Testing for Omitted Heterogeneity*

Eric Qian†

November 2, 2025

[Click here for the latest version]

Abstract

Structural estimation inevitably involves a choice of which parameters to treat as homogeneous across units, and incorrectly imposing homogeneity can lead to uninterpretable parameter estimates. To discipline this choice with data, I develop an optimal test for omitted unit-level heterogeneity applicable to moment condition models. Unlike existing semiparametric specification tests for heterogeneity, the test asymptotically maximizes a weighted average power criterion; for the special case of scalar heterogeneity, the test is the asymptotically uniformly most powerful test. Through simulations, I show that likelihood-based tests for parameter heterogeneity can severely over-reject when the likelihood function is misspecified. I study two applications. First, applied to income dynamics, I use the heterogeneity test as a diagnostic to determine the appropriate level of aggregation by education groups. Second, firm-level heterogeneity threatens estimates of production functions and ultimately estimates of the distribution of markups. Estimating production functions at the subindustry level, I use the test as a diagnostic for determining the subindustries whose moment conditions are compatible with the data.

Keywords: Moment condition model; semiparametric efficiency; income dynamics; markup estimation; random coefficients

*I am deeply grateful to my advisors Mikkel Plagborg-Møller, Mark Watson, and John Grigsby for their guidance and support. I also thank Ulrich Müller, Michal Kolesár, and Bo Honoré for their comments that substantially improved the paper. Thanks to the participants of the Princeton Student Macroeconomics Workshop and Princeton Student Econometrics Workshop for their helpful comments. All errors are my own.

†Princeton University. Email: ericqian@princeton.edu. Website: https://www.eric-qian.com.

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 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 statistic is then constructed using a second-derivative weighted sum of the normalized fitted moment conditions. After normalizing, the test statistic admits the interpretation as 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 (such as omnibus specification tests like the J test), the proposed test directs power toward the specific alternative hypotheses that arise from parameter heterogeneity giving rise to power gains. 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 specifications 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 to split the sample by education groups, helping ensure 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) give 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 contrasts 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 2 illustrates the main ideas of the proposed heterogeneity test in two motivating examples. Section 3 introduces the heterogeneity test. Section 4 presents a framework for semiparametric efficiency of the test. Through simulations, Section 5 examines the finite sample properties of the proposed test and compares with competing approaches. Section 6 applies the test to applications on income dynamics and markup dispersion. Section 7 concludes.

2 Motivating examples

Section 2.1 introduces a panel AR(2) to aid in discussing the considerations researchers face in estimating wide dynamic panel models. Section 2.2 illustrate 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.

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 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

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

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. 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 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.[1] 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.

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.

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

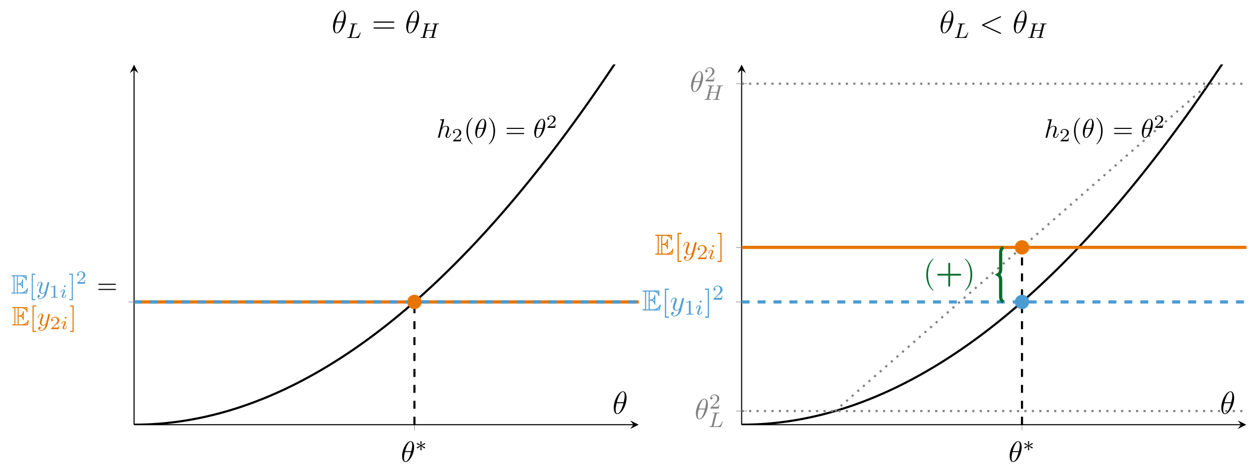

Figure 1: Plots of $h_2(\theta) = \theta^2$ and $\mathbb{E}(y_{2i})$ under parameter homogeneity (left panel) and under parameter heterogeneity (right panel).

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

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

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 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 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.

3 Testing for parameter heterogeneity

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

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 $\boldsymbol{\varepsilon}_i$:

$$\mathbf{y}_i = \mathbf{f}(\mathbf{x}_i, \boldsymbol{\varepsilon}_i; \boldsymbol{\theta}_i) \tag{2}$$

$$\theta_{ki} = \theta_k^* + u_{ki}\sigma_k \text{ for } k = 1, \dots p. \tag{3}$$

Random vector $\mathbf{u}_i = (u_{1i}, \dots, u_{pi})'$ 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 $\boldsymbol{\sigma} = (\sigma_1, \dots, \sigma_p)'$ 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 $\boldsymbol{\theta}^* = (\theta_1^*, \dots, \theta_p^*)'$ 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 inducing 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 $\boldsymbol{\theta}^*$

$$\mathbb{E}[\mathbf{g}(\mathbf{x}, \mathbf{y}^0; \boldsymbol{\theta}^*)] = \mathbf{0}_{m\times 1} \text{ where } \mathbf{y}^0 = \mathbf{f}(\mathbf{x}, \boldsymbol{\varepsilon}; \boldsymbol{\theta}^*). \tag{4}$$

The setup captures the scenario that a researcher has derived a set of moment conditions 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 3.1. $\boldsymbol{\sigma} = \mathbf{s}n^{-1/4}$ *where p-dimensional vector* $\mathbf{s}$ *is non-negative.*

Assumption 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 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).

3.2 Setup

Since Equation 4 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 found in Appendix A.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 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 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:

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

where the half-vectorization operator vech() stacks the elements of the lower triangular portion of a square matrix into a column vector. For what follows, the duplication matrix $\mathbf{D}$ maps the half-vectorization operator to the vectorization operator $\text{vec}(\mathbf{A}) = \mathbf{D}\text{vech}(\mathbf{A})$ for symmetric matrix $\mathbf{A}$.

Assumption A.2 of Appendix A.2 includes regularity conditions that are required for the consistency under the local alternative.

