pmatrix} \hat{C}_{11}\hat{C}'_{11} & \hat{C}_{11}\hat{C}'_{21} \\ \hat{C}_{21}\hat{C}'_{11} & \hat{C}_{21}\hat{C}'_{21} + \hat{C}_{22}^2 \\ \hat{C}_{31}\hat{C}'_{11} & \hat{C}_{31}\hat{C}'_{21} + \hat{C}_{32}\hat{C}_{22} \end{pmatrix}.$$

The partitioned inverse formula implies

$$\begin{pmatrix} \hat{\Sigma}_{11} & \hat{\Sigma}_{12} \\ \hat{\Sigma}_{21} & \hat{\Sigma}_{22} \end{pmatrix}^{-1} e_{j^*,j^*} = \frac{1}{\hat{C}_{21}\hat{C}_{21}' + \hat{C}_{22}^2 - \hat{C}_{21}\hat{C}_{11}'(\hat{C}_{11}\hat{C}_{11}')^{-1}\hat{C}_{11}\hat{C}_{21}'} \begin{pmatrix} -(\hat{C}_{11}\hat{C}_{11}')^{-1}\hat{C}_{11}\hat{C}_{21}' \\ 1 \end{pmatrix}$$

$$= \frac{1}{\hat{C}_{22}^2} \begin{pmatrix} -\hat{C}_{11}^{-1\prime}\hat{C}_{21}' \\ 1 \end{pmatrix},$$

so

$$(\hat{\Sigma}_{31}, \hat{\Sigma}_{32}) \begin{pmatrix} \hat{\Sigma}_{11} & \hat{\Sigma}_{12} \\ \hat{\Sigma}_{21} & \hat{\Sigma}_{22} \end{pmatrix}^{-1} e_{j^*,j^*} = \frac{1}{\hat{C}_{22}^2} \left(-\hat{C}_{31}\hat{C}_{11}'\hat{C}_{11}^{-1\prime}\hat{C}_{21}' + \hat{C}_{31}\hat{C}_{21}' + \hat{C}_{32}\hat{C}_{22} \right) = \frac{1}{\hat{C}_{22}}\hat{C}_{32}. \quad \square$$

C.5.3 Auxiliary asymptotic lemmas

Lemma C.5.6. *$T^{-1}\sum_{t=1}^{T} \|y_t - \tilde{y}_t\|^2 = O_p(T^{-2\zeta})$ and $T^{-1}\sum_{t=1}^{T} u_t(y_{t-1} - \tilde{y}_{t-1})' = O_p(T^{-2\zeta} + T^{-\zeta-1/2})$, where $u_t \equiv y_t - Ay_{t-1}$.*

Proof. Using Equation (3.2.1) and the definition $\tilde{y}_t \equiv (I_n - AL)^{-1}H\varepsilon_t$, we have

$$y_t - \tilde{y}_t = T^{-\zeta}(I_n - AL)^{-1}H\alpha(L)\varepsilon_t.$$

The lag polynomial $(I_n - AL)^{-1}H\alpha(L)$ is absolutely summable by virtue of being a product of absolutely summable polynomials, see Assumption 3.2.1(ii) and (v).

Brockwell and Davis (1991, Prop. 3.1.1) implies $E[T^{2\zeta}\|y_t - \tilde{y}_t\|^2] < \infty$. The first statement of the lemma then follows from Markov's inequality.

In order to establish the second part of the lemma, note that

$$\frac{1}{T}\sum_{t=1}^{T} u_t (y_{t-1} - \tilde{y}_{t-1})' = H\left(\frac{1}{T}\sum_{t=1}^{T} \varepsilon_t (y_{t-1} - \tilde{y}_{t-1})'\right) + T^{-\zeta}H\left(\frac{1}{T}\sum_{t=1}^{T} \alpha(L)\varepsilon_t (y_{t-1} - \tilde{y}_{t-1})'\right).$$

The product process $T^{\zeta}\varepsilon_t \otimes (y_{t-1} - \tilde{y}_{t-1})$ is a martingale difference sequence with finite variance by Lemma C.5.11 (note that this is a standard stochastic process and not a triangular array).

Hence, the first sample average in parenthesis on the right-hand side above is $O_p(T^{-\zeta-1/2})$ by Chebyshev's inequality. The second sample average in parenthesis on the right-hand side above is $O_p(T^{-\zeta})$ by Lemma C.5.11. Hence, the entire right-hand side in the above display is $O_p(T^{-\zeta-1/2}+T^{-2\zeta})$, as claimed. □

Lemma C.5.7.

$$T^{-1}\sum_{t=1}^{T}u_t y_{t-1}' = T^{-\zeta}H\sum_{\ell=1}^{\infty}\alpha_\ell DH'(A')^{\ell-1} + T^{-1}\sum_{t=1}^{T}H\varepsilon_t\tilde{y}_{t-1}' + o_p(T^{-\zeta}).$$

Proof.

