The purpose of an estimator is what it does: Misspecification, estimands, and over-identification

Isaiah Andrews, Jiafeng Chen, and Otavio Tecchio

Abstract. In over-identified models, misspecification—the norm rather than exception—fundamentally changes what estimators estimate. Different estimators imply different estimands rather than different efficiency for the same target. A review of recent applications of generalized method of moments in the *American Economic Review* suggests widespread acceptance of this fact: There is little formal specification testing and widespread use of estimators that would be inefficient were the model correct, including the use of "hand-selected" moments and weighting matrices. Motivated by these observations, we review and synthesize recent results on estimation under model misspecification, providing guidelines for transparent and robust empirical research. We also provide a new theoretical result, showing that Hansen's J-statistic measures, asymptotically, the range of estimates achievable at a given standard error. Given the widespread use of inefficient estimators and the resulting researcher degrees of freedom, we thus particularly recommend the broader reporting of J-statistics.

Date: September 30, 2025.

Andrews: Department of Economics, Massachusetts Institute of Technology, and NBER, iandrews@mit.edu; Chen: Department of Economics, Stanford University, jiafeng@stanford.edu; Teccio: Department of Economics, Massachusetts Institute of Technology, otecchio@mit.edu.

1. **Introduction**

Scientific and policy questions in economics frequently hinge on quantities such as counterfactual outcomes or effects on welfare. Recovering them often requires strong modeling assumptions. At the same time, many economists concede that economic models are at best approximately true (White, 1982; Hansen and Sargent, 2008; Manski, 2013; Rodrik, 2015). This leads to a gap between formal econometric recommendations and empirical practice for over-identified models.[1]

Many textbook treatments of over-identified models (e.g. Hayashi, 2000; Stock and Watson, 2018; Davidson and Mackinnon, 2021) recommend researchers test whether the model's over-identifying restrictions hold and use an efficient estimator when failing to reject. Experience reading economics papers and attending economics seminars, however, suggests that empirical practice often flouts these prescriptions. To document this, we examined empirical papers published in the *American Economic Review* between 2020 and 2024 which fit models using the generalized method of moments (GMM, Hansen, 1982) or related methods, obtaining a sample of 36 papers, of which 22 are over-identified. In this sample, we find that researchers commonly fit models using only a subset of the model's implications, for instance fitting a fully parametric model by moment-matching rather than maximum likelihood. Researchers also commonly weight moments differently than would the optimal estimator (e.g. 14 of the over-identified papers in our sample report no efficiently-weighted specifications). Both practices yield inefficient estimators under correct model specification. Formal tests of over-identifying restrictions are notably scarce (only 3 papers in our sample report J-tests of overidentifying restrictions), though various "eyeball tests" are common, with researchers e.g. plotting the gap between data- and model-implied moments (25 papers in our sample).

This gap between econometrics as taught and econometrics as practiced could appear puzzling if one thought economic models were plausibly correct—after all, it would be better to use a correct model, and to use the best estimator under that model, than to persist

[1]That is, models which admit different estimators for the same parameter and consequently imply testable restrictions on the data distribution, c.f. Chen and Santos 2018.

in using an incorrect model or an inefficient estimator. The common practices outlined above are, on the other hand, reasonable under the premise that any model we can take to the data is almost certainly wrong. Consistent with this premise, many papers in our data show evidence of substantial misspecification, with 8 papers reporting *just-identified* specifications where the parameter estimates cannot set the moments to zero. Under this view, it is natural to fit the model aiming for the "least-bad" answer. In judging model fit, it is moreover natural to draw quantitative judgments about where and how well the model fits, rather than focusing on the all-or-nothing question of whether the model fits exactly.

A similar perspective underlies a large and active recent literature in econometrics which studies the behavior and interpretation of popular estimators without assuming that the model is correct. This article provides a selective synthesis of this literature. While one can reasonably critique some of this literature as a reverse-engineering exercise, starting from estimators and working backwards in the hope of finding interpretable estimands (Mogstad and Torgovitsky, 2024), the perspective of this literature can also reasonably be summarized as "the purpose of an estimator is what it does."[2] That is, given that economic models are generally misspecified, it is important to understand how estimators actually behave in applications. Moreover, in our view, it can be reasonable to choose among estimators on this basis.

While reasonable, the common practice of choosing estimators based on misspecification concerns increases researcher degrees of freedom (Simmons et al., 2011). Different researchers estimating the same economic model on the same data may reach different conclusions due to differing misspecification concerns and subsequent estimation choices. Worse, dishonest researchers could use misspecification concern as a pretext for estimation choices which deliver a desired result, e.g. rejection of some null hypothesis.

As a first step towards quantifying the scope for such manipulation, and motivated by the wide range of weight-selection approaches observed in our survey of AER papers, we consider the choice of weight matrices for a given (over-identifying) set of moments. In

[2]Paraphrased from Beer (1985)—"The purpose of a system is what it does. There is, after all, no point in claiming that the purpose of a system is to do what is consistently fails to do."

models where the degree of misspecification is on the same order as sampling uncertainty (i.e. locally misspecified models) we prove that, asymptotically, the set of estimates obtainable at a given standard error is an interval centered at the efficient GMM estimate, with width proportional to the square root of Hansen (1982)'s J-statistic. Moreover, we show that a nefarious researcher can choose weights to engineer a t-statistic value of at least $\sqrt{J}$ for any null hypothesis. Thus, given a set of moments, J-statistics measure the scope for a nefarious researcher to "weight hack" their findings.

To summarize our conclusions, we recommend the following:

(1) Researchers should distinguish concerns and diagnostics for correct *econometric* specification (i.e. does the model correctly describe the relationship between the observable data and the objects of economic interest) vs. correct *statistical* specification (i.e. does the model fit the data distribution).

(2) Researchers should be explicit about what their choice of estimator implies for their estimand. If they use estimators which are inefficient under the model (e.g. due to non-optimal weighting), they should articulate why they believe such estimators are preferable.

(3) Researchers should use misspecification-robust standard errors, which excludes the usual GMM standard errors (Hansen, 1982) in over-identified models. Formulas are available for many problems, and when in doubt the bootstrap will often work.

(4) Researchers estimating over-identified moment condition models should report J-statistics, since they summarize the range of results may be obtained via alternative weighting choices. J-statistics are thus important to report even if one dispenses with J-tests.

The following sections develop these points in detail, highlighting relevant references from the literature as we proceed. Section 2 defines what we mean by econometric and statistical models, illustrating with examples, and introduces corresponding notions of misspecification. Section 3 defines the estimands resulting from given estimators, and reviews some important

estimand characterization results. Section 4 discusses misspecification-robust standard errors. Finally, Section 5 discusses the idea of local misspecification and derives what is to the best of our knowledge a new result, relating J-statistics to the range of estimates attainable under a standard error bound. Appendix A details a normal limit problem used to prove our J-statistic result, while Appendix B collects proofs for results stated in the main text.

2. What Do We Mean by "Misspecified?"

Suppose we observe a sample of n observations $\{X_1, ..., X_n\}$ drawn from a distribution P.[3] We are interested in using the data to learn an *economic parameter* $\theta \in \Theta$, such as a counterfactual outcome or optimal policy. For the data to convey information about θ, we must restrict which parameters can accompany a given data distribution. These restrictions could come from economic theory, from empirical evidence in other settings, or from ad-hoc restrictions chosen for analytical convenience. These restrictions imply an *econometric model* $\mathcal{M} \subseteq \Theta \times \Delta(\mathcal{X})$, for $\Delta(\mathcal{X})$ the set of distributions on the sample space $\mathcal{X}$. The model $\mathcal{M}$ collects pairs of (θ, P) that are consistent with our restrictions. $\mathcal{M}$ fully describes the relationship between the data distribution and the quantity of interest, and so determines the identification status of the parameter.[4] Specifically, the identified set for θ under data distribution P is the set of parameter values consistent with P: $\Theta_I(P) = \{\theta \in \Theta : (\theta, P) \in \mathcal{M}\}$. θ is point-identified if and only if $\Theta_I(P)$ is a singleton for all P consistent with $\mathcal{M}$.

We call the set of all P consistent with $\mathcal{M}$ the *statistical model*:

$$\mathcal{P} = \{P \in \Delta(\mathcal{X}) : \text{ there exists } \theta \in \Theta \text{ such that } (\theta, P) \in \mathcal{M}\}.$$

[3]Formally, we take P to be a distribution over the infinite sequence $X^\infty = \{X_1, X_2, ...\}$ and suppose the sample is the first n elements. Thus, our analysis covers both settings with iid data (in which case P is a product measure) and cases with dependent data, subject to restrictions we note below.

[4]While we discuss $\mathcal{M}$ as an econometric model given our focus on the economics literature, one might prefer another term, e.g. substantive model, since the same points arise when studying causal or counterfactual questions outside of economics. Spanos (2016) discusses related points under the heading of substantive adequacy.

Provided the parameter θ is point-identified under $\mathcal{M}$, we can define a *statistical parameter* $\theta^* : \mathcal{P} \to \Theta$ such that $\theta^*(P) = \theta$ for all $(\theta, P) \in \mathcal{M}$. By construction, θ^* is point-identified as well: $\Theta_I(P) = \{\theta^*(P)\}$ for all $P \in \mathcal{P}$.

Example 2.1 (Linear Causal Model)**.** Suppose we observe $X_i = (Y_i, W_i)$ drawn iid according to P. Here $Y_i \in \mathbb{R}$ is a scalar outcome while $W_i \in \mathcal{W} \subseteq \mathbb{R}$ is a continuous treatment. We believe that (i) each individual has potential outcome $Y_i(\cdot)$, where the observed outcome for individual i is equal to that individual's potential outcome evaluated at realized treatment, $Y_i = Y_i(W_i)$; that (ii) treatment is randomly assigned with W_i independent of $Y_i(\cdot)$; and that (iii) the potential outcomes are linear in w,

$$Y_i(w) = \alpha_i + \beta_i \cdot w.$$

This is a special case of the random coefficient models studied by e.g. Beran and Hall (1992).[5] Our parameter of interest is the average derivative $\theta = \mathbb{E}\left[\frac{\partial}{\partial w} Y_i(W_i)\right]$, which describes the average effect of a marginal increase in W_i. Under the model this average derivative can be written as $\mathbb{E}[\beta_i]$.

Hence, under an additional regularity condition on a variance, which we shorthand with $\Xi = \mathrm{Var}(W_i, \alpha_i, \beta_i)$, our econometric model is

$$\mathcal{M} = \left\{ \left(\mathbb{E}[\beta], \bigtimes_{i=1}^{\infty} \mathcal{L}(\alpha + \beta W, W) \right) : (\alpha, \beta) \perp\!\!\!\perp W, \det(\Xi) \in (0, \infty) \right\}$$

where we use $\mathcal{L}(A_i)$ to denote the law of a random variable A_i. The statistical model collects the corresponding data distributions,

$$\mathcal{P} = \left\{ \bigtimes_{i=1}^{\infty} \mathcal{L}(\alpha + \beta W, W) : (\alpha, \beta) \perp\!\!\!\perp W, \det(\Xi) \in (0, \infty) \right\}.$$

If the econometric model is correct, least squares regression of Y_i on W_i recovers θ, so we can define the statistical parameter as $\theta^*(P) = \frac{\mathrm{Cov}_P(W_i, Y_i)}{\mathrm{Var}_P(W_i)}$. ■

[5]Note that if we take $W_i \in \{0, 1\}$, rather than continuous, then this example is simply the potential outcomes model, $\alpha_i = Y_i(0)$ and $\beta_i = Y_i(1) - Y_i(0)$.