Assumption 3.2 (Smoothness)**.** *For the model in Equation 2, 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. For moment r, $\mathbf{0}_{T\times p} = \frac{\partial^2 \mathbf{g}_r(\mathbf{x},\mathbf{y};\boldsymbol{\theta})}{\partial \mathbf{y} \partial \boldsymbol{\theta}'}$.

Assumption 3.2 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 time continuous differentiability of the moment function imposed by the second condition is used in Lemma A.3 of Appendix A.2 for proving consistency of the sample expected Hessian matrix estimator $\widehat{\mathbf{H}}$ under local parameter heterogeneity. The requirement $\mathbf{0}_{T\times p} = \frac{\partial^2 \mathbf{g}_r(\mathbf{x},\mathbf{y};\boldsymbol{\theta})}{\partial \mathbf{y} \partial \boldsymbol{\theta}'}$ is a common feature of dynamic panel applications and allows the bias induced by parameter heterogeneity to be summarized by $\mathbf{H}$ alone (see the forthcoming results).[2]

Example. (Illustrative measurement error model, continued from Section 2.2.) Written in a moment condition framework, the moment function, expected Jacobian matrix, and expected Hessian matrix are

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

[2]The condition can be dropped, giving rise to an additional term containing the cross partial derivative of the generating model.

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 $\mathbf{\Sigma} = \mathbf{I}$.

3.3 Asymptotic bias and fitted moments

Lemma 3.1. *Impose Assumptions 3.1, 3.2, A.1, and A.2. Then,*

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

Lemma 3.1 shows the moment conditions evaluated at the mean parameter $\boldsymbol{\theta}^*$ are asymptotically biased. The rate of $n^{-1/4}$ in Assumption 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 $\boldsymbol{\varepsilon}_i$.[3] 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 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 $\mathbf{\Lambda}$ is non-zero). The duplication matrix $\mathbf{D}$ appears because the expected Hessian matrix $\mathbf{H}$ is written using the half-vectorization operator.

Proposition 3.1 (Asymptotic normality)**.** *Impose Assumptions 3.1, 3.2, A.1, and A.2. Then,*

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

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

Local parameter heterogeneity induces asymptotic bias in the GMM estimator, which ultimately affects inference. Proposition 3.1 shows that the asymptotic bias (which is at the same magnitude as the standard errors) is zero under parameter homogeneity ($\mathbf{\Lambda} = \mathbf{0}$)

[3] 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).

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 $\widehat{\boldsymbol{\theta}} \pm z^*_{1-\alpha/2}\sqrt{\text{avar}(\widehat{\boldsymbol{\theta}})/n}$ 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 3.2 (Fitted moments)**.** *Impose Assumptions 3.1, 3.2, A.1, and A.2. Then,*

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

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 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 $\mathbf{H}\mathbf{D}'\mathbf{D}\text{vech}(\boldsymbol{\Lambda})$ 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 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.

3.4 Constructing a test for parameter heterogeneity

Section 3.4.1 describes a limit experiment that forms the basis of the test for parameter heterogeneity. Section 3.4.2 introduces the heterogeneity test in the special case of testing for scalar heterogeneity. Section 3.4.3 generalizes the heterogeneity test to the multivariate case. See Section 4 for a discussion on semiparametric efficiency.

3.4.1 Limit experiment

The limiting distribution of the fitted GMM moments in Proposition 3.2 summarizes the over-identifying restrictions of a moment condition model and can be viewed as a limit experiment. Proposition 3.1 and Lemma A.3 of Appendix A.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 3.2 noted as $\mathbf{X}$:

$$\mathbf{X} \sim \mathcal{N}(-\frac{1}{2}\mathbf{M_W}\mathbf{H}\mathbf{D}'\mathbf{D}\text{vech}(\boldsymbol{\Lambda}), \mathbf{M_W}\boldsymbol{\Sigma}\mathbf{M_W}'). \tag{5}$$

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 4.

The score vector for testing $\text{vech}(\boldsymbol{\Lambda}) = \mathbf{0}$ in Equation 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 5 with respect to $\text{vech}(\boldsymbol{\Lambda})$ evaluated at $\text{vech}(\boldsymbol{\Lambda}) = \mathbf{0}$. Proposition A.2 of Appendix A.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

$$\mathbf{S} = (\mathbf{D}'\mathbf{D})\mathbf{H}'\boldsymbol{\Sigma}^{-1}\mathbf{X} \sim \mathcal{N}(-\frac{1}{2}\boldsymbol{\Omega}\text{vech}(\boldsymbol{\Lambda}), \boldsymbol{\Omega}) \tag{6}$$

where $\boldsymbol{\Omega} = (\mathbf{D}'\mathbf{D})\mathbf{H}'\boldsymbol{\Sigma}^{-1}\mathbf{M}\mathbf{H}(\mathbf{D}'\mathbf{D})$ and $\mathbf{M} = \mathbf{M}_{\boldsymbol{\Sigma}^{-1}}$. Proposition A.2 also implies that the score vector is invariant to the choice of weight matrix $\mathbf{W}$ so long as $\mathbf{G}'\mathbf{W}\mathbf{G}$ is full rank—the test statistic remains the same regardless of the choice of weight matrix.

Assumption 3.3. $\mathbf{\Omega}$ *is full rank.*

Assumption 3.3 can be viewed as a joint requirement of the existence of over-identifying restrictions (through $\mathbf{M}$) and nonlinearity of the moment function (through $\mathbf{H}$). Since $\mathbf{\Omega}$ is positive semidefinite by construction, the condition ensures $\mathbf{\Omega}$ is invertible and positive definite.

3.4.2 Test for scalar heterogeneity

When Λ is a scalar, a test for heterogeneity based on the limiting score vector of Equation 6 is one-sided. Under Assumption 3.3, the scalar score S defined in Equation 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 $\frac{\mathbf{H}'\mathbf{\Sigma}^{-1}\mathbf{X}}{\sqrt{\Omega}} < z_\alpha$ 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 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

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

Recall that the expected Hessian matrix $\mathbf{H}$ doesn't depend on on $\widehat{\theta}$ since the polynomial order of the two moment conditions with respect to θ is less than or equal to two (and $\mathbf{\Sigma}$ 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 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 $\widehat{\theta}^2$.

3.4.3 Test for multivariate heterogeneity

When $\mathbf{\Lambda}$ is instead a matrix (and again imposing Assumption 3.3), the one-sided scalar test can be generalized by computing the generalized likelihood ratio test statistic on the score vector of Equation 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 4.2. For now, begin with defining an infeasible test statistic that requires knowledge of the population objects

$$\mathcal{T} = \mathbf{S}'\mathbf{\Omega}^{-1}\mathbf{S} - \inf_{\mathbf{\Lambda}\in\mathcal{C}} \left[\mathbf{S} + \frac{1}{2}\mathbf{\Omega}\text{vech}(\mathbf{\Lambda})\right]' \mathbf{\Omega}^{-1}\left[\mathbf{S} + \frac{1}{2}\mathbf{\Omega}\text{vech}(\mathbf{\Lambda})\right] \tag{7}$$

where $\mathcal{C}$ gives the set of positive semidefinite matrices. 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 $\overline{\chi}^2(\mathbf{V}, \mathcal{C})$. 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.[4] 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 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 3.4.2 is a special case of the multivariate test outlined here when $\alpha < 0.5$.