$$\begin{aligned}
T^{-1}\sum_{t=1}^{T}u_t y_{t-1}' &= T^{-1}\sum_{t=1}^{T}u_t\tilde{y}_{t-1}' + \underbrace{T^{-1}\sum_{t=1}^{T}u_t(y_{t-1}-\tilde{y}_{t-1})'}_{=o_p(T^{-\zeta}) \text{ by Lemma C.5.6}} \\
&= T^{-1}\sum_{t=1}^{T}H\varepsilon_t\tilde{y}_{t-1}' + T^{-\zeta-1}\sum_{t=1}^{T}H\alpha(L)\varepsilon_t\tilde{y}_{t-1}' + o_p(T^{-\zeta}) \\
&= T^{-1}\sum_{t=1}^{T}H\varepsilon_t\tilde{y}_{t-1}' + T^{-\zeta}H\left(E[\alpha(L)\varepsilon_t\tilde{y}_{t-1}'] + o_p(1)\right) + o_p(T^{-\zeta}).
\end{aligned}$$

The last line invokes a law of large numbers, which applies because $\tilde{y}_t$ is an absolutely summable linear filter of the shocks, so we can apply Lemma C.5.11 in conjunction with Brockwell and Davis (1991, Thm. 7.1.1) and Chebyshev's inequality. Finally, note that

$$E[\alpha(L)\varepsilon_t\tilde{y}_{t-1}'] = \sum_{\ell=1}^{\infty}\sum_{s=0}^{\infty}\alpha_\ell E[\varepsilon_{t-\ell}\varepsilon_{t-s-1}']H'(A')^s = \sum_{\ell=1}^{\infty}\alpha_\ell DH'(A')^{\ell-1}. \qquad \square$$

Lemma C.5.8. $T^{-1}\sum_{t=1}^{T}y_{t-1}y_{t-1}' \xrightarrow{p} S$.

Proof. By Lemma C.5.6 and Cauchy-Schwarz, $T^{-1}\sum_{t=1}^{T}y_{t-1}y_{t-1}' = T^{-1}\sum_{t=1}^{T}\tilde{y}_{t-1}\tilde{y}_{t-1}' + o_p(1)$. Lemma C.5.11, Brockwell and Davis (1991, Thm. 7.1.1), and Chebyshev's inequality imply that the law of large numbers holds for $\{\tilde{y}_{t-1}\tilde{y}_{t-1}'\}$. □

Lemma C.5.9. *Omitting the subscript j^* in a slight abuse of notation, let $q_t \equiv (\underline{y}_{j^*,t}', y_{t-1}')'$. Then* *i)* $T^{-1}\sum_{t=1}^{T-h} q_t \varepsilon_{j^*,t} = O_p(T^{-\zeta} + T^{-1/2})$,

ii) $(T^{-1}\sum_{t=1}^{T-h} q_t q_t')^{-1} = O_p(1)$,

iii) $T^{-1}\sum_{t=1}^{T-h} q_t(\alpha_{j^*}(L)\varepsilon_t) = O_p(1)$,

where $\alpha_{j^*}(L)$ is the j^*-th row of $\alpha(L)$.

Proof. Let $\tilde{q}_t \equiv (\underline{\tilde{y}}_{j^*,t}', \tilde{y}_{t-1}')'$ and $\Delta_t \equiv q_t - \tilde{q}_t$. Note that

$$T^{-1}\sum_{t=1}^{T-h} q_t \varepsilon_{j^*,t} = T^{-1}\sum_{t=1}^{T-h} \Delta_t \varepsilon_{j^*,t} + T^{-1}\sum_{t=1}^{T-h} \tilde{q}_t \varepsilon_{j^*,t}. \tag{C.5.10}$$

Cauchy-Schwarz implies

$$\left\| T^{-1}\sum_{t=1}^{T-h} \Delta_t \varepsilon_{j^*,t} \right\| \leq \left(\frac{1}{T}\sum_{t=1}^{T-h} \|\Delta_t\|^2 \right)^{1/2} \left(\frac{1}{T}\sum_{t=1}^{T-h} \varepsilon_{j^*,t}^2 \right)^{1/2} = O_p(T^{-\zeta}) \times O_p(1),$$

using Lemma C.5.6 and Assumption C.3.1. The summands in the last sample average in (C.5.10) have mean zero due to shock orthogonality, so this sample average is $O_p\left(T^{-1/2}\right)$ by Lemma C.5.11 and Brockwell and Davis (1991, Thm. 7.1.1). This establishes part (i)) of the lemma.

For part (ii)) of the lemma, note that

$$\frac{1}{T}\sum_{t=1}^{T-h} q_t q_t' = \frac{1}{T}\sum_{t=1}^{T-h} \Delta_t \Delta_t' + \frac{1}{T}\sum_{t=1}^{T-h} \tilde{q}_t \Delta_t' + \frac{1}{T}\sum_{t=1}^{T-h} \Delta_t \tilde{q}_t' + \frac{1}{T}\sum_{t=1}^{T-h} \tilde{q}_t \tilde{q}_t'. \tag{C.5.11}$$

Lemma C.5.6 implies that the first term is $O_p\left(T^{-2\zeta}\right)$. Cauchy-Schwarz, along with Lemmas C.5.6 and C.5.8, imply that the second and third terms are $O_p(T^{-\zeta})$. The last term converges in probability to $\mathrm{var}(\tilde{q}_t)$, as in the proof of Lemma C.5.8. This matrix is non-singular, since $\tilde{q}_t = (\underline{\tilde{y}}_{j^*,t}', \tilde{y}_{t-1}')'$, where $\mathrm{var}(\tilde{y}_{t-1}) = S$ is non-singular by Assumption 3.2.1(iv), and Assumption 3.2.1(iii) implies that $\underline{\tilde{y}}_{j^*,t}$ equals a linear transformation of $\tilde{y}_{t-1}$ plus a non-singular orthogonal noise term.