Example 2.2 (Generalized Method of Moments)**.** Suppose the distribution of $(X_1, \ldots, X_n)$ under P is stationary. Suppose the parameter of interest θ is, under the model, some function of a full set of model parameters $\psi \in \Psi \subseteq \mathbb{R}^p$, $\theta = \vartheta(\psi)$ for a known function $\vartheta(\cdot)$. Our model implies that a vector of moment conditions

$$\mathbb{E}_P[g(X_i, \psi)] = 0 \in \mathbb{R}^k$$

holds at the true value of ψ.

Let the econometric model $\mathcal{M}$ denote the set (θ, P) consistent with the moment restrictions, along with additional regularity conditions (e.g., finite moments, stationarity) encoded by a set $\mathcal{P}^R$:

$$\mathcal{M} = \left\{(\vartheta(\psi), P) : \psi \in \Psi, P \in \mathcal{P}^R, \mathbb{E}_P[g(X_i, \psi)] = 0\right\}. \tag{1}$$

This econometric model implies a statistical model

$$\mathcal{P} = \left\{P \in \mathcal{P}^R : \text{ there exists } \psi \in \Psi \text{ such that } \mathbb{E}_P[g(X_i, \psi)] = 0\right\}$$

Provided the identified set for ψ is a singleton under each $P \in \mathcal{P}$—which in this context means that the moment equation $\mathbb{E}_P[g(X_i, \psi)] = 0$ is solved at a unique ψ—we can define $\psi^* : \mathcal{P} \to \Psi$ implicitly as the solution to the moment equations, $\mathbb{E}_P[g(X_i, \psi)] = 0$, and can define the statistical parameter as $\theta^*(P) = \vartheta(\psi^*(P))$. ■

So long as (a) we are confident in the specification of our model and (b) the parameter θ is point-identified under $\mathcal{M}$, the distinction between the econometric and statistical models is inconsequential, since the economic parameter θ and the statistical parameter $\theta^*(P)$ always agree. Once we consider misspecification, however, this distinction starts to matter.

We say that the model is *econometrically misspecified* if the true (θ, P) pair falls outside the econometric model, $(\theta, P) \notin \mathcal{M}$. We assume the parameter θ remains well-defined under misspecification, which restricts the sorts of parameters we consider. In particular, our analysis is well-suited to e.g. counterfactual outcomes or causal effects, which may be defined

(if not always estimated) non-parametrically. By contrast, for parameters which are ill-defined outside a specific model (e.g. preference parameters tied to a specific functional form), it makes little sense to contemplate a "true" θ outside of the model. Fortunately, in many economic analyses the objects of ultimate interest (e.g. counterfactuals, welfare, and optimal policies) are well-defined under much weaker assumptions than those often used to conduct estimation.

Analogously, the model is *statistically misspecified* if the observed data distribution rejects it, $P \notin \mathcal{P}$. Since we have thus far defined the statistical parameter θ^* only on the model, $\theta^*(P)$ is undefined under statistical misspecification. Statistical misspecification implies econometric misspecification but not the reverse.

Example 2.3 (Generalized Method of Moments, continued)**.** The moment condition model is econometrically misspecified at (θ, P) if and only if the moment conditions are violated at all values of ψ consistent with the true economic parameter θ, i.e.

$$0 \notin \{E_P[g(X_i, \psi)] : \psi \in \Psi, \vartheta(\psi) = \theta\}.$$

Under this definition, the reason for misspecification $(\theta, P) \notin \mathcal{M}$—incorrect relationship between ψ and P (i.e. the moments g), incorrect relationship between ψ and θ (i.e. the function ϑ), or some other source—is irrelevant. Indeed, since correct specification requires only that our model $\mathcal{M}$ allow the true (θ, P) pair, it is possible for the model to be correctly specified even if our underlying economic assumptions are incorrect and the model parameter ψ does not have a well-defined "true" value. Similarly, the model is statistically misspecified at P if there exists no value of ψ that solves the moment conditions

$$0 \notin \{E_P[g(X_i, \psi)] : \psi \in \Psi\}.$$

Since econometrically correct specification imposes an additional restriction ($\vartheta(\psi) = \theta$) relative to statistically correct specification, we see that the model may sometimes be econometrically, but not statistically, misspecified. In particular, when the model is just-identified with $\dim(g(X_i, \psi)) = k = p = \dim(\psi)$, the moment conditions involve the same number

of equations and unknowns and will often, though not always, have a solution. In over-identified models with $k > p$, by contrast, we have more moment equations than unknowns. ■

Example 2.4 (Linear Causal Model, continued)**.** As noted by e.g. Breunig and Hoderlein (2018), the random coefficient model implies a large collection of over-identifying restrictions. For instance, the model implies that

$$\mathrm{Var}_P(Y_i \mid W_i) = \mathrm{Var}(\alpha_i) + 2W_i \,\mathrm{Cov}(\alpha_i, \beta_i) + W_i^2 \,\mathrm{Var}(\beta_i),$$

and thus that the conditional variance of Y_i given W_i must quadratic in W_i. There are many ways for this (and other over-identifying restrictions) to fail, and thus for the model to be statistically misspecificed, $P \notin \mathcal{P}$.[6]

Despite this rich collection of over-identifying restrictions, there exist forms of econometric misspecification which do not imply statistical misspecification. For instance, there are ways that random assignment may fail which are statistically undetectable but nevertheless distort our conclusions about $\theta = E[\frac{\partial}{\partial w} Y_i(W_i)]$.[7] ■

To formally account for the possibility of misspecification, suppose we are willing to assume that the true (θ, P) pair is contained in a *nesting model* $\mathcal{M}^*$ which satisfies $\mathcal{M} \subset \mathcal{M}^*$, and has a corresponding statistical model $\mathcal{P}^*$. This nesting condition is vacuously satisfied when we consider the maximal nesting model $\mathcal{M}^* = \Theta \times \Delta(\mathcal{X})$ in which all distributions are consistent with all parameter values, but to obtain useful results we will need to impose further restrictions. The econometrics literature uses nesting models for two primary purposes:

(1) To explore what statements can be made about θ under the nesting model $\mathcal{M}^*$. For instance, we could characterize and estimate the identified set for θ under $\mathcal{M}^*$, or

[6]The support of W matters for this conclusion. If we instead consider binary $W_i \in \{0, 1\}$ (and took e.g. $\theta = E[Y_i(1) - Y_i(0)]$, which is again equal to $E[\beta_i]$ under the model) then any P with finite, nonzero $\mathrm{Var}_P(W_i Y_i)$ and $\mathrm{Var}_P((1 - W_i)Y_i)$ is consistent with $\mathcal{P}$.

[7]For instance, it could be that $W_i = 1\{F_{\beta|\alpha}(\beta_i|\alpha_i) > \frac{1}{2}\}\tilde{W}_i$ for a continuously distributed variable $\tilde{W}_i$ independent of (α_i, β_i) and $F_{\beta|\alpha}$ the conditional distribution function for β_i given α_i. Thus W_i is set to zero when β_i exceeds its conditional median given α_i, and the mean of β_i cannot be recovered from the observed data.

derive confidence intervals which ensure coverage for θ under all $(\theta, P) \in \mathcal{M}^*$. In many cases the strength of the conclusions obtained will depend on the details of the nesting model considered, so it is also common to explore sensitivity to the choice of nesting model $\mathcal{M}^*$.

(2) To explore how specific estimators, motivated by the initial model $\mathcal{M}$, perform under the nesting model. This could include interpreting the parameters these estimators recover for $(\theta, P) \in \mathcal{M}^* \setminus \mathcal{M}$ and deriving standard errors and confidence intervals which are valid for inference on these estimands.

This article opts to focus on the latter approach.

2.1. **An aside on the many meanings of models.** Our definition of a statistical model follows a large literature in semiparametric statistics and econometrics, e.g. Chapter 25 of van der Vaart (2000) and Chen and Santos (2018). By contrast, phrases like "economic model" and "econometric model" are used in many different ways in the literature. Here we offer an incomplete survey and briefly relate some other definitions to ours. Our definition is in a sense minimal so long as one wishes to describe the properties of estimators for θ. For instance, if we wish to characterize the set of parameter-value and bias pairs, $(\theta, \mathbb{E}_P\left[\hat{\theta}\right] - \theta)$, implied by any possible estimator $\hat{\theta}$ for θ, it is necessary and sufficient to know the set of possible values for the pair (θ, P). This, however, is precisely our definition of the econometric model.

A classical definition of a structural model is given by Koopmans and Reiersol (1950). They define a *structure* Q as a particular joint distribution over latent and observed variables.[8] A *model* $\mathcal{Q}$ in Koopmans and Reiersol (1950) is then a set of structures. The observed variables follow a distribution $P = P(Q)$, generated from a structure $Q \in \mathcal{Q}$. A *structural parameter* is a functional of the structure $\theta = \theta(Q)$. In our notation, a structural model in the Koopmans and Reiersol (1950) sense implies the econometric model

$$\mathcal{M} = \{(\theta(Q), P(Q)) : Q \in \mathcal{Q}\}.$$

[8]Precisely, Koopmans and Reiersol (1950) define a structure as a distribution of latent variables and an equation linking observed and latent variables.

To permit misspecification, we may consider a nesting model $\mathcal{M}^*$ generated by a larger set of structures $\mathcal{Q}^* \supset \mathcal{Q}$.

Potential-outcomes notation and structural causal models also provide natural nesting models (Neyman, 1923/1990; Rubin, 1974; Pearl, 2009; Richardson and Robins, 2013; Heckman and Pinto, 2015). Many parameters of economic interest are functionals of the joint distribution Q of all potential outcomes and treatments, suitably defined.[9] The distribution of potential outcomes then implies a distribution of the observed data $P(Q)$. Our approach nests this as well: $\mathcal{M} = \{(\theta(Q), P(Q)) : Q \in \mathcal{Q}\}$ constitutes a model in our sense.

These notions are clearer and more versatile than some others. Some textbooks and reviews define an econometric model as deterministic economic relations plus stochastic disturbances (Intriligator, 1983; Woolridge, 2009), borrowing from the simultaneous equations literature. This usage is ambiguous, however, because the error term's meaning is unclear: For instance, the equation

$$Y_i = \alpha + \beta W_i + \epsilon_i$$

may model linear, homogeneous causal effects $Y_i(w) = \alpha + \beta w + \epsilon_i$; or it may simply define β as a statistical functional equal to the population OLS coefficient, leaving the joint distribution of (Y, W) unrestricted. Under our terminology, the two cases correspond to distinct econometric models.[10]

3. Choice of Estimator and Estimand

Consider estimators $\hat{\theta}(X^n) \in \Theta$ that are consistent under the model, i.e., for all $(\theta, P) \in \mathcal{M}$, $\hat{\theta} \xrightarrow{p} \theta$ as $n \to \infty$. Under standard regularity conditions on the nesting model $\mathcal{M}^*$, $\hat{\theta}$

[9] In particular, structural equation models are often generative models for potential outcomes. Relating to Koopmans and Reiersol (1950), one could think of the joint distribution of potential outcomes and treatment variables as the structure.