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 $\inf_{\mathbf{\Lambda}\in\mathcal{C}} \left[\mathbf{S} + \frac{1}{2}\mathbf{\Omega}\text{vech}(\mathbf{\Lambda})\right]' \mathbf{\Omega}^{-1}\left[\mathbf{S} + \frac{1}{2}\mathbf{\Omega}\text{vech}(\mathbf{\Lambda})\right]$ directs power toward the specific alternative hypothesis of parameter heterogeneity. To see this, suppose 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 3.1 is the feasible analogue to the infeasible test statistic of Equation 7, replacing

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

the population objects with their corresponding estimators:

Definition 3.1 (Heterogeneity test statistic)**.** *Impose Assumptions 3.1, 3.2, A.1, A.2, and 3.3. Let* $\mathbf{S}_n = \mathbf{D}'\mathbf{D}\widehat{\mathbf{H}}'\widehat{\boldsymbol{\Sigma}}^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbf{g}(\mathbf{x}_i, \mathbf{y}_i; \widehat{\boldsymbol{\theta}}^{\text{eff}})$*. Then, the* ***heterogeneity test statistic*** *is*

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

where $\widehat{\boldsymbol{\theta}}^{\text{eff}}$ *is an efficient GMM estimator,* $\widehat{\boldsymbol{\Omega}} = (\mathbf{D}'\mathbf{D})\widehat{\mathbf{H}}'\widehat{\boldsymbol{\Sigma}}^{-1}\widehat{\mathbf{M}}\widehat{\mathbf{H}}(\mathbf{D}'\mathbf{D})$ *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 $\boldsymbol{\Omega}$ with $\widehat{\boldsymbol{\Omega}}$.
Remarks

1. The test can be adapted to situations where the researcher is only interested in testing for parameter heterogeneity for a subset of the estimated parameters. Such a situation could arise if elements of the parameter vector were already a random coefficient (like when certain parameters are already modeled as random coefficients) or if a researcher were interested in directing power in detecting parameter heterogeneity toward a subset of the parameter vector. In these cases, the rows and columns of the covariance matrix $\boldsymbol{\Lambda}$ that correspond to the nuisance parameter are taken to be zero.

   Simplifying matrix computations, such a procedure is equivalent to substituting the Hessian matrix $\mathbf{H}$ in the above derivations with a smaller Hessian matrix that *only* includes the second derivatives and partial derivatives corresponding to the parameters under test.

2. A normalized version of the score vector can be interpreted as a single draw from a multivariate normal distribution centered at vech($\boldsymbol{\Lambda}$)

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

   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 3.1—in particular, that parameter heterogeneity is not too large and that parameter heterogeneity is independent of the underlying shock.

3. The proposed heterogeneity 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.

4 Semiparametrically efficient testing

Section 4.1 sets up the framework for evaluating semiparametric efficiency. Section 4.2 discusses the main result, showing that the proposed test maximizes a weighted average power criterion for distant alternatives. Section 4.3 contains additional technical details.

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 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

$$\boldsymbol{\theta}_t = \boldsymbol{\theta}^* + \mathbf{c}t \text{ and } \boldsymbol{\sigma}_t = \boldsymbol{\sigma}^* + \mathbf{s}\sqrt{t} \tag{8}$$

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 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 4.

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 $\mathbf{u}$

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

Unlike parametric tests for the presence of mixture distributions, the specific distribution for $p_{xy,t}$ is unknown. These challenges are discussed in Section 4.3.

Score function. For computation of the semiparametric score function, Assumption A.3 of Appendix A.3 imposes two times continuous differentiability of the density of $(\mathbf{x}',\mathbf{y}')'$ with respect to $\boldsymbol{\theta}$. Lemma 4.1 shows that the semiparametric score function can be decomposed into parametric and non-parametric components:

Lemma 4.1. *Let operator $\frac{\partial_+}{\partial t}$ give the right-sided partial derivative with respect to t. Impose Assumptions A.3 and A.4. Let $(\mathbf{x}',\mathbf{y}')'$ be on the support of the probability model $P_{\boldsymbol{\theta}^*,\boldsymbol{\sigma}^*,\mathrm{vech}(\mathbf{C}),p_{xy},p_u}$. Then,*

$$\frac{\partial_+}{\partial t}\log q_t(\mathbf{x},\mathbf{y})\mid_{t=0} = \mathbf{a}'\boldsymbol{\ell}(\mathbf{x},\mathbf{y}) + \eta(\mathbf{x},\mathbf{y}) \text{ and } \boldsymbol{\ell}(\mathbf{x},\mathbf{y}) = \begin{bmatrix} \frac{\partial}{\partial\boldsymbol{\theta}}p(\mathbf{x},\mathbf{y};\boldsymbol{\theta}^*) \\ \frac{1}{2}(\mathbf{D}'\mathbf{D})\mathrm{vech}(\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}p(\mathbf{x},\mathbf{y};\boldsymbol{\theta}^*)) \end{bmatrix}$$

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

$\ell(\mathbf{x},\mathbf{y})$ can be interpreted as the score function for the parametric part of the model holding fixed the non-parametric part, where $\frac{\partial}{\partial\boldsymbol{\theta}}p(\mathbf{x},\mathbf{y};\boldsymbol{\theta}^*)$ is the score for $\boldsymbol{\theta}^*$ and $\frac{1}{2}(\mathbf{D}'\mathbf{D})\mathrm{vech}(\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}p(\mathbf{x},\mathbf{y};\boldsymbol{\theta}^*))$ is the score for $\boldsymbol{\sigma}^*$. The latter 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)). $\eta(\mathbf{x},\mathbf{y})$ 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 of Appendix A.3 is required for local asymptotic normality to hold.