Part (iii)) follows from Cauchy-Schwarz, Lemma C.5.8, and Assumption 3.2.1(v). □

Lemma C.5.10. *Use the same notation as Lemma C.5.9, and let*

$$\hat{x}_{h,t} \equiv (\vartheta - \hat{\vartheta}_h)' q_t + \varepsilon_{j^*,t} + T^{-\zeta} \alpha_{j^*}(L) \varepsilon_t.$$

Then

$$T^{-1} \sum_{t=1}^{T-h} (\hat{x}_{h,t} - \varepsilon_{j^*,t})^2 = O_p(T^{-2\zeta}) + o_p(T^{-1/2}). \tag{C.5.12}$$

Proof. It suffices by the c_r-inequality to show that

a) $T^{-1} \sum_{t=1}^{T-h} \left((\vartheta - \hat{\vartheta}_h)' q_t \right)^2 = O_p(T^{-2\zeta}) + o_p\left(T^{-1/2}\right)$,

b) $T^{-1} \sum_{t=1}^{T-h} \left(\alpha_{j^*}(L)\varepsilon_t\right)^2 = O_p\left(1\right)$.

To establish (a)), note that Cauchy-Schwarz implies

$$\frac{1}{T} \sum_{t=1}^{T-h} \left((\vartheta - \hat{\vartheta}_h)' q_t \right)^2 \le \left\| \vartheta - \hat{\vartheta}_h \right\|^2 \left(\frac{1}{T} \sum_{t=1}^{T-h} \|q_t\|^2 \right) = O_p\left(\left(T^{-\zeta} + T^{-1/2} \right)^2 \right) \times O_p(1),$$

by Lemmas C.5.8 and C.5.9. Hence, the right-hand side is $O_p(T^{-2\zeta}) + o_p(T^{-1/2})$.

Statement (b)) follows from Assumption 3.2.1(v) and Markov's inequality. □

Lemma C.5.11. *Let $\psi(L)$ and $\varphi(L)$ be two absolutely summable, univariate, two-sided lag polynomials. Then for any $j, k \in \{1, \dots, m\}$, the product process $z_t \equiv [\psi(L)\varepsilon_{j,t}] \times [\varphi(L)\varepsilon_{k,t}]$ has absolutely summable autocovariance function.*

Proof. Bound $\sum_{\ell=-\infty}^{\infty} |\operatorname{cov}(z_t, z_{t+\ell})|$ by

$$\begin{aligned} &\sum_{\ell=-\infty}^{\infty} \sum_{\tau_1=-\infty}^{\infty} \sum_{\tau_2=-\infty}^{\infty} \sum_{\tau_3=-\infty}^{\infty} \sum_{\tau_4=-\infty}^{\infty} |\psi_{\tau_1}||\varphi_{\tau_2}||\psi_{\tau_3}||\varphi_{\tau_4}| |\operatorname{cov}(\varepsilon_{j,t+\tau_1}\varepsilon_{k,t+\tau_2}, \varepsilon_{j,t+\tau_3+\ell}\varepsilon_{k,t+\tau_4+\ell})| \\ &\le \left(\sum_{\tau=-\infty}^{\infty} |\psi_\tau| \right)^2 \left(\sum_{\tau=-\infty}^{\infty} |\varphi_\tau| \right)^2 \left(\sup_{\tau_1,\dots,\tau_4} \sum_{\ell=-\infty}^{\infty} |\operatorname{cov}(\varepsilon_{j,t+\tau_1}\varepsilon_{k,t+\tau_2}, \varepsilon_{j,t+\tau_3+\ell}\varepsilon_{k,t+\tau_4+\ell})| \right). \end{aligned} \tag{C.5.13}$$

To finish the proof, we show that the supremum above is finite. By stationarity, the sum inside the supremum can be written as

$$\sum_{r=-\infty}^{\infty} |\operatorname{cov}(\varepsilon_{j,0}\varepsilon_{k,s}, \varepsilon_{j,r}\varepsilon_{k,r+\tau})|, \tag{C.5.14}$$

where we have substituted $s = \tau_2 - \tau_1$, $\tau = \tau_4 - \tau_3$, and $r = \ell + \tau_3 - \tau_1$. Now fix s and τ. Let $N\colon \mathbb{Z} \to \{1, 2, 3, 4\}$ denote the function that assigns each integer r to the number of times the maximum value of the tuple $(0, s, r, r + \tau)$ appears in the tuple. We group the terms indexed by r in the sum (C.5.14) according to their value of $N(r)$. First, since $\{\varepsilon_t\}$ is a martingale difference sequence, all terms r with $N(r) = 1$ yield a covariance of 0 and so do not contribute to the sum. Second, a simple enumeration of cases shows that there is at most 1 value of r with $N(r) = 3$ and at most one with $N(r) = 4$. Finally, consider terms r with $N(r) = 2$. If $s \neq 0$ and $\tau \neq 0$, then there are at most 4 such terms (since this requires $r \in \{0, s, -\tau, s - \tau\}$). If $s = 0$ and/or $\tau = 0$, terms with $N(r) = 2$ must be of the form $|\operatorname{cov}(\varepsilon_{j,0}\varepsilon_{k,0}, \varepsilon_{j,r}\varepsilon_{k,r+\tau})|$ (with $r, r + \tau < 0$) and/or $|\operatorname{cov}(\varepsilon_{j,0}\varepsilon_{k,0}, \varepsilon_{j,-r}\varepsilon_{k,s-r})|$ (with $r, r - s > 0$). The preceding arguments show that the sum (C.5.14) is bounded by