[10] The former describes a model from homogeneous and linear potential outcomes

$$\mathcal{M} = \left\{ (\beta, P) : \bigtimes_{i=1}^{n} (Y_i(W), W_i) \sim P, Y_i(w_1) - Y_i(w_2) = \beta(w_1 - w_2) \right\}.$$

The latter defines a model in which

$$\mathcal{M} = \left\{ \left(\frac{\mathrm{Cov}_P(W, Y)}{\mathrm{Var}_P(W)}, P \right) : \bigtimes_{i=1}^{n} (Y_i, W_i) \sim P, \ P \text{ has second moments} \right\}.$$

still converges in probability to some limit when $P \in \mathcal{P}^* \setminus \mathcal{P}$. Denote this limit by $\text{plim}(\hat{\theta}, P)$. Because $\hat{\theta}$ is consistent under the model, $\text{plim}(\hat{\theta}, P) = \theta^*(P)$ for every $P \in \mathcal{P}$. However, different choices of estimator will typically imply different limits $\text{plim}(\hat{\theta}, P)$ for P outside the initial model, $P \in \mathcal{P}^* \setminus \mathcal{P}$.[11] Thus, when choosing which estimator to use under the model $\mathcal{M}$, we are also (implicitly) choosing which parameter we wish to estimate when the model is statistically misspecified.

Example 3.1 (Linear Causal Model, continued)**.** Under the model $\mathcal{M}$, we may estimate θ using the ordinary least squares (OLS) estimator $\hat{\theta}_{\text{OLS}} = \frac{\text{Cov}_n(W_i, Y_i)}{\text{Var}_n(W_i)}$, for Cov_n and Var_n the sample covariance and variance, respectively. Alternatively, we may instead estimate θ using the weighted least squares (WLS) estimator $\hat{\theta}_{\text{WLS}} = \frac{\text{Cov}_{n,\omega}(W_i, Y_i)}{\text{Var}_{n,\omega}(W_i)}$, where $\text{Cov}_{n,\omega}$ and $\text{Var}_{n,\omega}$ weight observation i by $\omega_i = \omega(W_i)$ for a weight function ω. By the Gauss–Markov theorem, the efficient estimator uses $\omega_i = \text{Var}_P(Y_i \mid W_i)^{-1}$. Consistency implies that $\text{plim}(\hat{\theta}_{\text{OLS}}, P) = \text{plim}(\hat{\theta}_{\text{WLS}}, P)$ for all $P \in \mathcal{P}$. By contrast, under statistical misspecification $\text{plim}(\hat{\theta}_{\text{OLS}}, P) = \frac{\text{Cov}_P(W_i, Y_i)}{\text{Var}_P(W_i)}$ while $\text{plim}(\hat{\theta}_{\text{WLS}}, P) = \frac{\text{Cov}_{P,\omega}(W_i, Y_i)}{\text{Var}_{P,\omega}(W_i)}$, where these limits can differ for $P \notin \mathcal{P}$. Hence, once we allow for misspecification WLS is not more efficient than OLS: the two simply estimate different things. ■

Example 3.2 (Generalized Method of Moments, continued)**.** Moment condition models are usually estimated by the generalized method of moments(GMM). For $g_n(X^n, \psi) = \frac{1}{n}\sum g(X_i, \psi)$ the sample average moments and $\hat{\Omega} = \Omega_n(X^n)$ a data-dependent weighting matrix, GMM takes $\hat{\theta}_\Omega = \vartheta\left(\hat{\psi}_\Omega\right)$ for

$$\hat{\psi}_\Omega = \hat{\psi}_\Omega(X^n) = \arg\min_{\psi \in \Psi} g_n(X^n, \psi)' \hat{\Omega} g_n(X^n, \psi).$$

For instance, the two-step efficient GMM estimator, which minimizes the large-sample variance of the estimate under the model, sets $\hat{\Omega}$ equal to the inverse of an estimate of the asymptotic variance of $\sqrt{n} g_n(\psi)$.

[11] Indeed, as noted by Chen and Santos (2018) there is a close connection between testability of a statistical model $\mathcal{P}$ and the existence of distinct estimators for statistical functionals defined on the model.

If we define $\Omega_P = \text{plim}(\hat{\Omega}, P)$—assumed to exist—and $g_P(\psi) = \mathbb{E}_P[g(X_i, \psi)]$, then under mild conditions (see Imbens (1997); Hall and Inoue (2003)),

$$\text{plim}(\hat{\psi}_\Omega, P) = \arg\min_{\psi \in \Psi} g_P(\psi)' \Omega_P g_P(\psi) \text{ and } \text{plim}(\hat{\theta}_\Omega, P) = \vartheta(\text{plim}(\hat{\psi}_\Omega, P)).$$

When the model is statistically correct the argmin in the GMM objective does not depend on the weighting matrix, and thus neither does the limit of $\hat{\theta}_\Omega$. In contrast, when the model is statistically misspecified different weighting matrices yield different estimands. Hence, under misspecification the two-step GMM estimator is no more efficient than any other estimator: Once again each weighting leads us to recover a different parameter. ■

Example 3.3 (Parametric models)**.** Previously, the estimators we discussed differed only by how they weighted a common set of objects (W_i values in one case and moments in the other). In other settings, estimators exploit the data in qualitatively different ways. To illustrate, suppose we have a parametric model which implies that the data are drawn from P_ψ, for a finite-dimensional parameter ψ, where $\theta = \vartheta(\psi)$. Assume P_ψ is a continuous distribution with density p_ψ. Our econometric model is thus $\mathcal{M} = \{(\vartheta(\psi), P_\psi) : \psi \in \Psi\}$, while our statistical model is $\mathcal{P} = \{P_\psi : \psi \in \Psi\}$.

Under the model $\mathcal{M}$ we may estimate θ by maximum likelihood, taking $\hat{\psi}_{\text{MLE}} = \arg\max_{\psi \in \Psi} \log p_\psi(X^n)$ and $\hat{\theta}_{\text{MLE}} = \vartheta(\hat{\psi}_{\text{MLE}})$. Alternatively, we may estimate the model by the method of simulated moments, choosing a set of $\mathbb{R}^k$-valued functions h, constructing moments comparing the realized values of $h(X_i)$ to their model-implied means $g(X_i, \psi) = h(X_i) - \mathbb{E}_{P_\psi}[h(X_i)]$, and computing GMM estimates $\hat{\psi}_\Omega$ and $\hat{\theta}_\Omega$ for some choice of weighting matrix. Provided the moment equations have a unique solution under P_ψ for all $\psi \in \Psi$, these estimators have the same probability limit under the model but maximum likelihood will be (weakly) more efficient.

As previously discussed, under misspecification and provided regularity conditions hold, the GMM estimator $\hat{\psi}_\Omega$ converges to the value which minimizes the Ω_P-weighted deviation of the moment conditions from zero. In contrast, $\text{plim}(\hat{\psi}_{\text{MLE}}, P) = \arg\max_{\psi \in \Psi} \mathbb{E}_P[\log p_\psi(X^n)]$,

which minimizes the Kullback–Leibler divergence between the distribution of the data and the model-implied distribution (White, 1982). ■

Once we allow that the model $\mathcal{M}$ is econometrically misspecified, we will usually find that none of the estimands we consider recovers the precise parameter of economic interest θ.[12] Nevertheless, there are numerous results in the literature which characterize the interpretation of specific estimands under particular nesting models. If one of these interpretations is "close" to the parameter of interest θ in a qualitative sense, this may increase the appeal of the corresponding estimator.

Example 3.4 (Linear Causal Model, continued)**.** The results of Yitzhaki (1996) imply that the least-squares estimand $\text{plim}(\hat{\theta}_{OLS}, P)$ can be written as a weighted average of the partial derivative of $\mathbb{E}_P[Y_i \mid W_i]$ with respect to W_i,

$$\text{plim}(\hat{\theta}_{OLS}, P) = \mathbb{E}_P\left[\eta_P(W_i)\frac{\partial}{\partial w}\mathbb{E}_P[Y_i \mid W_i]\right]$$

for particular weights $\eta_P(W_i) \geq 0$ with $\mathbb{E}_P[\eta_P(W_i)] = 1$. If the nesting model $\mathcal{M}^*$ maintains random assignment and differentiability of potential outcomes $Y_i(w)$, it follows that $\text{plim}(\hat{\theta}_{OLS}, P) = \mathbb{E}_P[\eta_P(W_i)\frac{\partial}{\partial w}Y_i(W_i)]$, so OLS recovers a weighted average of causal effects, though typically not the unweighted average specified by θ. One can show an analogous result for weighted least squares, so the choice between the OLS and WLS estimands reduces to preference between the two sets of weights. ■

Example 3.5 (Parametric Models, continued)**.** In some economic applications of parametric models, researchers appear to prefer the estimates from simulated method of moments to maximum likelihood.[13] This view has a long history in economics, with many researchers preferring SMM or related "calibration" procedures to maximum likelihood because of the

[12]For one thing, if θ is only set-identified under the nesting model $\mathcal{M}^*$, then it is by definition impossible to construct an estimator that consistently recovers θ.

[13]For instance, Bordalo et al. (2020) write "We prefer this method to maximum likelihood for two reasons. First, one advantage of our model is that it is simple and transparent. However, this simplicity comes at the cost of likely misspecification, and it is well known that with misspecification concerns moment estimators are often more reliable."

ability to select moments that, in their judgment, more directly reflect the economic quantity of interest—see Dawkins et al. (2001) for an overview of calibration and extensive references.

It is natural to interpret these statements as (implicit) restrictions on the nesting model researchers have in mind, which ensure that some estimands are more closely related to the parameters of interest that others. Such restrictions are rarely stated formally, however. Since these considerations appear to actively guide estimation choices, more explicit articulation of the rationale for a given estimator is useful, and formalizing such arguments seems a potentially fruitful area for future econometric research. ■

Example 3.6 (Generalized Method of Moments, continued)**.** As with OLS, the literature has characterized the GMM estimand for specific choices of moments, weights, and nesting models. For example, Imbens and Angrist (1994) and Angrist and Imbens (1995) show that, under monotonicity, two-stage least squares—a special case of GMM—recovers a positive-weight average of treatment effects. In finance, Hansen and Jagannathan (1997) show that in a misspecified asset pricing model the GMM objective with a specific weighting matrix measures the squared error in approximating an admissible stochastic discount factor, and so measures the "size" of pricing errors in a precise sense. ■

As our discussion of parametric models suggests, even beyond settings covered by existing theoretical guarantees, researchers' choice of weighting matrices for GMM and related methods (e.g. simulated method of moments, minimum distance) are often informed by concerns about misspecification. To systematically explore this, we reviewed empirical papers published in the *American Economic Review* between 2020 and 2024 which a keyword search suggested used moment-based methods (excluding methods like ordinary least squares, two-stage least squares, and maximum likelihood which may be cast as special cases of GMM but which the authors did not describe as such). The search turned up 36 papers. We refer to these papers collectively as our "GMM sample." In this sample, 22 of the papers were over-identified, and in 14 of these no results were reported based on an efficient weighting matrix.[14] For those models where results based on efficient GMM are not reported, the

[14]In addition, some of the papers which notionally use efficient weights plug in parameters estimated in a first step. This changes the form of the "correct" optimal weighting matrix, but is not accounted for in the

most common (non-exclusive) choices were (i) the inverse of the diagonal of the asymptotic variance-covariance matrices (4 papers) (ii) identity weighting matrix (6 papers), and (iii) user-selected weighting matrices, often explicitly chosen based on domain knowledge (7 papers).