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 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 $t \mapsto P_{t,\mathbf{a}'\boldsymbol{\ell}+\eta}$ be the submodel described in Section 4.1 (indexed by t) with arbitrary score function $\mathbf{a}'\boldsymbol{\ell} + \eta$ matching the convention of Chapter 25.6 of van der Vaart (1998). The "tangent set" describes the set of score functions and is $(\mathbf{c}', \boldsymbol{\lambda}')'\boldsymbol{\ell} + \eta$ for $\mathbf{c} \in \mathbb{R}^p$, $\text{vech}^{-1}(\boldsymbol{\lambda}) \in \mathcal{C}$, 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.4 of Appendix A.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 4.1 (to follow) is the asymptotic representation theorem (see e.g. Theorem 9.3 of van der Vaart (1998)).[5] A statistic that converges weakly in the model $P_{n^{-1/2},\mathbf{a}'\boldsymbol{\ell}+\eta}$ is "matched" by some statistic in the limit experiment—in this case, the limit experiment is a single observation $\mathbf{Z}$ of a multivariate normal distribution with mean $\text{vech}(\boldsymbol{\Lambda})$ and variance $4\boldsymbol{\Omega}^{-1}$:

$$\mathbf{Z} \sim \mathcal{N}(\text{vech}(\boldsymbol{\Lambda}), 4\boldsymbol{\Omega}^{-1}). \tag{9}$$

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 9 for establishing the asymptotic properties of tests. Equation 9 also implies that the discussion of Section 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 $-\frac{1}{2}\boldsymbol{\Omega}\mathbf{Z}$ is equal in distribution to $\mathbf{S}$.

Since the matrix $\boldsymbol{\Lambda}$ is unknown, Equation 9 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 9 is to compare the power of a test for *one* particular pointwise alternative. Call such an alternative $\text{vech}(\overline{\boldsymbol{\Lambda}})$. Then, the Neyman-Pearson lemma implies that any other test is less powerful than the likelihood ratio test comparing the pointwise alternative $\text{vech}(\overline{\boldsymbol{\Lambda}})$ 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,

[5] 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.

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 $w_r(\boldsymbol{\lambda})$ where $\boldsymbol{\lambda}$ is a $p(p+1)/2$-dimensional vector. $\boldsymbol{\lambda}$ corresponds to the lower triangular portion of the heterogeneity covariance matrix $\boldsymbol{\Lambda}$. For radius $r > 0$, let $p_\lambda(\boldsymbol{\lambda}; r)$ be the probability density function of $r\boldsymbol{\Omega}^{-1/2}\boldsymbol{\zeta}$ where $\boldsymbol{\zeta}$ is uniformly distributed on the $p(p+1)/2$-dimensional unit sphere. Let vech^{-1} reconstruct a symmetric matrix from its half-vectorized form. 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.4 of Appendix A.3 gives technical conditions that allow the interchange of integrals and derivatives. Theorem 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 4.1. *Impose Assumptions 3.2, 3.3, A.1, A.2, and A.4. Suppose $P_{t,(\mathbf{c}',\boldsymbol{\lambda}')\boldsymbol{\ell}+\eta}$ is differentiable in quadratic mean at $t = 0$ and $(\mathbf{c}', \boldsymbol{\lambda}')\boldsymbol{\ell} + \eta$ 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,*

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

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,*

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

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 found in Appendix A.3 shows that the scalar test for heterogeneity described in Section 3.4.2 is the locally uniformly most powerful test. The result can be seen in Equation 9 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 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 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 5. The expansion of Lemma 4.1 (to follow) reduces to that of Gu (2016) if the likelihood were instead fully specified (making the moment restrictions unnecessary).

3. In general, it is difficult to make arguments in favor of one weight function over another (Andrews, 1998). The particular weight function in Theorem 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 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 4.1 extend these same arguments to the particular multivariate hypothesis testing problem considered in this paper.

4.3 Proof sketch

This subsection gives theoretical context for Theorem 4.1 by describing the computation of the efficient influence function and efficient information matrix, which can be viewed as an information bound for $\text{vech}(\mathbf{\Lambda}_0)$.

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

$$H_0\colon \boldsymbol{\psi}(P) = \mathbf{0}_{\frac{p(p+1)}{2}\times 1} \text{ vs. } H_1\colon \boldsymbol{\psi}(P) \in \{\boldsymbol{\lambda} : \text{vech}^{-1}(\boldsymbol{\lambda}) \text{ is positive semidefinite}\} \setminus \mathbf{0}_{\frac{p(p+1)}{2}\times 1}. \quad (10)$$

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

Analogous to the information matrix in likelihood theory, the efficient information matrix can be interpreted as the summary of the information available on $\text{vech}(\mathbf{\Lambda}_0)$. To compute, note that $\boldsymbol{\psi}(P_{t,(\mathbf{c}',\boldsymbol{\lambda}')\boldsymbol{\ell}+\eta})$ is differentiable in the ordinary sense at $t = 0$ as $\partial_+\boldsymbol{\psi}(P_{t,(\mathbf{c}',\boldsymbol{\lambda}')\boldsymbol{\ell}+\eta})/\partial t \mid_{t=0} =$

$\boldsymbol{\lambda}$. Then, the Riesz representation theorem implies $\text{vech}(\boldsymbol{\Lambda}_0)$ is differentiable *as a parameter on the model*[6] if and only if there exists some function $\widetilde{\boldsymbol{\psi}}(\mathbf{x}, \mathbf{y})$, the *efficient influence function*, such that

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

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.6 in Appendix A.3 finds such a function $\widetilde{\boldsymbol{\psi}}(\mathbf{x}, \mathbf{y})$, showing that the efficient information matrix and efficient influence function are

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

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 $\text{vech}(\boldsymbol{\Lambda}_0)$ which is exactly the covariance of the limit experiment $\mathbf{Z}$ described in Equation 9.

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 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

$$y_{it} = \phi_{1i}y_{it-1} + \phi_{2i}y_{it-2} + \varepsilon_{it}, \quad \phi_{1i} \sim \mathcal{N}(.5, \delta^2), \quad \phi_{2i} \sim \mathcal{N}(.3, \delta^2). \tag{11}$$

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

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

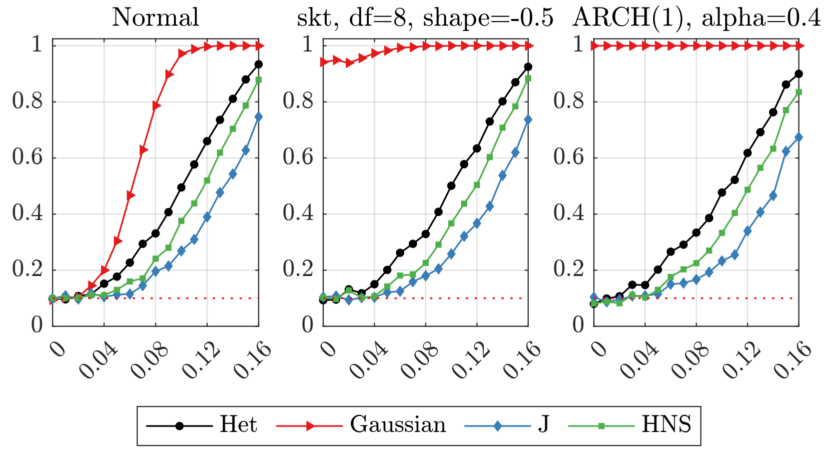

Figure 2: Empirical rejection probabilities by test and shock distribution. The horizontal axis gives the standard deviation of the random coefficient δ and the vertical axis gives the empirical rejection probability. "Het" gives this paper's moment-based heterogeneity test, "Gaussian" gives a Gaussian likelihood $C(\alpha)$ test, "J" gives a GMM Sargan-Hansen J-test, and HNS gives the Hahn et al. (2014) test.