$$6E[\|\varepsilon_t\|^4] + 2\sum_{r=1}^{\infty}\sum_{b=1}^{\infty} \|\operatorname{cov}(\varepsilon_0 \otimes \varepsilon_0, \varepsilon_{-r} \otimes \varepsilon_{-b})\|,$$

which is finite by Assumption C.3.1 and does not depend on s or τ. Thus, the supremum in (C.5.13) is finite. $\square$

Bibliography

Ackerberg, D. A., Caves, K., and Frazer, G. (2015). Identification Properties of Recent Production Function Estimators. *Econometrica*, 83(6):2411–2451.

Adrian, T., Grinberg, F., Liang, N., Malik, S., and Yu, J. (2022). The Term Structure of Growth-at-Risk. *American Economic Journal: Macroeconomics*, 14(3):283–323.

Ahn, S. C. and Schmidt, P. (1995). Efficient estimation of models for dynamic panel data. *Journal of Econometrics*, 68(1):5–27.

Ai, C. and Chen, X. (2007). Estimation of possibly misspecified semiparametric conditional moment restriction models with different conditioning variables. *Journal of Econometrics*, 141(1):5–43. Special issue on "Semiparametric methods in econometrics".

Alan, S., Browning, M., and Ejrnæs, M. (2018). Income and Consumption: A Micro Semistructural Analysis with Pervasive Heterogeneity. *Journal of Political Economy*, 126(5):1827–1864.

Alessi, L., Barigozzi, M., and Capasso, M. (2011). Non-fundamentalness in structural econometric models: A review. *International Statistical Review*, 79(1):16–47.

Allen, F., Babus, A., and Carletti, E. (2009). Financial Crises: Theory and Evidence. *Annual Review of Financial Economics*, 1(1):97–116.

Alvarez, F., Lippi, F., and Souganidis, P. (2023). Price setting with strategic complementarities as a mean field game. *Econometrica*, 91(6):2005–2039.

Andrews, D. W. (1998). Hypothesis testing with a restricted parameter space. *Journal of Econometrics*, 84(1):155–199.

Andrews, D. W. K. (1991). Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation. *Econometrica*, 59(3):817.

Andrews, D. W. K. (1996). Admissibility of the Likelihood Ratio Test when the Parameter Space is Restricted under the Alternative. *Econometrica*, 64(3):705.

Andrews, D. W. K. (2001). Testing When a Parameter is on the Boundary of the Maintained Hypothesis. *Econometrica*, 69(3):683–734.

Andrews, D. W. K. and Ploberger, W. (1994). Optimal Tests when a Nuisance Parameter is Present Only Under the Alternative. *Econometrica*, 62(6):1383.

Andrews, I., Gentzkow, M., and Shapiro, J. M. (2017). Measuring the Sensitivity of Parameter Estimates to Estimation Moments*. *The Quarterly Journal of Economics*, 132(4):1553–1592.

Aquaro, M., Bailey, N., and Pesaran, M. H. (2021). Estimation and inference for spatial models with heterogeneous coefficients: An application to US house prices. *Journal of Applied Econometrics*, 36(1):18–44.

Arellano, M. and Bond, S. (1991). Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations. *The Review of Economic Studies*, 58(2):277.

Arellano, M. and Bonhomme, S. (2012). Identifying Distributional Characteristics in Random Coefficients Panel Data Models. *Review of Economic Studies*, 79(3):987–1020.

Arellano, M. and Bover, O. (1995). Another look at the instrumental variable estimation of error-components models. *Journal of Econometrics*, 68(1):29–51.

Armstrong, T. B. and Kolesár, M. (2021). Sensitivity analysis using approximate moment condition models. *Quantitative Economics*, 12(1):77–108.

Armstrong, T. B., Kolesár, M., and Plagborg-Møller, M. (2022). Robust empirical Bayes confidence intervals. *Econometrica*, 90(6):2567–2602.

Auclert, A., Monnery, H., Rognlie, M., and Straub, L. (2023). Managing an Energy Shock: Fiscal and Monetary Policy. Working paper.

Banafti, S. and Lee, T.-H. (2022). Inferential Theory for Granular Instrumental Variables in High Dimensions. Working paper, arXiv.

Bang, H. and Robins, J. M. (2005). Doubly robust estimation in missing data and causal inference models. *Biometrics*, 61(4):962–972.

Bartholomew, D. J. (1959). A Test Of Homogeneity for Ordered Alternatives. *Biometrika*, 46(1/2):36.