One reason the "optimal" weighting matrix may not be used in some cases is that in settings where the sample size is small relative to the number of moments, it is known that estimation of the optimal weighting matrix can lead to bias in both the estimate and the standard errors (see e.g. Altonji and Segal, 1996; Windmeijer, 2005). One suggestion in such cases is to use the inverse of the diagonal alone, which Altonji and Segal (1996) found yielded "a substantial bias reduction" their problem.[15] However, together with optimal weighting this still covers only about half of the over-identified papers in our GMM sample.

For the remaining two choices, identity weights and hand-selected weights, researchers are directly choosing the weight assigned to different moments, and thus the estimand. This is most obvious for hand-selected weighting matrices, where in many cases the authors explicitly motivate their choice of weighting matrix based on their goals for estimation.[16] Perhaps more subtly, for identity weighting matrices the scale that the authors choose for each moment (e.g. whether to measure a monetary outcome in dollars or thousands of dollars) determines the weight.[17] In Lise and Postel-Vinay (2020), for instance, the authors present a standard error formula based on identity weighting (see their Footnote 31), while their code uses a hand-selected diagonal but non-identity weighting (which is consistent with the discussion in their footnote under an appropriate choice of units for the moments).

implemented weighting matrices. Consequently, the number of papers in fact using theoretically efficient weights is still lower.

[15]Consistent with this intepretation, Abebe et al. (2021a) write that "In line with the recent literature, we use this simple weighting matrix instead of the theoretically optimal weighting matrix, which may suffer from small sample bias (Altonji and Segal, 1996)."

[16]For instance, Antràs and De Gortari (2020) write that "In order to guarantee that our model provides a proper quantitative evaluation of the general-equilibrium workings of the world economy, we place a higher weight on matching the empirical moments of larger trade flows." Similarly, Einav et al. (2020) write that "Because identification is primarily driven by the discontinuous change in screening rates at age 40, we weight more heavily moments that are closer to age 40 than moments that are associated with younger and older ages."

[17]Weighting by the inverse of the diagonal of the variance eliminates dependence on the scale of individual moments, but is still sensitive to more general linear transformations of the moments, for instance replacing control and treatment means by control means and treatment-control contrasts.

Remark 3.7 (Further Results on Estimand Characterization)**.** A growing literature characterizes estimands obtained from misspecified models in additional settings. An early positive result by Angrist (1998) shows that in linear models where a binary treatment is as good as randomly assigned given a discrete covariate, regressions on the treatment and saturated dummies for the covariate recover a positive weighted average of causal effects. More recently, however, a large literature warns that estimands from linear (regression or instrumental variables) models fail to recover positively weighted averages of causal effects in important cases. Examples include difference in difference and event-study models estimated via two-way fixed effect regressions (de Chaisemartin and D'Haultfœuille, 2020; Callaway and Sant'Anna, 2021; Sun and Abraham, 2021), IV models estimated by two stage least squares regressions with multiple endogenous variables or insufficiently rich controls (Słoczyński, 2020; Bhuller and Sigstad, 2024; Blandhol et al., 2025), nonlinear estimators for linear IV, such as LIML and continuously updating GMM (Kolesar, 2013; Andrews, 2019), and linear regressions including both controls and multiple treatments (Goldsmith-Pinkham et al., 2024). Chen (2025) develops a general method for characterizing the causal estimand, if any, recovered by linear regression estimates.

Other recent papers characterize estimands in nonlinear models, including Andrews et al. (2025) who study a class of nonlinear structural models and provide necessary and sufficient conditions for instrumental variables estimates to recover linear transformations of causal effects, and Kolesár and Plagborg-Møller (2025) who consider estimation of causal effects in time-series models and derive positive results for certain linear estimators, along with negative results for some popular nonlinear estimators. Other recent work extends the concept of a pseudo-true value to set-identified models (Andrews and Kwon, 2023; Kaido and Molinari, 2024). ■

4. **Inference on Estimands**

After fixing the estimator—and hence a statistical functional $\theta^*(P)$ defined on $\mathcal{P}^*$—we naturally seek to quantify sampling uncertainty, but misspecification can complicate inference.

Under misspecification, standard error formulas derived under $\mathcal{P}$ may no longer be correct; corresponding confidence intervals may undercover the functional $\theta^*(P)$ when $P \in \mathcal{P}^* \setminus \mathcal{P}$. This is a well-known motivation for misspecification-robust standard errors for maximum likelihood (White, 1982).

Similar results exist for GMM (Imbens, 1997; Hall and Inoue, 2003; Hansen and Lee, 2021), yet they are seldom used in practice (analytic misspecification-robust standard errors do not appear to be used in any of the papers in our GMM sample). Since these results seem to be less well-known, we sketch an asymptotic expansion illustrating how misspecification impacts GMM, as well as the intuition for robust standard errors. Let $\hat{\psi}_\Omega$ be the GMM estimator with weighting matrix $\hat{\Omega} = \Omega + o_P(1)$ and ψ^*_Ω be its corresponding statistical functional.[18] Under statistically correct specification, ψ^*_Ω is simply the GMM parameter defined by $\mathbb{E}_P[g(X_i, \psi^*_\Omega)] = 0$ (assumed unique). Under misspecification, ψ^*_Ω minimizes the population GMM criterion $\phi \mapsto (\mathbb{E}_P[g(X_i, \phi)])'\Omega\mathbb{E}_P[g(X_i, \phi)]$. Let $\Gamma = \mathbb{E}_P[\nabla g(X_i, \psi^*_\Omega)]$ and

$$\hat{\Gamma} = \frac{1}{n}\sum_{i=1}^{n} \nabla g(X_i, \hat{\psi}_\Omega) = \Gamma + O_P(1/\sqrt{n}).$$

By the first-order conditions for GMM, $\hat{\Gamma}'\hat{\Omega} g_n(\hat{\psi}_\Omega) = 0 = \Gamma'\Omega\mathbb{E}_P[g_n(\psi^*_\Omega)]$.

To derive the standard errors for the GMM estimator under misspecification, note that by a first-order Taylor expansion and $\sqrt{n}$ consistency of $\hat{\psi}_\Omega$,

$$\begin{aligned}\sqrt{n} g_n(\hat{\psi}_\Omega) &= \sqrt{n} g_n(\psi^*_\Omega) + \underbrace{\left(\frac{1}{n}\sum_{i=1}^{n} \nabla g(X_i, \psi^*_\Omega)\right)}_{\Gamma + o_P(1)} \sqrt{n}(\hat{\psi}_\Omega - \psi^*_\Omega) + o_P(\sqrt{n}\|\hat{\psi}_\Omega - \psi^*_\Omega\|) \\ &= \sqrt{n} g_n(\psi^*_\Omega) + \Gamma\sqrt{n}(\hat{\psi}_\Omega - \psi^*_\Omega) + o_P(1). \end{aligned} \tag{2}$$

Rearranging and multiplying by $\hat{\Gamma}'\hat{\Omega}$,

$$-(\hat{\Gamma}'\hat{\Omega}\Gamma)\sqrt{n}(\hat{\psi}_\Omega - \psi^*_\Omega) = \sqrt{n}\Gamma'\Omega g_n(\psi^*_\Omega) + \sqrt{n}(\hat{\Gamma}'\hat{\Omega} - \Gamma'\Omega) g_n(\psi^*_\Omega) + o_P(1). \tag{3}$$

[18] Since we consider behavior at a fixed P, where possible we suppress the dependence on P to simplify notation.

When the model is well-specified $g_n(\psi^*_\Omega)$ has mean zero, so the second term in (3) is negligible, $\sqrt{n}(\hat{\Gamma}'\hat{\Omega} - \Gamma'\Omega)g_n(\psi^*_\Omega) = o_P(1)$, and the estimator is asymptoticaly normal with the conventional GMM asymptotic variance:

$$\begin{aligned}\sqrt{n}(\hat{\psi}_\Omega - \psi^*_\Omega) &= -(\hat{\Gamma}'\hat{\Omega}\Gamma)^{-1}\sqrt{n}\Gamma'\Omega g_n(\psi^*_\Omega) + o_P(1) \\ &\xrightarrow{d} \mathcal{N}(0, (\Gamma'\Omega\Gamma)^{-1}\Gamma'\Omega\Sigma(P)\Omega\Gamma(\Gamma'\Omega\Gamma)^{-1}),\end{aligned}$$

where $\Sigma(P) \equiv \text{aVar}_P(\sqrt{n}g_n(\psi^*_\Omega))$, and $\text{aVar}_P(\cdot)$ denotes the asymptotic variance under P. However, when the model is statistically misspecified $\sqrt{n}(\hat{\Gamma}'\hat{\Omega} - \Gamma'\Omega)g_n(\psi^*_\Omega)$ is non-negligible and, under regularity conditions, converges to a Gaussian with nonzero, consistently estimable variance. Thus, the familiar sandwich formula is no longer correct and misspecification-robust standard errors should be used instead.

Since many economic models are statistically misspecified, we are skeptical that the conventional GMM formula accurately captures the sampling variability of GMM estimators in practice. Parallel to White (1982)'s widely-adopted suggestion for maximum likelihood, we recommend that robust standard errors be used for GMM as well. Again parallel to maximum likelihood, robust standard errors for GMM do no harm, at least in large samples: They are equal to the conventional standard errors when P happens to be well-specified, but remain valid for inference on $\psi^*_\Omega(P)$ when $P \notin \mathcal{P}$. Consequently, provided $\vartheta(\psi)$ is smooth in ψ, delta-method standard errors will be valid for inference on $\theta^*_\Omega(P) = \vartheta(\psi^*_\Omega(P))$.

The robust standard error formula (e.g. Hall and Inoue, 2003) depends on additional terms not present in the usual formula, and can be complicated in some cases though software implementations are available.[19] An alternative for settings with iid data is the standard nonparametric bootstrap, which would approximate the limit distribution in (3) when $\theta^*(P)$ varies smoothly with P.[20] For bootstrapping GMM models, some propose recentering the bootstrapped moments so as to enforce the moment restriction in the "bootstrap world"

[19]For instance in the dcivreg command in Stata, which implements the misspecification-robust approach of Hwang et al. (2022) for linear IV.

[20]As discussed by Hong and Li (2025), the smoothness requirements for $\sqrt{n}$ asymptotic normality of GMM under misspecification are stronger than those for the correctly specified case. The authors provide a bootstrap which is robust to insufficient smoothness.