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 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 11 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.

From Figure 2, 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

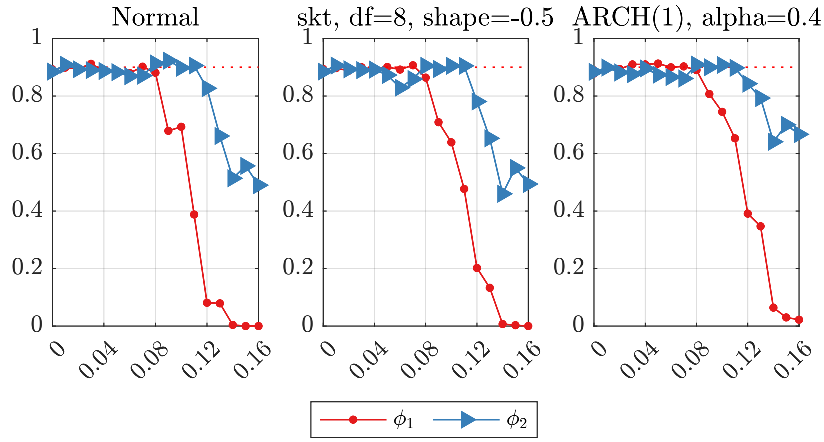

Figure 3: Empirical coverage by parameter and shock type. The horizontal axis gives the standard deviation of the random coefficient δ and the vertical axis gives the empirical coverage for 95% confidence intervals.

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 3, Figure 3 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.

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.

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

	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: 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.

the 1968-1980 and 1980-1992 subsamples.[7]

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 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 potentially miss it in this context. Rescaling this paper's heterogeneity test statistic (see Remark 2 of Section 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.

[7] 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.

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 is substantive. 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.[8]

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 heterogeneity 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

[8]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_{it}^m divided by the revenue share of materials

$$\mu_{it} = \frac{\beta_{it}^m}{s_{it}^m}. \tag{12}$$

While s_{it}^m is taken from the data, β_{it}^m 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

$$y_{it} = \beta_j^k k_{it} + \beta_j^\ell \ell_{it} + \beta_j^m m_{it} + \omega_{it} + \varepsilon_{it} \text{ and } \omega_{it} = \rho_{j0} + \rho_{j1}\omega_{it-1} + \zeta_{it}. \tag{13}$$

The elasticity β_j^m is used to estimate markups according to Equation 12. 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

$$y_{it} = \phi_t(\ell_{it}, m_{it}, k_{it}) + \varepsilon_{it} \tag{14}$$

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.[9] For the second set, Equations 13 and 14 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:

$$0 = \mathbb{E}\left[\zeta_{it}(1,\, k_{it},\, \ell_{it-1},\, m_{it-1},\, \phi_{t-1}(\ell_{it-1}, m_{it-1}, k_{it-1}),\, \beta^k k_{it-1},\, \beta^\ell \ell_{it-2},\, \beta^m m_{it-2})\right]. \tag{15}$$

The first five moments in Equation 15 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 15, 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^m_{it}. 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

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

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 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.

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 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 2 lists the ISIC4 categories that reject the J test at 5% (see Tables 3 and 4 found in Appendix B.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 vehicles) that potentially don't satisfy the requirement of input flexibility—like through unions[10] 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 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 3 and 4 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 2 lists the ISIC4 categories that fail to reject both the J test and

[10]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.

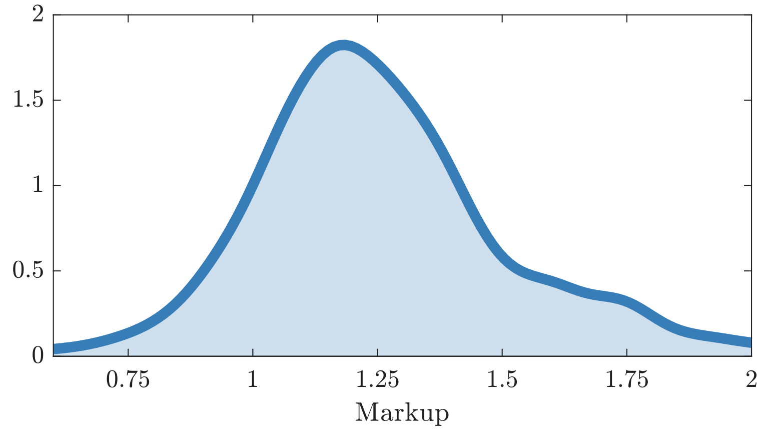

Figure 4: Estimated sales-weighted distribution of plant-level markups μ_{it} for plants belonging to categories listed in Column 3 of Table 2.

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 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 2, Figure 4 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.[11]

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

[11] 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.

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 12, 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.

7 Conclusion

This paper proposes a test for 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.

A Theoretical results

Section A.1 contains matrix results used in subsequent proofs. Section A.2 contains the assumptions and results for the test of multivariate heterogeneity. Section A.3 contains results on semiparametric efficiency.

A.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 $\mathbf{M_W} = \mathbf{I} - \mathbf{G}(\mathbf{G}'\mathbf{W}\mathbf{G})^{-1}\mathbf{G}'\mathbf{W}$. Let † be the generalized matrix inverse.

Proposition A.1. *$\mathbf{M}'_\mathbf{W}(\mathbf{M_W}\mathbf{\Sigma}\mathbf{M_W}')^\dagger\mathbf{M_W}$ is invariant to the choice of $\mathbf{W}$.*

Proof. Begin by showing $\mathrm{col}(\mathbf{M}) = \mathrm{col}(\mathbf{G})$. From the rank-nullity theorem, it suffices to show that $\mathrm{null}(\mathbf{M}') = \mathrm{null}(\mathbf{G}')$. Define projection matrix $\mathbf{P_W} = \mathbf{G}(\mathbf{G}'\mathbf{W}\mathbf{G})^{-1}\mathbf{G}'\mathbf{W}$ 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, $\mathrm{null}(\mathbf{M_W}') = \mathrm{null}(\mathbf{G}')$ and $\mathrm{col}(\mathbf{M_W}) = \mathrm{col}(\mathbf{G})$.

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

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

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

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

□

Proposition A.2. *For duplication matrix $\mathbf{D}$, let $\mathbf{H}$ and $\mathbf{\Lambda}$ be matrices such that*

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

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

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

Proposition A.1 implies $\mathbf{M_W}'(\mathbf{M_W \Sigma M_W}')^\dagger \mathbf{M_W}$ is invariant to the choice of $\mathbf{W}$, so $\mathbf{\Omega_W}$ is also invariant to $\mathbf{W}$.