Baumeister, C. and Hamilton, J. D. (2023). A Full-Information Approach to Granular Instrumental Variables. Working paper.

Bayoumi, T. and Eichengreen, B. (1992). Shocking Aspects of European Monetary Unification. Technical Report w3949, National Bureau of Economic Research, Cambridge, MA.

Beach, R. H., Bullock, A. M., Heller, K. B., and Domanico, J. L. (2000). Lime Production: Industry Profile. Technical report, Environmental Protection Agency.

Belloni, A. and Chernozhukov, V. (2013). Least squares after model selection in high-dimensional sparse models. *Bernoulli*, 19(2):521–547.

Blundell, R. and Bond, S. (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics*.

Blundell, R. and Bond, S. (2000). GMM Estimation with persistent panel data: An application to production functions. *Econometric Reviews*, 19(3):321–340.

Bonhomme, S., Lamadon, T., and Manresa, E. (2019). A Distributional Framework for Matched Employer Employee Data. *Econometrica*, 87(3):699–739.

Bonhomme, S. and Manresa, E. (2015). Grouped Patterns of Heterogeneity in Panel Data: Grouped Patterns of Heterogeneity. *Econometrica*, 83(3):1147–1184.

Bonhomme, S. and Weidner, M. (2022). Minimizing sensitivity to model misspecification. *Quantitative Economics*, 13(3):907–954.

Braun, P. A. and Mittnik, S. (1993). Misspecifications in vector autoregressions and their effects on impulse responses and variance decompositions. *Journal of Econometrics*, 59(3):319–341.

Brockwell, P. J. and Davis, R. A. (1991). *Time Series: Theory and Methods*. Springer Series in Statistics. Springer, 2nd edition.

Browning, M., Ejrnaes, M., and Alvarez, J. (2010). Modelling Income Processes with Lots of Heterogeneity. *Review of Economic Studies*, 77(4):1353–1381.

Camanho, N., Hau, H., and Rey, H. (2022). Global Portfolio Rebalancing and Exchange Rates. *The Review of Financial Studies*, 00(0):47.

Carroll, C. D. and Samwick, A. A. (1997). The nature of precautionary wealth. *Journal of Monetary Economics*, 40(1):41–71.

Chamberlain, G. (1992). Efficiency Bounds for Semiparametric Regression. *Econometrica*, 60(3):567.

Chant, D. (1974). On asymptotic tests of composite hypotheses in nonstandard conditions. *Biometrika*, 61(2):291–298.

Chari, V. V., Kehoe, P. J., and McGrattan, E. R. (2008). Are structural VARs with long-run restrictions useful in developing business cycle theory? *Journal of Monetary Economics*, 55(8):1337–1352.

Chen, J., Shin, Y., and Zheng, C. (2022). Estimation and inference in heterogeneous spatial panels with a multifactor error structure. *Journal of Econometrics*, 229(1):55–79.

Chernoff, H. (1954). On the Distribution of the Likelihood Ratio. *The Annals of Mathematical Statistics*, 25(3):573–578.

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., and Robins, J. (2018). Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal*, 21(1):C1–C68.

Chernozhukov, V., Escanciano, J. C., Ichimura, H., Newey, W. K., and Robins, J. M. (2022). Locally robust semiparametric estimation. *Econometrica*, 90(4):1501–1535.

Chesher, A. (1984). Testing for Neglected Heterogeneity. *Econometrica*, 52(4):865.

Chodorow-Reich, G., Ghent, A., and Haddad, V. (2021). Asset Insulators. *The Review of Financial Studies*, 34(3):1509–1539.

Conesa, J. C., Kitao, S., and Krueger, D. (2009). Taxing Capital? Not a Bad Idea After All! *American Economic Review*, 99(1):25–48.

Constantinides, G. M. and Duffie, D. (1996). Asset Pricing with Heterogeneous Consumers. *Journal of Political Economy*, 104(2):219–240.

Davidson, J. (1994). *Stochastic Limit Theory: An Introduction for Econometricians*. Advanced Texts in Econometrics. OUP Oxford.

De Loecker, J., Eeckhout, J., and Unger, G. (2020). The Rise of Market Power and the Macroeconomic Implications*. *The Quarterly Journal of Economics*, 135(2):561–644.

De Loecker, J. and Warzynski, F. (2012). Markups and Firm-Level Export Status. *American Economic Review*, 102(6):2437–2471.

Edmond, C., Midrigan, V., and Xu, D. Y. (2023). How Costly Are Markups? *Journal of Political Economy*, 131(7):1619–1675.

Edwards, S., Edwards, A. C., Bentolila, S., and Galí, J. (2000). Economic reforms and labour markets: Policy issues and lessons from Chile. *Economic Policy*, 15(30):182–229.

Elliott, G., Rothenberg, T. J., and Stock, J. H. (1996). Efficient Tests for an Autoregressive Unit Root. *Econometrica*, 64(4):813–836.

Evdokimov, K. S. and Zeleneev, A. (2023). Simple Estimation of Semiparametric Models with Measurement Errors.

Fagereng, A., Holm, M. B., and Natvik, G. J. (2021). MPC Heterogeneity and Household Balance Sheets. *American Economic Journal: Macroeconomics*, 13(4):1–54.