(Brown and Newey 2002; also see Section 3.7 in Horowitz 2001). When combined with a t-ratio bootstrap, this approach generates asymptotic refinements under statistically correct specification. Unfortunately, however, recentering yields invalid inference under statistical misspecification. Intuitively, recentering imposes correct specification in the bootstrap data generating process. Thus, the asymptotic variance of the estimator under the recentered bootstrap matches the conventional (invalid) sandwich formula.[21] We thus recommend *against* recentering moments.

Remark 4.1 (A Bayesian Interpretation of Bootstrap Percentile Intervals)**.** While the asymptotic expansion (3) is valid under misspecification it, and the resulting standard errors, still relies on regularity conditions, such as the existence of a unique interior solution $\psi^*_\Omega(P)$ in the population GMM problem, strong identification of $\psi^*_\Omega(P)$, and differentiability of ϑ. Even when these conditions fail, however, the nonparametric bootstrap distribution for the estimate $\hat{\theta}_\Omega$ is asymptotically equivalent to the Bayesian posterior distribution for the estimand $\theta^*_\Omega(P)$ under a particular nonparametric and non-informative limit of priors.[22] Thus, if we use bootstrap percentile intervals for inference on θ^*_Ω, they offer correct large-sample coverage for well-behaved estimands $\theta^*_\Omega(P)$ regardless of model misspecification, while also providing a "fallback" interpretation as a Bayesian credible set for $\theta^*_\Omega(P)$ in less well-behaved cases. The bootstrap also provides a natural tool for diagnosing the failure of the conventional, non-misspecification-robust GMM standard errors: if the bootstrap distribution appears approximately normal and centered at $\hat{\theta}_\Omega$, but differs from the normal suggested by the conventional standard error, this suggests the expansion (3) may be reasonable but that the conventional, non-misspecification-robust asymptotic approximation is not. See Andrews and Shapiro (2024) for formal results on this approach to evaluating asymptotic approximations. ■

[21] Similarly, bootstrapping the moments but using the conventional formula to approximate the estimator variance, as done in some papers in our GMM sample, yields non-robust standard errors.

[22] See e.g. Andrews and Shapiro (2024) for discussion of the relevant priors, and Appendix D of that paper for sufficient conditions for asymptotic equivalence. If we use the Bayes rather than nonparametric bootstrap, then the equivalence is exact.

Remark 4.2 (Further Results on Misspecification-Robust Inference)**.** While White (1980) is primarily framed around heteroskedasticity robustness, these standard errors for linear regression are also robust to nonlinearity of the conditional mean function and so represent a very widely used application of the misspecification-robust approach. More recently, Lee (2018) highlights the invalidity of conventional GMM standard errors in two stage least squares with heterogeneous effects and multiple instruments, and specializes the formula of Hall and Inoue (2003) to this case. Lee (2014) shows that bootstrapping a misspecification-robust t-statistic yields asymptotic refinements for inference on the GMM estimand, offering an alternative to recentering, and Hansen and Lee (2021) derive misspecification-robust standard errors for the iterated-to-convergence GMM estimator. Beyond standard errors, misspecification also invalidates many identification-robust procedures, and Kleibergen and Zhan (2025) derive an identification- and misspecification-robust confidence interval for the continuously updating GMM estimand. Finally, Müller (2013) considers Bayesian decision making in misspecified models and shows that for decision problems concerning ML estimands, decision rules based on an artificial posterior constructed using robust standard errors asymptotically dominate conventional Bayesian inference. ■

5. **Local Misspecification**

Our study of statistical properties (e.g. probability limits and standard errors) has so far focused on fixed, or "global" misspecification, since we have held P fixed and considered behavior as the sample size grows. As n grows, however, any fixed distribution $P \notin \mathcal{P}$ becomes detectable with probability one. In a GMM setting, for instance, J-statistics

$$J = \min_{\psi \in \Psi} n g_n(\psi)' \hat{\Sigma}^{-1} g_n(\psi)$$

diverge to infinity under global misspecification, rather than to their limiting χ^2_{k-p} distribution under statistically correct specification (Hansen, 1982).

In many economic applications the degree of statistical misspecification appears to be roughly on the same order as statistical uncertainty. For instance, J-statistics, while larger

than expected under correct specification, are not always vastly so. Estimates from different weighting matrices may be statistically distinguishable from one another, yet numerically and substantively fairly close. Motivated by this observation, and following Newey (1985), a growing econometrics literature studies models of *local* misspecification, which assume that the sample of size n is generated by a data-generating process P_n such that at some fixed value ψ for the GMM parameters,

$$\mathbb{E}_{P_n}\left[g(X_i, \psi)\right] = \frac{\eta}{\sqrt{n}}.$$

Because standard errors shrink at rate $\sqrt{n}$, the misspecification magnitude matches their scale.

As with other local asymptotic approximations, for instance in the literature studying weak instruments (Staiger and Stock, 1997), we do not imagine that the data distribution would mechanically change were we able to gather more data. Instead, local misspecification is a modeling device to ensure that our asymptotic approximation reflects a salient feature of many finite-sample problems, namely that the apparent degree of misspecification is of the same order as sampling uncertainty. Smaller misspecification would vanish compared to sampling noise; larger misspecification would swamp—and invalidate, as discussed in the last section—standard errors. Hence local misspecification is the appropriate regime for studying trade-offs between estimand choice and estimator precision.

Local misspecification also allows important analytical simplifications. Specifically, in locally misspecified settings many statistics behave like their correctly-specified analogs, save for the addition of extra bias terms. In particular, Newey (1985) Theorem 4 shows that under local misspecification the GMM estimate is asymptotically linear in the moments evaluated at ψ,

$$\hat{\psi}_\Omega - \psi = -(\Gamma'\Omega\Gamma)^{-1}\Gamma'\Omega g_n(\psi) + o_p\left(\frac{1}{\sqrt{n}}\right) \tag{4}$$

where Γ and Σ represent the limits of $\Gamma(P_n)$ and $\Sigma(P_n)$ respectively. Thus, the asymptotic expansion under local misspecification is the same as that in the correctly specified case. In particular, conventional GMM standard errors remain valid under local misspecification.

The local misspecification framework has been used in a number of recent papers. For instance, Andrews et al. (2017) suggest researchers looking to give intuition for the behavior of their estimates may report estimates of $-(\Gamma'\Omega\Gamma)^{-1}\Gamma'\Omega$, since this describes how misspecification in the moments (i.e. different values of η) feeds forward into bias in the estimates. Armstrong and Kolesár (2021) impose the further restriction that $\theta = \vartheta(\psi)$, so the GMM model is econometrically correctly specified in the case where $\eta = 0$, and derive optimal confidence intervals for θ under restrictions on the possible values for η. Bonhomme and Weidner (2022) consider models with additional structure where some model components may be locally misspecified, and derive asymptotically optimal estimators under bounds on the degree of misspecification. Similarly, for a class of moment-condition models imposing parametric assumptions Christensen and Connault (2023) derive bounds on quantities of interest under both local and global misspecification.

Despite this recent activity using the local misspecification framework, there remains scope for new and interesting results in this area. To illustrate, in the remainder of this section we use the local misspecification framework to develop what is to the best of our knowledge a new result, characterizing the relationship between the J-statistic and the class of GMM estimates obtained by varying the weighting matrix.

5.1. **J-Statistics, not J-Tests.** Relatively few economics papers report J-statistics. In our GMM sample, for instance, only 3 out of the 22 over-identified papers report J-statistics, either in the paper or the online appendix.[23] One reason may be that few economists expect their models to hold exactly. Under that view, a non-rejection signals insufficient power of the J-test, not correct specification. This suggests a focus on "economic" rather than statistical measures of model fit. Consistent with this, 19 of the 22 over-identified papers in our GMM sample report some version of an informal identification check, often by plotting the model-implied values of a set of moments against either (i) the data moments that were used to estimate the model or (ii) a "held out" set of moments not used in estimation, which may

[23]One additional paper, Mueller et al. (2021) reports the value of the two-step SMM objective function at its minimum. Because they use the optimal weighting matrix (estimated by bootstrap), this objective value is equal to the J-statistic absent simulation error.

be interpreted as estimation moments assigned zero weight by $\hat{\Omega}$. Both of these approaches convey information about the extent to which the model's over-identifying restrictions appear to hold in the sample, but lack the formal statistical interpretation of the J-statistic.

Such informal checks, while potentially useful, are an imperfect substitute for J-statistics. To highlight this point, in this section we show that J-statistics are useful for much more than J-tests. As discussed in the introduction, the fact that researchers may choose different weighting matrices due to misspecification concern introduces an additional researcher degree of freedom. We show below that (under local misspecification) J-statistics capture, in a precise sense, the range of estimates, t-statistics, and confidence sets a researcher may obtain by varying their choice of weight matrix. This is valuable information for readers, and we encourage wider reporting of J-statistics in applications of GMM.

To better understand the behavior of J statistics in practice, we attempted to calculate J statistics for the papers in our GMM sample which did not already report them. In this exercise we were limited to those papers with accessible data, and where the optimally-weighted GMM objective was sufficiently well-behaved that we were confident in the convergence of our numerical optimization routines. Subject to these restrictions, we were able to compute J-statistics for an additional three papers beyond those which initially reported them. Table 1 collects the J-statistics for these papers, along with the original three. Strikingly, for the 6 papers in our sample where we were able to obtain J statistics, in only one case is the model rejected at the 5% level (although three papers have specifications rejected at the 10% level). Also striking, while the number of papers is limited, there is not obvious selection in which papers report J-statistic. Indeed, the papers for which we computed J-statistics have, on the whole, larger p-values (and thus, in that sense, less evidence of misspecification) than the papers which originally reported J-statistics.

To develop our result, we consider classes of weighting matrices with minimum and maximum eigenvalue bounded below and above by κ^{-1} and κ, respectively, for $\kappa \in [1, \infty)$ a

TABLE 1. XX

Paper	Model	J-statistic	dF	p-value	Reported in the paper?
	Table 4, Column 1	3.05	1	0.08	
Abebe et al. (2021b)	Table 4, Column 2	4.67	3	0.20	No
	Table 4, Column 4	4.28	2	0.12	
Bourreau et al. (2021)	Table 4, Column 2	50.88	7	0.00	J-statistic and dF
Mueller et al. (2021)	Table 5	0.462	3	0.93	No
García-Jimeno et al. (2022)	Table 2, Column 2	2.16	5	0.83	J-statistic and p-value
	Table 6, Column 7	0.361	1	0.55	
Karahan et al. (2024)	Table 7, Column 4	5.813	2	0.05	p-value
	Table 7, Column 5	0.327	1	0.57	
Kelley et al. (2024)	Table 7	0.005	1	0.94	No

Notes: This table reports J-statistics and associated degrees of freedom and p-values for six papers in our "GMM sample." See the text for inclusion criteria.

positive constant:

$$\mathcal{W}_\kappa = \left\{\Omega \in \mathbb{R}^{k\times k} : \Omega = \Omega', \kappa^{-1} \leq \lambda_{\min}(\Omega) \leq \lambda_{\max}(\Omega) \leq \kappa\right\}.$$

Provided the linear approximation (4) holds uniformly over such classes as $n \to \infty$, the J-statistic characterizes, in a precise sense, the extent to which GMM results vary across weighting matrix choices.[24]

Assumption 5.1. *The sample moment conditions evaluated at ψ are asymptotically normal as $n \to \infty$, $\sqrt{n}g_n(\psi) \xrightarrow{d} N(\eta, \Sigma)$. Equation (4) holds uniformly, in the sense that*

$$\sup_{\Omega\in\mathcal{W}_\kappa} \sqrt{n}\left|\hat{\psi}_\Omega - \psi + (\Gamma'\Omega\Gamma)^{-1}\Gamma'\Omega g_n(\psi)\right| \xrightarrow{p} 0$$

as $n \to \infty$ for all fixed $\kappa > 0$. Moreover, we have estimators $\hat{\Gamma}_\Omega$ and $\hat{\Sigma}_\Omega$ such that $\hat{\Gamma}_\Omega \xrightarrow{p} \Gamma$ and $\hat{\Sigma}_\Omega \xrightarrow{p} \Sigma$ uniformly over $\mathcal{W}_\kappa$, where both limits are full rank and do not depend on Ω, and $\sqrt{n}\|g_n(\hat{\psi}_{\Sigma^{-1}}) - g_n(\psi) + \Gamma(\hat{\psi}_{\Sigma^{-1}} - \psi)\| \xrightarrow{p} 0$.