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

$$(\mathbf{D'D})(\mathbf{MH})'\mathbf{\Sigma}^{-1}\mathbf{X} \sim \mathcal{N}(-\frac{1}{2}(\mathbf{D'D})(\mathbf{MH})'\mathbf{\Sigma}^{-1}\mathbf{MH}(\mathbf{D'D})\text{vech}(\mathbf{\Lambda}), (\mathbf{D'D})(\mathbf{MH})'\mathbf{\Sigma}^{-1}\mathbf{MH}(\mathbf{D'D})). \tag{16}$$

Also,

$$(\mathbf{D'D})\mathbf{H}'\mathbf{\Sigma}^{-1}\mathbf{X} \sim \mathcal{N}(-\frac{1}{2}(\mathbf{D'D})\mathbf{H}'\mathbf{\Sigma}^{-1}\mathbf{MH}(\mathbf{D'D})\text{vech}(\mathbf{\Lambda}), (\mathbf{D'D})\mathbf{H}'\mathbf{\Sigma}^{-1}\mathbf{MH}(\mathbf{D'D})). \tag{17}$$

The distributions of Equations 16 and 17 are identical since $\mathbf{\Sigma}^{-1}\mathbf{M} = \mathbf{M}'\mathbf{\Sigma}^{-1}\mathbf{M}$. ∎

A.2 Framework and results for local parameter heterogeneity

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

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

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

i) *For* $r \in \{1, \ldots, m\}$, $\mathbb{E}[\sup_{\boldsymbol{\theta}\in\mathbf{\Theta}} \|\frac{\partial g_r(\mathbf{x}, \mathbf{f}(\mathbf{x}, \boldsymbol{\varepsilon}; \boldsymbol{\theta}); \boldsymbol{\theta})}{\partial\boldsymbol{\theta}'}\|]$ *and* $\mathbb{E}[\sup_{\boldsymbol{\theta}\in\mathbf{\Theta}} \|\frac{\partial^2 g_r(\mathbf{x}, \mathbf{f}(\mathbf{x}, \boldsymbol{\varepsilon}; \boldsymbol{\theta}); \boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\|]$ *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 3.1

Proof. Throughout, Assumption 3.2 allows for differentiability of the moment function and generating model. Let the r'th moment condition be $\mathbf{g}_r(\mathbf{x}, \mathbf{y}; \boldsymbol{\theta})$. Then, consider the moment's Taylor expansion about $\boldsymbol{\theta}^*$:

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

Using Assumption 3.2, 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

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

The first parenthetical term of the above display is $o_p(1)$ from Chebyshev's inequality, Condition (ii) of Assumption A.1, and Condition ii) of Assumption A.2. Then, the above display can be written as

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

The second term of the above display simplifies to

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

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.2. The desired result follows from stacking moment function $\mathbf{g}_r(\mathbf{Z}_i, \boldsymbol{\theta}^*)$.

□

Lemma A.2. *Impose Assumptions 3.2, A.1, and A.2. Then,*

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

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

Proof. For the result in the first display, consider the moment condition $\mathbb{E}[g_r(\mathbf{f}(\boldsymbol{\varepsilon}_i, \boldsymbol{\theta}^*), \boldsymbol{\theta}^*)] = 0$. Computing the total derivative with respect to $\boldsymbol{\theta}$ and evaluating at $\boldsymbol{\theta} = \boldsymbol{\theta}^*$, Condition i) of Assumption A.2, Assumption 3.2, and the multivariate chain rule imply

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

The result in the second display follows analogously. Note that the terms corresponding to the cross partial derivative of the moment function g_r with respect to $\mathbf{y}$ and $\boldsymbol{\theta}$ equal zero from Condition ii) of Assumption 3.2. □

Proposition A.3. *Impose Assumptions 3.1, 3.2, A.1, and A.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 3.2 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

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

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

□

Proof of Proposition 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 3.2 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 and Assumption 3.2), 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}} \|\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}}]\| = \\
&\quad \sup_{\boldsymbol{\theta}\in\mathcal{N}} \|\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}}]\| \mid_{\boldsymbol{\tau}=\widetilde{\boldsymbol{\theta}}} \\
&\leq \quad \sup_{\boldsymbol{\theta}\in\mathcal{N}} \|\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}}]\| \\
&\quad + \sup_{\boldsymbol{\theta}\in\mathcal{N}} \|\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}]\| \\
&\quad + \sup_{\boldsymbol{\theta}\in\mathcal{N}} \|\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}]\| \mid_{\boldsymbol{\tau}=\widetilde{\boldsymbol{\theta}}} .
\end{aligned}
$$

The first term after the inequality is $o_p(1)$ from Condition (v) of Assumption A.1. The second and third terms after the inequality are $o_p(1)$ from Condition iv) of Assumption A.2, Chebyshev's inequality, and Condition (ii) of Assumption A.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 3.1,

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

□

Proof of Proposition 3.2

Proof. Throughout, Assumption 3.2 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

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

The second line follows from Proposition 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.2, Chebyshev's inequality, and Condition (ii) of Assumption A.1. □

Lemma A.3. *Impose Assumptions 3.1, 3.2, A.1, and A.2. Then,*

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

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

Proof. Proposition A.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 3.2 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. The second term after the inequality is $o_p(1)$ from Condition vi) of Assumption A.2, Chebyshev's inequality, and Condition (ii) of Assumption A.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. The second term after the inequality is $o_p(1)$ from Chebyshev's inequality, Condition (ii) of Assumption A.1, and Condition iii) of Assumption A.2. Analogously, the third term after the inequality is $o_p(1)$ from Chebyshev's inequality, Condition (ii) of Assumption A.1, and Conditions iv) and v) of Assumption A.2.

□

A.3 Semiparametric

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

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

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

Proof of Lemma 4.1

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

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

The first equality follows from Condition i) of Assumption A.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 $\int \frac{\partial}{\partial t} p_{u,t}(\mathbf{u}) du = 0$.

$\square$

Lemma A.4. *Impose Assumptions A.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.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.5. *Impose Assumptions A.4 and A.1. 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 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

$$\mathbf{h} - \Pi(\mathbf{h}|\dot{\mathcal{P}}) = \tag{18}$$

$$\iint \mathbf{h}\mathbf{g}(\mathbf{x},\mathbf{y};\boldsymbol{\theta}^*)' p_{xy}(\mathbf{x},\mathbf{y};\boldsymbol{\theta}^*) d\mathbf{x}d\mathbf{y} \left[\iint \mathbf{g}(\mathbf{x},\mathbf{y};\boldsymbol{\theta}^*)\mathbf{g}(\mathbf{x},\mathbf{y};\boldsymbol{\theta}^*)' p_{xy}(\mathbf{x},\mathbf{y};\boldsymbol{\theta}^*) d\mathbf{x}d\mathbf{y} \right]^{-1} \mathbf{g}(\mathbf{x},\mathbf{y};\boldsymbol{\theta}^*). \tag{19}$$

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 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 19.