Fama, E. F. and French, K. R. (2015). A five-factor asset pricing model. *Journal of Financial Economics*, 116(1):1–22.

Fernald, J., Gandhi, A., Ruzic, D., and Traina, J. (2025). Micro and Macro Perspectives on Production-Based Markups. Working paper.

Flynn, J. P. and Sastry, K. (2022). The Macroeconomics of Narratives. *SSRN Electronic Journal*.

Fuster, A., Kaplan, G., and Zafar, B. (2021). What Would You Do with $500? Spending Responses to Gains, Losses, News, and Loans. *The Review of Economic Studies*, 88(4):1760–1795.

Gabaix, X. and Koijen, R. S. J. (2021). In Search of the Origins of Financial Fluctuations: The Inelastic Markets Hypothesis. Technical Report w28967, National Bureau of Economic Research.

Gabaix, X. and Koijen, R. S. J. (2024). Granular Instrumental Variables. *Journal of Political Economy*, 132(7):2274–2303.

Gandhi, A., Navarro, S., and Rivers, D. A. (2020). On the Identification of Gross Output Production Functions. *Journal of Political Economy*, 128(8):2973–3016.

Genovese, C. and Wasserman, L. (2008). Adaptive confidence bands. *The Annals of Statistics*, 36(2):875 – 905.

González-Casasús, O. and Schorfheide, F. (2025). Misspecification-robust shrinkage and selection for VAR forecasts and IRFs. Technical Report 33474, National Bureau of Economic Research.

Graham, B. S. and Powell, J. L. (2012). Identification and Estimation of Average Partial Effects in "Irregular" Correlated Random Coefficient Panel Data Models. *Econometrica*, 80(5):2105–2152.

Granger, C. W. J. and Morris, M. J. (1976). Time series modelling and interpretation. *Journal of the Royal Statistical Society. Series A (General)*, 139(2):246–257.

Gu, J. (2016). Neyman's C(α) Test for Unobserved Heterogeneity. *Econometric Theory*, 32(6):1483–1522.

Gu, J. and Koenker, R. (2017). Unobserved Heterogeneity in Income Dynamics: An Empirical Bayes Perspective. *Journal of Business & Economic Statistics*, 35(1).

Guvenen, F. (2007). Learning Your Earning: Are Labor Income Shocks Really Very Persistent? *American Economic Review*, 97(3):687–712.

Guvenen, F. (2009). An empirical investigation of labor income processes. *Review of Economic Dynamics*, 12(1):58–79.

Guvenen, F., Karahan, F., Ozkan, S., and Song, J. (2021). What Do Data on Millions of U.S. Workers Reveal About Lifecycle Earnings Dynamics? *Econometrica*, 89(5):2303–2339.

Hahn, J., Newey, W. K., and Smith, R. J. (2014). Neglected heterogeneity in moment condition models. *Journal of Econometrics*, 178:86–100.

Hall, R. E. (1988). The Relation between Price and Marginal Cost in U.S. Industry. *Journal of Political Economy*, 96(5):921–947.

Hall, R. E. (1990). Invariance Properties of Solow's Productivity Residual. In Solow, R. M. and Diamond, P. A., editors, *Growth, Productivity, Unemployment: Essays to Celebrate Bob Solow's Birthday*, pages 71–112. MIT Press, Cambridge, Mass.

Hall, R. E., Blanchard, O. J., and Hubbard, R. G. (1986). Market Structure and Macroeconomic Fluctuations. *Brookings Papers on Economic Activity*, 1986(2):285.

Hallin, M., Ingenbleek, J.-F., and Puri, M. L. (1989). Asymptotically most powerful rank tests for multivariate randomness against serial dependence. *Journal of Multivariate Analysis*, 30(1):34–71.

Hallin, M. and Puri, M. L. (1988). Optimal rank-based procedures for time series analysis: Testing an ARMA model against other ARMA models. *The Annals of Statistics*, 16(1):402–432.

Hansen, B. E. (1994). Autoregressive Conditional Density Estimation. *International Economic Review*, 35(3):705.

Hansen, L. P. (1982). Large Sample Properties of Generalized Method of Moments Estimators. *Econometrica*, 50(4):1029.

Hansen, L. P., Heaton, J., and Yaron, A. (1996). Finite-Sample Properties of Some Alternative GMM Estimators. *Journal of Business & Economic Statistics*, 14(3):262.

Hausman, J. (1978). Specification tests in econometrics. *Econometrica*, 46(6):1251–1271.

Hausman, J. A. (1983). Chapter 7 Specification and estimation of simultaneous equation models. In *Handbook of Econometrics*, volume 1, pages 391–448. Elsevier.

Heathcote, J., Storesletten, K., and Violante, G. L. (2010). The Macroeconomic Implications of Rising Wage Inequality in the United States. *Journal of Political Economy*, 118(4):681–722.

Herbst, E. P. and Johannsen, B. K. (2024). Bias in local projections. *Journal of Econometrics*, 240(1):105655.

Hsieh, C.-T. and Klenow, P. J. (2009). Misallocation and Manufacturing TFP in China and India. *The Quarterly Journal of Economics*, 124(4):1403–1448.