[24]We limit attention to $\Omega \in \mathcal{W}_\kappa$ because there is no hope that the linear approximation is uniformly valid over classes that whose closure includes e.g. weight matrices with severely reduced rank.

Let

$$\hat{\sigma}^2_\Omega = \frac{1}{n}\hat{h}' \left(\hat{\Gamma}'_\Omega \Omega \hat{\Gamma}_\Omega\right)^{-1} \hat{\Gamma}'_\Omega \Omega \hat{\Sigma}_\Omega \Omega \hat{\Gamma}_\Omega \left(\hat{\Gamma}'_\Omega \Omega \hat{\Gamma}_\Omega\right)^{-1} \hat{h} \quad \hat{h} = \frac{\partial \vartheta(\hat{\psi}_\Omega)}{\partial \psi}$$

denote the delta-method variance estimator for $\hat{\theta}_\Omega$, and $\hat{\Theta}(\kappa, \tau)$ the set of GMM estimates that are attainable with (i) $\Omega \in \mathcal{W}_\kappa$ and (ii) standard error less than $\sqrt{1+\tau^2}$ times that of the efficient GMM estimator,

$$\hat{\Theta}(\kappa, \tau) = \left\{\hat{\theta}_\Omega : \Omega \in \mathcal{W}_\kappa, \hat{\sigma}^2_\Omega \leq (1+\tau^2)\hat{\sigma}^2_{\Sigma^{-1}}\right\}.$$

Finally, let $\hat{\Theta}(\tau)$ be the efficient GMM confidence interval with critical value $\tau\sqrt{J}$

$$\hat{\Theta}(\tau) = \left[\hat{\theta}_{\Sigma^{-1}} \pm \tau\sqrt{J}\hat{\sigma}_{\Sigma^{-1}}\right].$$

One can show that these sets are asymptotically equivalent for large κ.

Proposition 5.2. *Under Assumption 5.1, and for*

$$d_H(A, B) = \max\left\{\sup_{a\in A}\inf_{b\in B}|a-b|, \sup_{b\in B}\inf_{a\in A}|a-b|\right\}$$

the Hausdorff distance between the sets A and B, for any $\varepsilon > 0$,

$$\lim_{\kappa\to\infty}\limsup_{n\to\infty} \mathrm{P}\left(\sqrt{n}\cdot d_H\left(\hat{\Theta}(\kappa, \tau), \hat{\Theta}(\tau)\right) > \varepsilon\right) = 0.$$

Proposition 5.2 states that, asymptotically, GMM estimates with, e.g., approximately 10% higher asymptotic variance than the efficient estimator form an interval around $\hat{\theta}_{\Sigma^{-1}}$ with radius $\sqrt{10\%}\cdot\sqrt{J}\cdot\hat{\sigma}_{\Sigma^{-1}}$. Thus, if we consider the set of estimates that a researcher can obtain at a given standard error "cost" relative to the efficient GMM estimate, this set is (asymptotically) characterized by the J-statistic. Note that while we state this result using the "oracle" efficient estimator $\hat{\theta}_{\Sigma^{-1}}$ and standard error $\hat{\sigma}_{\Sigma^{-1}}$ for simplicity, under Assumption 5.1 it is equivalent to use the feasible variance estimate $\hat{\Sigma}_\Omega$ for any fixed, full-rank Ω.

This result has immediate implications for the behavior of t-statistics and confidence intervals.

Corollary 5.3. *Under the assumptions of Proposition 5.2, consider the t-statistic for testing $H_0 : \theta = \theta_0$ based on the GMM estimator with weighting matrix Ω:*

$$t_\Omega(\theta_0) = \frac{\hat{\theta}_\Omega - \theta_0}{\hat{\sigma}_\Omega}.$$

Then for all $\varepsilon > 0$,

$$\lim_{\kappa\to\infty} \limsup_{n\to\infty} \mathrm{P}\left(\left|\inf_{\theta_0\in\mathbb{R}} \sup_{\Omega\in\mathcal{W}_\kappa} |t_\Omega(\theta_0)| - \sqrt{J}\right| > \varepsilon\right) = 0.$$

This result shows that, asymptotically, a researcher can find a weighting matrix that yields an absolute t-statistic of at least $\sqrt{J}$ for any possible null hypothesis. Hence, for instance, when $\sqrt{J} > 1.96$ a nefarious researcher who is unconstrained in their choice of weighting matrix can reject any null hypothesis at the 5% level. Consequently, in order to limit the scope for "weight-hacking" to yield statistically significant results, it must be that either J statistics are small or that other restrictions constrain the choice of weight matrix.[25]

Absent restrictions on the weighting matrix the J-statistic threshold necessary for a result to be "weight-hacking" proof at the 5% level, $\sqrt{J} \leq 1.96$ does not scale with the degree of over-identification, and so can be much more stringent than conventional J-test critical values. In Table 1, for instance, four of ten specifications, and 3 of 6 papers, have $\sqrt{J} > 1.96$, despite only one specification being rejected at the 5% level by the J-test.

Restated in terms of confidence intervals, we obtain the following result.

Corollary 5.4. *Under the assumptions of Proposition 5.2, for a critical value c, consider confidence sets centered around the GMM estimator $\hat{\theta}_\Omega$, $\hat{C}_\Omega(c) = [\hat{\theta}_\Omega \pm c\hat{\sigma}_\Omega]$. Then for all $\varepsilon > 0$,*

$$\lim_{\kappa\to\infty} \limsup_{n\to\infty} \mathrm{P}\left(\left|\inf\left\{c \geq 0 : \bigcap_{\Omega\in\mathcal{W}_\kappa} C_\Omega(c) \neq \varnothing\right\} - \sqrt{J}\right| > \varepsilon\right) = 0.$$

The intersection point is asymptotically equal to the efficient GMM estimator:

$$\lim_{\nu\to 0} \lim_{\kappa\to\infty} \limsup_{n\to\infty} \mathrm{P}\left(\left|\sqrt{n} \cdot d_H\left(\bigcap_{\Omega\in\mathcal{W}_\kappa} C_\Omega(\sqrt{J}+\nu), \left\{\hat{\theta}_{\Sigma^{-1}}\right\}\right)\right| > \varepsilon\right) = 0.$$

[25] Among other results, Honoré et al. (2020) characterize the derivative of σ^2_Ω with respect to elements of Ω. Their results may thus be useful for building intuition about which Ω satisfy $\hat{\sigma}^2_\Omega \leq (1+\tau^2)\hat{\sigma}^2_{\Sigma^{-1}}$.

Hence, to ensure at least one point of overlap between the confidence sets of researchers considering different weighting matrices, researchers would need to use critical values no smaller than $\sqrt{J}$. This result bears an interesting resemblance to one by Andrews et al. (2024). There, working in a Bayesian framework and studying linear-in-parameters moment condition models, the authors show that confidence intervals with width $\sqrt{J}$ have correct average coverage under a class of priors motivated by the GMM criterion.

Our asymptotic results in this section follow from exact, finite-sample results in a model, discussed in the Appendix, where the moments are linear in parameters and have known variance. Armstrong and Kolesár (2021) show that this linear model corresponds to the limit experiment for weakly identified GMM.

Returning to Table 1, of the 10 specifications, 4 have J-statistics larger than $1.96^2 = 3.84$; for these specifications, it is possible to choose weight matrices to reject *any* point null hypothesis. More quantitatively, using Proposition 5.2, we can interpret conventional confidence sets around the efficient estimator $\hat{\theta}_{\Sigma^{-1}}$ as a set of GMM *estimates* at a given standard error inflation $\sqrt{1+\tau^2}-1$ for $\tau = 1.96/\sqrt{J}$. The interquartile range of this implied standard error inflation is 35% to 230%.

Remark 5.5 (Further Results on Local Misspecification)**.** Other recent work includes Kitamura et al. (2013), who establish a local-misspecification-robustness property for a generalized empirical likelihood estimator, and Andrews et al. (2020) who use a local misspecification framework to examine the relationship between structural estimates and other, non-moment statistics, again with the goal of helping researchers clarify what "drives" their estimates. Fessler and Kasy (2019) show how one can use the implications of locally misspecified economic models to improve unbiased but potentially noisy estimates, while Armstrong et al. (2024) consider the problem of optimally combining biased and unbiased estimators under local misspecification. ■

References

Abebe, Girum, A Stefano Caria, and Esteban Ortiz-Ospina, "The selection of talent:

Experimental and structural evidence from ethiopia," *American Economic Review*, 2021, *111* (6), 1757–1806. 16

_____, **A. Stefano Caria, and Esteban Ortiz-Ospina**, "The Selection of Talent: Experimental and Structural Evidence from Ethiopia," *American Economic Review*, 2021, *111* (6), 1757–1806. 25

Altonji, Joseph G and Lewis M Segal, "Small-sample bias in GMM estimation of covariance structures," *Journal of Business & Economic Statistics*, 1996, *14* (3), 353–366. 16

Andrews, Donald W K and Soonwoo Kwon, "Misspecified Moment Inequality Models: Inference and Diagnostics," *The Review of Economic Studies*, 2023, *91* (1), 45–76. 17

Andrews, Isaiah, "On the structure of IV estimands," *Journal of Econometrics*, 2019, *211* (1), 294–307. 17

_____ **and Jesse M Shapiro**, "Communicating Scientific Uncertainty via Approximate Posteriors," Working Paper 32038, National Bureau of Economic Research January 2024. 20

_____, **Harvey Barnhard, and Jake Carlson**, "True and Pseudo-True Values," 2024. 28

_____, **Matthew Gentzkow, and Jesse M Shapiro**, "Measuring the sensitivity of parameter estimates to estimation moments," *The Quarterly Journal of Economics*, 2017, *132* (4), 1553–1592. 23

_____, _____, **and Jesse M. Shapiro**, "On the Informativeness of Descriptive Statistics for Structural Estimates," *Econometrica*, 2020, *88* (6), 2231–2258. 28