□

Lemma A.6. *Impose Assumptions 3.2, 3.3, A.4, A.1, and A.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* $\mathrm{vech}(\boldsymbol{\Lambda})$ *is* $\widetilde{\mathbf{I}} = (\mathbf{D}'\mathbf{D})\mathbf{H}'\boldsymbol{\Sigma}^{-1}\mathbf{M}\mathbf{H}(\mathbf{D}'\mathbf{D})/4$. *The efficient influence function for* $\mathrm{vech}(\boldsymbol{\Lambda})$ *is*

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

Proof. Begin with computing $\iint \boldsymbol{\ell}(\mathbf{x},\mathbf{y})\mathbf{g}(\mathbf{x},\mathbf{y})' p_{xy}(\mathbf{x},\mathbf{y};\boldsymbol{\theta}^*) d\mathbf{x}d\mathbf{y}$. The moment condition and Condition iii) of Assumption A.4 implies

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

Next fix $\mathbf{u} = \mathbf{1}_{p\times 1}$ so $\mathbf{U} = \mathbf{I}_p$. Then, Condition iv) of Assumption A.4 implies

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

Equation 2 and Condition ii) of Assumption A.2 implies that the integral can equivalently be expressed as

$$\begin{aligned}
&\frac{\partial^2}{\partial\boldsymbol{\sigma}\partial\boldsymbol{\sigma}'}\iint g_r(\mathbf{x},\mathbf{f}(\mathbf{x},\boldsymbol{\varepsilon};\boldsymbol{\theta}^*+\boldsymbol{\sigma});\boldsymbol{\theta}^*)p_{x\varepsilon}(\mathbf{x},\boldsymbol{\varepsilon})d\boldsymbol{\varepsilon}d\mathbf{x}\ |_{\boldsymbol{\sigma}=\mathbf{0}} \\
&\quad= \iint\Big(\frac{\mathbf{f}(\mathbf{x},\boldsymbol{\varepsilon};\boldsymbol{\theta}^*+\boldsymbol{\sigma})}{\partial\boldsymbol{\theta}'}\Big)'\frac{\partial^2 g_r(\mathbf{x},\mathbf{f}(\mathbf{x},\boldsymbol{\varepsilon};\boldsymbol{\theta}^*+\boldsymbol{\sigma});\boldsymbol{\theta}^*)}{\partial\mathbf{y}\partial\mathbf{y}'}\Big(\frac{\mathbf{f}(\mathbf{x},\boldsymbol{\varepsilon};\boldsymbol{\theta}^*+\boldsymbol{\sigma})}{\partial\boldsymbol{\theta}'}\Big)p_{x\varepsilon}(\mathbf{x},\boldsymbol{\varepsilon})d\boldsymbol{\varepsilon}d\mathbf{x} \\
&\quad+ \iint\frac{\partial}{\partial\boldsymbol{\tau}}\Big[\frac{\partial g_r(\mathbf{x},\mathbf{f}(\mathbf{x},\boldsymbol{\varepsilon};\boldsymbol{\theta}^*+\boldsymbol{\sigma});\boldsymbol{\theta}^*)}{\partial\mathbf{y}'}\Big(\frac{\mathbf{f}(\mathbf{x},\boldsymbol{\varepsilon};\boldsymbol{\theta}^*+\boldsymbol{\tau})}{\partial\boldsymbol{\theta}'}\Big)\Big]p_{x\varepsilon}(\mathbf{x},\boldsymbol{\varepsilon})d\boldsymbol{\varepsilon}d\mathbf{x}\ |_{\boldsymbol{\sigma}=\mathbf{0},\boldsymbol{\tau}=0} \\
&\quad= -\iint\frac{\partial^2\mathbf{g}_r(\mathbf{x},\mathbf{f}(\mathbf{x},\boldsymbol{\varepsilon};\boldsymbol{\theta}^*);\overline{\boldsymbol{\tau}})}{\partial\overline{\boldsymbol{\tau}}\partial\overline{\boldsymbol{\tau}}'}p_{x\varepsilon}(\mathbf{x},\boldsymbol{\varepsilon})d\mathbf{x}d\boldsymbol{\varepsilon}\ |_{\overline{\boldsymbol{\tau}}=\boldsymbol{\theta}^*}\ .
\end{aligned}$$

The final line follows from Lemma A.2. 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 $\iint\boldsymbol{\ell}(\mathbf{x},\mathbf{y})\mathbf{g}(\mathbf{x},\mathbf{y};\boldsymbol{\theta}^*)'p_{xy}(\mathbf{x},\mathbf{y};\boldsymbol{\theta}^*)d\mathbf{x}d\mathbf{y} = \begin{bmatrix}-\mathbf{G}'\\ -\frac{1}{2}(\mathbf{D}'\mathbf{D})\mathbf{H}'\end{bmatrix}$.

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

$$\begin{aligned}
&\iint\widetilde{\boldsymbol{\psi}}(\mathbf{x},\mathbf{y})(\mathbf{a}'\boldsymbol{\ell}(\mathbf{x},\mathbf{y})+\eta(\mathbf{x},\mathbf{y}))p_{xy}(\mathbf{x},\mathbf{y};\boldsymbol{\theta}^*)d\mathbf{x}d\mathbf{y} \\
&\quad= -\frac{1}{2}\widetilde{\mathbf{I}}^{-1}(\mathbf{D}'\mathbf{D})\mathbf{H}'\boldsymbol{\Sigma}^{-1}\mathbf{M}\Big[\iint\mathbf{g}(\mathbf{x},\mathbf{y};\boldsymbol{\theta}^*)\boldsymbol{\ell}(\mathbf{x},\mathbf{y})'p_{xy}(\mathbf{x},\mathbf{y};\boldsymbol{\theta}^*)d\mathbf{x}d\mathbf{y}\Big]\mathbf{a} \\
&\quad= \frac{1}{2}\widetilde{\mathbf{I}}(\mathbf{D}'\mathbf{D})\mathbf{H}'\boldsymbol{\Sigma}^{-1}\underbrace{\mathbf{M}\mathbf{G}}_{=\mathbf{0}}\mathbf{c}+\frac{1}{4}\widetilde{\mathbf{I}}^{-1}(\mathbf{D}'\mathbf{D})(\mathbf{M}\mathbf{H})'\boldsymbol{\Sigma}^{-1}\mathbf{H}(\mathbf{D}'\mathbf{D})\text{vech}(\boldsymbol{\Lambda}) = \text{vech}(\boldsymbol{\Lambda}).
\end{aligned}$$

The efficient information matrix and score function for vech($\boldsymbol{\Lambda}$) immediately follow.