Hubbard, R., Skinner, J., and Zeldes, S. P. (1994). The importance of precautionary motives in explaining individual and aggregate saving. *Carnegie-Rochester Conference Series on Public Policy*, 40:59–125.

Inoue, A. and Kilian, L. (2002). Bootstrapping smooth functions of slope parameters and innovation variances in VAR(∞) models. *International Economic Review*, 43(2):309–331.

Inoue, A. and Kilian, L. (2020). The uniform validity of impulse response inference in autoregressions. *Journal of Econometrics*, 215(2):450–472.

Jordà, Ò. (2005). Estimation and Inference of Impulse Responses by Local Projections. *American Economic Review*, 95(1):161–182.

Jordà, Ò. (2023). Local projections for applied economics. *Annual Review of Economics*, 15(1):607–631.

Känzig, D. R. (2021). The macroeconomic effects of oil supply news: Evidence from OPEC announcements. *American Economic Review*, 111(4):1092–1125.

Kilian, L. (1998). Small-sample confidence intervals for impulse response functions. *Review of Economics and Statistics*, 80(2):218–230.

Kilian, L. and Kim, Y. J. (2011). How reliable are local projection estimators of impulse responses? *Review of Economics and Statistics*, 93(4):1460–1466.

Kilian, L. and Lütkepohl, H. (2017). *Structural Vector Autoregressive Analysis*. Cambridge University Press.

Kitamura, K., Otsu, T., and Evdokimov, K. S. (2013). Robustness, Infinitesimal Neighborhoods, and Moment Restrictions. *Econometrica*, 81(3):1185–1201.

Klubock, T. M. (2014). *La Frontera: Forests and Ecological Conflict in Chile's Frontier Territory*. Duke University Press.

Kuersteiner, G. M. (2001). Optimal instrumental variables estimation for ARMA models. *Journal of Econometrics*, 104(2):359–405.

Lazarus, E., Lewis, D. J., Stock, J. H., and Watson, M. W. (2018). HAR Inference: Recommendations for Practice. *Journal of Business & Economic Statistics*, 36(4):541–559.

Lee, W. (2022). Identification and estimation of dynamic random coefficient models. Working paper.

Leeb, H. and Pötscher, B. M. (2005). Model selection and inference: Facts and fiction. *Econometric Theory*, 21(01):21–59.

Leeper, E. M., Sims, C. A., Zha, T., Hall, R. E., and Bernanke, B. S. (1996). What Does Monetary Policy Do? *Brookings Papers on Economic Activity*, 1996(2):1.

Lehmann, E. and Romano, J. P. (2022). *Testing Statistical Hypotheses*. Springer Texts in Statistics. Springer Cham, 4 edition.

Levinsohn, J. and Petrin, A. (2003). Estimating Production Functions Using Inputs to Control for Unobservables. *Review of Economic Studies*, 70(2):317–341.

Lewis, D., Melcangi, D., and Pilossoph, L. (2026). Latent heterogeneity in the marginal propensity to consume. *The Review of Economic Studies*, page rdaf102.

Li, D., Plagborg-Møller, M., and Wolf, C. K. (2024). Local projections vs. VARs: Lessons from thousands of DGPs. *Journal of Econometrics*, 244(2):105722. Themed Issue: Macroeconometrics.

Lindsay, B. G. (1995). Mixture Models: Theory, Geometry and Applications. *NSF-CBMS Regional Conference Series in Probability and Statistics*, 5:i–163.

Liu, L. (1993). Entry-exit, learning, and productivity change: Evidence from Chile. *Journal of Development Economics*, 42(2):217–242.

Llorca-Jaña, M., Nazer, R., Morales, D., and Navarrete-Montalvo, J. (2020). Milk and meat consumption and production in Chile, c. 1930-2017: A history of a successful nutrition transition. *Historia Agraria Revista de agricultura e historia rural*, 82:245–285.

Lütkepohl, H. (1984). Linear aggregation of vector autoregressive moving average processes. *Economics Letters*, 14(4):345–350.

Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis*. Springer.

MaCurdy, T. E. (1982). The use of time series processes to model the error structure of earnings in a longitudinal data analysis. *Journal of Econometrics*, 18(1):83–114.

Magnus, J. and Neudecker, H. (2007). *Matrix Differential Calculus with Applications in Statistics and Econometrics*. Wiley Series in Probability and Statistics. John Wiley & Sons, 3rd edition.

Manski, C. F. (1993). Identification of Endogenous Social Effects: The Reflection Problem. *The Review of Economic Studies*, 60(3):531.

Montiel Olea, J. L. and Plagborg-Møller, M. (2021). Local projection inference is simpler and more robust than you think. *Econometrica*, 89(4):1789–1823.

Montiel Olea, J. L., Plagborg-Møller, M., Qian, E., and Wolf, C. K. (2025). Local projections or VARs? A primer for macroeconomists. *NBER Macroeconomics Annual*. Forthcoming.

Müller, U. K. and Stock, J. H. (2011). Forecasts in a slightly misspecified finite order VAR. Manuscript, Princeton University.

Nakamura, E. and Steinsson, J. (2018). Identification in Macroeconomics. *Journal of Economic Perspectives*, 32(3):59–86.