_____, **Nano Barahona, Matthew Gentzkow, Ashesh Rambachan, and Jesse M Shapiro**, "Structural Estimation Under Misspecification: Theory and Implications for Practice," *The Quarterly Journal of Economics*, 2025. 17

Angrist, Joshua D., "Estimating the Labor Market Impact of Voluntary Military Service Using Social Security Data on Military Applicants," *Econometrica*, 1998, *66* (2), 249–288. 17

Angrist, Joshua D and Guido W Imbens, "Two-stage least squares estimation of average causal effects in models with variable treatment intensity," *Journal of the American statistical Association*, 1995, *90* (430), 431–442. 15

Antràs, Pol and Alonso De Gortari, "On the geography of global value chains," *Econometrica*, 2020, *88* (4), 1553–1598. 16

Armstrong, Timothy B and Michal Kolesár, "Sensitivity analysis using approximate moment condition models," *Quantitative Economics*, 2021, *12* (1), 77–108. 23, 28, 36

Armstrong, Timothy B., Patrick Kline, and Liyang Sun, "Adapting to Misspecification," 2024. 28

Beer, Stafford, *Diagnosing the system for organizations*, John Wiley and Sons, 1985. 3

Beran, Rudolf and Peter Hall, "Estimating coefficient distributions in random coefficient regressions," *The annals of Statistics*, 1992, pp. 1970–1984. 6

Bhuller, Manudeep and Henrik Sigstad, "2SLS with multiple treatments," *Journal of Econometrics*, 2024, *242* (1), 105785. 17

Blandhol, Christine, John Bonney, Magne Mogstad, and Alexander Torgovitsky, "When is TSLS Actually LATE?," Working Paper 29709, National Bureau of Economic Research 2025. 17

Bonhomme, Stéphane and Martin Weidner, "Minimizing sensitivity to model misspecification," *Quantitative Economics*, 2022, *13* (3), 907–954. 23

Bordalo, Pedro, Nicola Gennaioli, Yueran Ma, and Andrei Shleifer, "Overreaction in macroeconomic expectations," *American Economic Review*, 2020, *110* (9), 2748–2782. 14

Bourreau, Marc, Yutec Sun, and Frank Verboven, "Market Entry, Fighting Brands, and Tacit Collusion: Evidence from the French Mobile Telecommunications Market," *American Economic Review*, 2021, *111* (11), 3459–99. 25

Breunig, Christoph and Stefan Hoderlein, "Specification testing in random coefficient models," *Quantitative Economics*, 2018, *9* (3), 1371–1417. 9

Brown, Bryan W and Whitney K Newey, "Generalized method of moments, efficient

bootstrapping, and improved inference," *Journal of Business & Economic Statistics*, 2002, *20* (4), 507–517. 20

Callaway, Brantly and Pedro H.C. Sant'Anna, "Difference-in-Differences with multiple time periods," *Journal of Econometrics*, 2021, *225* (2), 200–230. 17

Chen, Jiafeng, "Potential weights and implicit causal designs in linear regression," 2025. 17

Chen, Xiaohong and Andres Santos, "Overidentification in regular models," *Econometrica*, 2018, *86* (5), 1771–1817. 2, 10, 12

Christensen, Timothy and Benjamin Connault, "Counterfactual Sensitivity and Robustness," *Econometrica*, 2023, *91* (1), 263–298. 23

Davidson, Russell and James Mackinnon, *Econometric Theory and Methods*, Oxford University Press, 2021. 2

Dawkins, Christina, Thirukodikaval Nilakanta Srinivasan, and John Whalley, "Calibration," in "Handbook of econometrics," Vol. 5, Elsevier, 2001, pp. 3653–3703. 15

de Chaisemartin, Clément and Xavier D'Haultfœuille, "Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects," *American Economic Review*, 2020, *110* (9), 2964–96. 17

Einav, Liran, Amy Finkelstein, Tamar Oostrom, Abigail Ostriker, and Heidi Williams, "Screening and selection: The case of mammograms," *American Economic Review*, 2020, *110* (12), 3836–3870. 16

Fessler, Pirmin and Maximilian Kasy, "How to Use Economic Theory to Improve Estimators: Shrinking Toward Theoretical Restrictions," *The Review of Economics and Statistics*, 10 2019, *101* (4), 681–698. 28

García-Jimeno, Camilo, Angel Iglesias, and Pinar Yildirim, "Information Networks and Collective Action: Evidence from the Women's Temperance Crusade," *American Economic Review*, 2022, *112* (1), 41–80. 25

Goldsmith-Pinkham, Paul, Peter Hull, and Michal Kolesár, "Contamination Bias in Linear Regressions," *American Economic Review*, 2024, *114* (12), 4015–51. 17

Hall, Alastair R and Atsushi Inoue, "The large sample behaviour of the generalized method of moments estimator in misspecified models," *Journal of Econometrics*, 2003, *114* (2), 361–394. 13, 18, 19, 21

Hansen, Bruce E and Seojeong Lee, "Inference for iterated GMM under misspecification," *Econometrica*, 2021, *89* (3), 1419–1447. 18, 21

Hansen, Lars Peter, "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 1982, *50* (4), 1029–1054. 2, 4, 21

_____ **and Ravi Jagannathan**, "Assessing specification errors in stochastic discount factor models," *The Journal of Finance*, 1997, *52* (2), 557–590. 15

_____ **and Thomas Sargent**, *Robustness*, Princeton University Press, 2008. 2

Hayashi, Fumio, *Econometrics*, Princeton University Press, 2000. 2

Heckman, James and Rodrigo Pinto, "Causal analysis after Haavelmo," *Econometric Theory*, 2015, *31* (1), 115–151. 11

Hong, Han and Jessie Li, "Rate-adaptive Bootstrap for Possibly Misspecified GMM," *Econometric Theory*, 2025, *41* (2), 421–471. 19

Honoré, Bo, Thomas Jørgensen, and Áureo de Paula, "The informativeness of estimation moments," *Journal of Applied Econometrics*, 2020, *35* (7), 797–813. 27

Horowitz, Joel L, "The bootstrap," in "Handbook of econometrics," Vol. 5, Elsevier, 2001, pp. 3159–3228. 20

Hwang, Jungbin, Byunghoon Kang, and Seojeong Lee, "A doubly corrected robust variance estimator for linear GMM," *Journal of Econometrics*, 2022, *229* (2), 276–298. 19

Imbens, Guido W, "One-step estimators for over-identified generalized method of moments models," *The Review of Economic Studies*, 1997, *64* (3), 359–383. 13, 18

Imbens, Guido W. and Joshua D. Angrist, "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 1994, *62* (2), 467–475. 15

Intriligator, Michael D, "Economic and econometric models," *Handbook of econometrics*, 1983, *1*, 181–221. 11

Kaido, Hiroaki and Francesca Molinari, "Information Based Inference in Models with

Set-Valued Predictions and Misspecification," 2024. 17

Karahan, Fatih, Benjamin Pugsley, and Ayşegül Şahin, "Demographic Origins of the Start-up Deficit," *American Economic Review*, 2024, *114* (7), 1986–2023. 25

Kelley, Erin M., Gregory Lane, and David Schönholzer, "Monitoring in Small Firms: Experimental Evidence from Kenyan Public Transit," *American Economic Review*, 2024, *114* (10), 3119–60. 25

Kitamura, Yuichi, Taisuke Otsu, and Kirill Evdokimov, "Robustness, Infinitesimal Neighborhoods, and Moment Restrictions," *Econometrica*, 2013, *81* (3), 1185–1201. 28

Kleibergen, Frank and Zhaoguo Zhan, "Double robust inference for continuous updating GMM," *Quantitative Economics*, 2025, *16* (1), 295–327. 21

Kolesar, Michal, "Estimation in an Instrumental Variables Model with Treatment Effect Heterogenity," 2013. 17

Kolesár, Michal and Mikkel Plagborg-Møller, "Dynamic Causal Effects in a Nonlinear World: the Good, the Bad, and the Ugly," 2025. 17

Koopmans, Tjalling C and Olav Reiersol, "The identification of structural characteristics," *The Annals of Mathematical Statistics*, 1950, *21* (2), 165–181. 10, 11

Lee, Seojeong, "Asymptotic refinements of a misspecification-robust bootstrap for generalized method of moments estimators," *Journal of Econometrics*, 2014, *178*, 398–413. 21

_____, "A Consistent Variance Estimator for 2SLS When Instruments Identify Different LATEs," *Journal of Business & Economic Statistics*, 2018, *36* (3), 400–410. 21

Lise, Jeremy and Fabien Postel-Vinay, "Multidimensional Skills, Sorting, and Human Capital Accumulation," *American Economic Review*, 2020, *110* (8), 2328–76. 16

Manski, Charles F., *Public Policy in an Uncertain World*, Harvard University Press, 2013. 2

Mogstad, Magne and Alexander Torgovitsky, "Chapter 1 - Instrumental variables with unobserved heterogeneity in treatment effects," in Christian Dustmann and Thomas Lemieux, eds., *Christian Dustmann and Thomas Lemieux, eds.*, Vol. 5 of *Handbook of*

Labor Economics, Elsevier, 2024, pp. 1–114. 3

Mueller, Andreas I., Johannes Spinnewijn, and Giorgio Topa, "Job Seekers' Perceptions and Employment Prospects: Heterogeneity, Duration Dependence, and Bias," *American Economic Review*, 2021, *111* (1), 324–63. 23, 25

Müller, Ulrich K., "Risk of Bayesian Inference in Misspecified Models, and the Sandwich Covariance Matrix," *Econometrica*, 2013, *81* (5), 1805–1849. 21

Newey, Whitney K., "Generalized method of moments specification testing," *Journal of Econometrics*, 1985, *29* (3), 229–256. 22

Neyman, Jerzy, "On the application of probability theory to agricultural experiments. Essay on principles. Section 9," *Statistical Science*, 1923/1990, pp. 465–472. 11

Pearl, Judea, *Causality*, Cambridge university press, 2009. 11

Richardson, Thomas S and James M Robins, "Single world intervention graphs (SWIGs): A unification of the counterfactual and graphical approaches to causality," *Center for the Statistics and the Social Sciences, University of Washington Series. Working Paper*, 2013, *128* (30), 2013. 11

Rodrik, Dani, *Economics Rules: The Rights and Wrongs of The Dismal Science*, W.W. Norton, 2015. 2

Rubin, Donald B, "Estimating causal effects of treatments in randomized and nonrandomized studies.," *Journal of educational Psychology*, 1974, *66* (5), 688. 11

Simmons, Joseph P., Leif D. Nelson, and Uri Simonsohn, "False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant," *Psychological Science*, 2011, *22* (11), 1359–1366. 3

Słoczyński, Tymon, "When should we (not) interpret linear iv estimands as late?," *arXiv preprint arXiv:2011.06695*, 2020. 17

Spanos, Aris, "Transforming structural econometrics: substantive vs. statistical premises of inference," *Review of Political Economy*, 2016, *28* (3), 426–437. 5

Staiger, Douglas and James H. Stock, "Instrumental Variables Regression with Weak Instruments," *Econometrica*, 1997, *65* (3), 557–586. 22