□

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

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

where the second equality follows from the form of the efficient influence function (see Lemma A.6). 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 $(\mathbf{c}_1', \boldsymbol{\lambda}_1')'\boldsymbol{\ell} + \eta_1$. Assume that $\mathbb{E}[((\mathbf{c}_1', \boldsymbol{\lambda}_1')'\boldsymbol{\ell} + \eta_1)^2] = 1$ to ease notation.

   Define the orthonormal base $\mathcal{G} = ((\mathbf{c}_1', \boldsymbol{\lambda}_1')'\boldsymbol{\ell} + \eta_1, \ldots, (\mathbf{c}_k', \boldsymbol{\lambda}_k')'\boldsymbol{\ell} + \eta_k)'$ of an arbitrary finite-dimensional subspace of the tangent space. Note that the first element of the orthonormal base is $(\mathbf{c}_1', \boldsymbol{\lambda}_1')'\boldsymbol{\ell} + \eta_1$. Lemma 25.14 of van der Vaart (1998) implies that for any $(\mathbf{c}', \boldsymbol{\lambda}')'\boldsymbol{\ell} + \eta \in \operatorname{lin}\mathcal{G}$, $P_{t,(\mathbf{c}',\boldsymbol{\lambda}')'\boldsymbol{\ell}+\eta}$ 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,

   $$(P^n_{h/\sqrt{n},\mathbf{b}'\mathcal{G}} \colon h > 0, \mathbf{b} \in S^{k-1}) \rightsquigarrow (N_m(h\mathbf{b}, \mathbf{I}) \colon h > 0, \mathbf{b} \in S^{k-1}, \operatorname{vech}^{-1}((\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_k)\mathbf{b}) \in \mathcal{C}).$$

   The last restriction restricts $\mathbf{b}$ so that $\mathbf{b}'\mathcal{G}$ lies within the tangent set. For score function $(\mathbf{c}_1', \boldsymbol{\lambda}_1')'\boldsymbol{\ell} + \eta_1$ and $h = h_1$, fix a subsequence for which $\limsup_{n\to\infty} \pi_n(P_{h_1 n^{-1/2},(\mathbf{c}_1',\boldsymbol{\lambda}_1')'\boldsymbol{\ell}+\eta_1})$ 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

   $$\mathbf{Z} \sim \mathcal{N}(\boldsymbol{\lambda}_1, 4\boldsymbol{\Omega}^{-1}) \tag{20}$$

   where $\mathbb{E}[\widetilde{\boldsymbol{\psi}}\widetilde{\boldsymbol{\psi}}'] = 4\boldsymbol{\Omega}^{-1}$ from Lemma A.6. $\pi(\boldsymbol{\lambda}_1)$ 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 $\limsup_{n\to\infty} \pi_n(P_{h/\sqrt{n},(\mathbf{c}',\boldsymbol{\lambda}')'\boldsymbol{\ell}+\eta}) \leq \pi(\boldsymbol{\lambda})$, 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 $P_{1/\sqrt{n},(\mathbf{c}',\boldsymbol{\lambda}')'\boldsymbol{\ell}+\eta}$ to $\mathcal{N}(\boldsymbol{\lambda}, \mathbb{E}[\widetilde{\boldsymbol{\psi}}\widetilde{\boldsymbol{\psi}}'])$, which matches the limit experiment. Since $\mathbf{Z}_n'(\widehat{\boldsymbol{\Omega}}/4)\mathbf{Z}_n - \inf_{\mathrm{vech}^{-1}(\boldsymbol{\lambda})\in\mathcal{C}}\Big[\mathbf{Z}_n - \boldsymbol{\lambda}\Big]'(\widehat{\boldsymbol{\Omega}}/4)\Big[\mathbf{Z}_n - \boldsymbol{\lambda}\Big] = \mathbf{S}_n'\widehat{\boldsymbol{\Omega}}^{-1}\mathbf{S}_n - \inf_{\mathrm{vech}^{-1}(\boldsymbol{\lambda})\in\mathcal{C}}\Big[\mathbf{S}_n + \frac{1}{2}\widehat{\boldsymbol{\Omega}}\boldsymbol{\lambda}\Big]'\widehat{\boldsymbol{\Omega}}^{-1}\Big[\mathbf{S}_n + \frac{1}{2}\widehat{\boldsymbol{\Omega}}\boldsymbol{\lambda}\Big]$, 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 20 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

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

   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 $\mathbf{Z}$)

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

□

Corollary A.1. *Impose Assumptions 3.2, 3.3, A.4, A.1, and A.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 10 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,*

$$\lim_{n\to\infty} \mathbb{P}_{\Lambda,n^{-1/2}}(\phi_n = 1) = 1 - \Phi\left(z_{1-\alpha} - \frac{\Lambda}{\sqrt{\widetilde{I}^{-1}}}\right). \tag{21}$$

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 4.1 to the scalar case. □

B Additional empirical results

B.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 3: 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 elasticitiy 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 4: 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 elasticitiy 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.

References

Ackerberg, D. A., Caves, K., and Frazer, G. (2015). Identification Properties of Recent Production Function Estimators. *Econometrica*, 83(6):2411–2451.

Ahn, S. C. and Schmidt, P. (1995). Efficient estimation of models for dynamic panel data. *Journal of Econometrics*, 68(1):5–27.

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.

Andrews, D. W. (1998). Hypothesis testing with a restricted parameter space. *Journal of Econometrics*, 84(1):155–199.

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.

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.

Bartholomew, D. J. (1959). A Test Of Homogeneity for Ordered Alternatives. *Biometrika*, 46(1/2):36.

Beach, R. H., Bullock, A. M., Heller, K. B., and Domanico, J. L. (2000). Lime Production: Industry Profile. Technical report, Environmental Protection Agency.

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.

Browning, M., Ejrnaes, M., and Alvarez, J. (2010). Modelling Income Processes with Lots of Heterogeneity. *Review of Economic Studies*, 77(4):1353–1381.

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.

Chernoff, H. (1954). On the Distribution of the Likelihood Ratio. *The Annals of Mathematical Statistics*, 25(3):573–578.

Chesher, A. (1984). Testing for Neglected Heterogeneity. *Econometrica*, 52(4):865.

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.

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.

Fernald, J., Gandhi, A., Ruzic, D., and Traina, J. (2025). Micro and Macro Perspectives on Production-Based Markups.

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.

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.

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.

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.

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.

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.

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.

Lee, W. (2022). Identification and estimation of dynamic random coefficient models.

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.

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.

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.

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.

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.

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.

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.

Staiger, D. and Stock, J. H. (1997). Instrumental Variables Regression with Weak Instruments. *Econometrica*, 65(3):557–586.

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.