Newey, W. (1985). Generalized method of moments specification testing. *Journal of Econometrics*, 29(3):229–256.

Newey, W. K. and McFadden, D. (1994). Chapter 36 Large sample estimation and hypothesis testing. In *Handbook of Econometrics*, volume 4, pages 2111–2245. Elsevier.

Neyman, J. (1959). Optimal asymptotic tests of composite statistical hypotheses. In Grenander, U., editor, *Probability and Statistics, the Herald Cramer Volume*, pages 213–234. John Wiley & Sons, New York.

Olley, G. S. and Pakes, A. (1996). The Dynamics of Productivity in the Telecommunications Equipment Industry. *Econometrica*, 64(6):1263.

Ottonello, P. and Winberry, T. (2020). Financial Heterogeneity and the Investment Channel of Monetary Policy. *Econometrica*, 88(6):2473–2502.

Pakes, A. and Pollard, D. (1989). Simulation and the Asymptotics of Optimization Estimators. *Econometrica*, 57(5):1027.

Perlman, M. D. (1969). One-Sided Testing Problems in Multivariate Analysis. *The Annals of Mathematical Statistics*, 40(2):549–567.

Pesaran, M. and Smith, R. (1995). Estimating long-run relationships from dynamic heterogeneous panels. *Journal of Econometrics*, 68(1):79–113.

Pesaran, M. H. and Yang, L. (2024). Heterogeneous Autoregressions in Short T Panel Data Models.

Plagborg-Møller, M. and Wolf, C. K. (2021). Local projections and VARs estimate the same impulse responses. *Econometrica*, 89(2):955–980.

Pope, A. L. (1990). Biases of estimators in multivariate non-Gaussian autoregressions. *Journal of Time Series Analysis*, 11(3):249–258.

Ramey, V. A. (2016). Macroeconomic shocks and their propagation. In Taylor, J. B. and Uhlig, H., editors, *Handbook of Macroeconomics*, volume 2, chapter 2, pages 71–162. Elsevier.

Rao, C. R. and Mitra, S. K. (1972). Generalized inverse of a matrix and its applications. *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*, 1:601–620.

Raval, D. (2023). Testing the Production Approach to Markup Estimation. *Review of Economic Studies*, 90(5):2592–2611.

Robertson, T., Wright, F., and Dykstra, R. (1988). *Order Restricted Statistical Inference*. Wiley.

Robins, J. and Rotnitzky, A. (2001). Comments on "inference for semiparametric models: Some questions and an answer", by p.j. bickel and j. kwon. *Statistica Sinica*, 11(4):920–936.

Robins, J. M., Mark, S. D., and Newey, W. K. (1992). Estimating exposure effects by modelling the expectation of exposure conditional on confounders. *Biometrics*, 48(2):479–495.

Robins, J. M. and Rotnitzky, A. (1995). Semiparametric efficiency in multivariate regression models with missing data. *Journal of the American Statistical Association*, 90(429):122–129.

Robins, J. M., Rotnitzky, A., and van der Laan, M. (2000). On profile likelihood: Comment. *Journal of the American Statistical Association*, 95(450):477–482.

Rothenberg, T. J. (1984). Chapter 15 Approximating the distributions of econometric estimators and test statistics. In *Handbook of Econometrics*, volume 2, pages 881–935. Elsevier.

Schorfheide, F. (2005). VAR forecasting under misspecification. *Journal of Econometrics*, 128(1):99–136.

Silvapulle, M. J. and Sen, P. K. (2001). *Constrained Statistical Inference: Inequality, Order, and Shape Restrictions*. Wiley Series in Probability and Statistics. John Wiley & Sons.

Silvapulle, M. J. and Silvapulle, P. (1995). A Score Test against One-Sided Alternatives. *Journal of the American Statistical Association*, 90(429):342–349.

Sims, C. A. (1980). Macroeconomics and reality. *Econometrica*, 48(1):1–48.

Staiger, D. and Stock, J. H. (1997). Instrumental Variables Regression with Weak Instruments. *Econometrica*, 65(3):557–586.

Stock, J. H. and Watson, M. W. (2016). Dynamic factor models, factor-augmented vector autoregressions, and structural vector autoregressions in macroeconomics. In Taylor, J. B. and Uhlig, H., editors, *Handbook of Macroeconomics*, volume 2, chapter 8, pages 415–525. Elsevier.

Stock, J. H. and Watson, M. W. (2018). Identification and estimation of dynamic causal effects in macroeconomics using external instruments. *Economic Journal*, 128(610):917–948.

Su, L., Wang, W., and Xu, X. (2023). Identifying latent group structures in spatial dynamic panels. *Journal of Econometrics*, 235(2):1955–1980.

van der Vaart, A. (1998). *Asymptotic Statistics*. Cambridge University Press.

Wooldridge, J. M. (2005). Fixed-Effects and Related Estimators for Correlated Random-Coefficient and Treatment-Effect Panel Data Models. *Review of Economics and Statistics*, 87(2):385–390.

Wooldridge, J. M. (2009). On estimating firm-level production functions using proxy variables to control for unobservables. *Economics Letters*, 104(3):112–114.

Xu, K.-L. (2023). Local projection based inference under general conditions. Manuscript, Indiana University Bloomington.