Stock, James and Mark Watson, *Introduction to Econometrics*, Pearson, 2018. 2

Sun, Liyang and Sarah Abraham, "Estimating dynamic treatment effects in event studies with heterogeneous treatment effects," *Journal of Econometrics*, 2021, *225* (2), 175–199. 17

van der Vaart, Aad W, *Asymptotic statistics*, Vol. 3, Cambridge university press, 2000. 10

White, Halbert, "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," *Econometrica*, 1980, *48* (4), 817–838. 21

_____, "Maximum likelihood estimation of misspecified models," *Econometrica: Journal of the econometric society*, 1982, pp. 1–25. 2, 14, 18, 19

Windmeijer, Frank, "A finite sample correction for the variance of linear efficient two-step GMM estimators," *Journal of econometrics*, 2005, *126* (1), 25–51. 16

Woolridge, Jeffrey M, "Introductory econometrics: A modern approach," *Mason, OH: South Western Cengage Learning*, 2009. 11

Yitzhaki, Shlomo, "On using linear regressions in welfare economics," *Journal of Business & Economic Statistics*, 1996, *14* (4), 478–486. 14

Appendix A. Normal Limit Problem

Armstrong and Kolesár (2021) show that the problem of estimating θ under local misspecification reduces, asymptotically, to that of observing (Y, Γ, Σ, h) where we wish to estimate $\theta = h'\phi$ and

$$Y = -\Gamma\phi + \eta + \Sigma^{1/2}\epsilon \quad \epsilon \sim \mathcal{N}(0, I_k). \tag{5}$$

Quantities associated with GMM have natural analogues in the limit problem (5), which we collect in Table 2. These quantities are linked in the sense that the asymptotic behavior of GMM quantities (left-side of Table 2) agree with the behavior of the limit problem analogues under (5). In particular, GMM estimators for ψ with weight matrix Ω correspond to estimators of ϕ in the form

$$\hat{\phi}_\Omega = -(\Gamma'\Omega\Gamma)^{-1}\Gamma'\Omega Y.$$

The finite-sample distribution of $\hat{\phi}_\Omega$ in the model (5)

$$\hat{\phi}_\Omega \sim \mathcal{N}(\phi - (\Gamma'\Omega\Gamma)^{-1}\Gamma'\Omega\eta, (\Gamma'\Omega\Gamma)^{-1}\Gamma'\Omega\Sigma\Omega\Gamma(\Gamma'\Omega\Gamma)^{-1})$$

is exactly the GMM limit distribution under local misspecification. The corresponding estimators for θ is then $\hat{\theta}_\Omega = h'\hat{\phi}_\Omega$. Similarly, the J-statistic has counterpart

$$\begin{aligned} J &= \min_u (Y + \Gamma u)\Sigma^{-1}(Y + \Gamma u) \\ &= (Y + \Gamma\hat{\phi}_{\Sigma^{-1}})\Sigma^{-1}(Y + \Gamma\hat{\phi}_{\Sigma^{-1}}) \end{aligned}$$

These statistics have even simpler representations in a canonical reparametrization of (5). See Proposition A.5.

Remark A.1 (The problem (5) as a limit version of GMM)**.** To further motivate the limit problem (5), we briefly sketch the connection to the asymptotics of GMM. We first localize the problem to a $1/\sqrt{n}$-neighborhood of ψ. Suppose we are given some (nonrandom) $\tilde{\psi}$ with

$$-\phi \equiv \sqrt{n}(\tilde{\psi} - \psi) = O(1).$$

TABLE 2. Quantities in GMM estimation and their analogues in the limit problem (5)

GMM problem	Analogue in the limit problem (5)
ψ	ϕ
$\theta = \vartheta(\psi)$	$\theta = h'\phi$
Jacobian of moments, $\mathbb{E}_P\left[\frac{\partial}{\partial\psi} g(X_i, \psi)\right]$	Γ
Variance of moments, $\mathrm{Var}_P(g(X_i, \psi))$	Σ
Gradient of θ w.r.t. ψ, $\frac{\partial\vartheta}{\partial\psi}$	h
GMM estimator of ψ with weight matrix Ω_P	$-(\Gamma'\Omega\Gamma)^{-1}\Gamma'\Omega Y$
GMM estimator of θ with weight matrix Ω_P	$-h'(\Gamma'\Omega\Gamma)^{-1}\Gamma'\Omega Y$
J-statistic $\min_{\phi\in\Psi} g_n(\phi)' \mathrm{Var}_P(g(X_i, \psi))^{-1} g_n(\phi)$	$\min_u (Y + \Gamma u)\Sigma^{-1}(Y + \Gamma u)$

We can view ϕ as the parameter of interest instead. Any estimator $\hat{\psi}$ of ψ induces an estimator $\hat{\phi}$ of ϕ since

$$\sqrt{n}(\hat{\psi} - \psi) = \underbrace{\sqrt{n}(\hat{\psi} - \tilde{\psi})}_{\equiv\hat{\phi}} + \underbrace{\sqrt{n}(\tilde{\psi} - \psi)}_{-\phi} = \hat{\phi} - \phi.$$

For $h \in \mathbb{R}^p$ the gradient of $\vartheta(\cdot)$, we also have by linearization

$$\sqrt{n}(\vartheta(\hat{\psi}) - \vartheta(\psi)) = h'(\hat{\phi} - \phi) + o_P(1) \tag{6}$$

for $\hat{\psi}$ that are $\sqrt{n}$-consistent.

Given $\tilde{\psi}$, we can summarize the statistical information in the GMM problem with the value of the moments at $\tilde{\psi}$, $g_n(\tilde{\psi})$. By a Taylor expansion analogous to (2),

$$\sqrt{n} g_n(\tilde{\psi}) = \sqrt{n} g_n(\psi) + \Gamma\sqrt{n}(\tilde{\psi} - \psi) + o_P(1) = \sqrt{n} g_n(\psi) - \Gamma\phi + o_P(1).$$

Finally, by the central limit theorem, $\sqrt{n} g_n(\psi) \xrightarrow{d} \mathcal{N}(\eta, \Sigma)$ for Σ the limit of $\mathrm{Var}_P[g(\cdot, \psi)]$. Thus, the observed quantity $Y \equiv \sqrt{n} g_n(\tilde{\psi})$ has asymptotic distribution

$$Y \sim \mathcal{N}(-\Gamma\phi + \eta, \Sigma).$$

The above display, combined with (6), is equivalent to the problem (5).[26] ■

[26]Without local misspecification, terms like $(\hat{\Gamma}'\hat{\Omega} - \Gamma_P'\Omega_P)\eta$ are no longer asymptotically negligible, since η grows with n.

In the limit problem (5), we obtain exact, non-asymptotic analogs of Proposition 5.2, Corollary 5.3, and Corollary 5.4.

Proposition A.2. *In the limit problem (5), let $\hat{\theta}_\Omega = h'\hat{\phi}_\Omega$ and assume Γ, Σ, h all have full rank. For $\tau \geq 0$ and $\sigma_\Omega^2 = \mathrm{Var}(\hat{\theta}_\Omega)$, the set of GMM estimates for θ with variance at most $(1+\tau^2)\sigma_{\Sigma^{-1}}^2$ is (almost surely) an interval centered at $\hat{\theta}_{\Sigma^{-1}}$ with radius $\sigma_{\Sigma^{-1}}\tau\sqrt{J}$:*

$$\left\{h'\hat{\phi}_\Omega : \sigma_\Omega^2 \leq (1+\tau^2)\sigma_{\Sigma^{-1}}^2\right\} = \left[\hat{\theta}_{\Sigma^{-1}} \pm \sigma_{\Sigma^{-1}}\tau\sqrt{J}\right].$$

Corollary A.3. *In the setup of Proposition A.2, consider the t-statistic for $H_0 : \theta = \theta_0$ of the GMM estimator with weighting matrix Ω:*

$$t_\Omega(\theta_0) = \frac{h'\hat{\phi}_\Omega - \theta_0}{\sqrt{\mathrm{Var}(h'\phi_\Omega)}}.$$

Then

$$\inf_{\theta_0}\sup_{\Omega}|t_\Omega(\theta_0)| = \sqrt{J}.$$

Corollary A.4. *In the setup of Proposition A.2, for a critical value c, consider confidence sets centered around the GMM estimator $\hat{\theta}_\Omega$, $\hat{C}_\Omega(c) = [\hat{\theta}_\Omega \pm c\sigma_\Omega]$. Then $\sqrt{J}$ is the smallest critical value such that all confidence sets intersect:*

$$\sqrt{J} = \inf\left\{c \geq 0 : \bigcap_\Omega C_\Omega(c) \neq \varnothing\right\}. \tag{7}$$

The intersection point is the efficient GMM estimator: $\bigcap_\Omega C_\Omega(\sqrt{J}) = \{\hat{\theta}_{\Sigma^{-1}}\}$.

To prove these results, it is helpful to work with a canonical, normalized form of the limit experiment:

$$Y^* = \begin{bmatrix}\hat{\phi}_{\Sigma^{-1}} \\ Z\end{bmatrix} = \begin{bmatrix}\phi \\ 0\end{bmatrix} + \eta^* + \Sigma^{*1/2}\epsilon^* \quad \epsilon^* \sim \mathcal{N}(0, I_k). \tag{8}$$

for some block-diagonal $\Sigma^* = \mathrm{diag}(\Sigma_\phi^*, \Sigma_Z^*)$.

Proposition A.5. *Suppose Γ is full-rank and Σ is positive definite. Then (5) is equivalent to (8) in the sense that there is some known and invertible matrix Q whereby*

$$Y^* = QY \quad \eta^* = Q\eta \quad \Sigma^* = Q\Sigma Q'.$$

Moreover, recall that $\hat{\theta}_\Omega = -h'(\Gamma_P'\Omega_P\Gamma)^{-1}\Gamma_P'\Omega_P Y$, $\sigma_\Omega^2 = \mathrm{Var}(\hat{\theta}_\Omega)$, $\hat{\theta}_{\Sigma^{-1}} = \hat{\theta}_{\Sigma^{-1}}$, $\sigma_{\Sigma^{-1}}^2 = \sigma_{\Sigma^{-1}}^2$. Then,

(1) GMM estimators can be written as $\hat{\theta}_\Omega = h'\hat{\phi}_\Omega = h'\hat{\phi}_{\Sigma^{-1}} + v_\Omega' Z$, where any v_Ω corresponds to some choice of the weighting matrix

$$\{v_\Omega : \Omega \text{ is positive definite}\} = \mathbb{R}^{k-p}.$$

(2) $\sigma_\Omega^2 = \sigma_{\Sigma^{-1}}^2 + v_\Omega'\tilde{\Sigma}_Z^* v_\Omega$

(3) The J-statistic is the norm of Z: $J = Z'\left(\Sigma_Z^*\right)^{-1} Z$.

A.1. **Proofs for Normal Limit Problem.**

Proposition A.2. *In the limit problem (5), let $\hat{\theta}_\Omega = h'\hat{\phi}_\Omega$ and assume Γ, Σ, h all have full rank. For $\tau \geq 0$ and $\sigma_\Omega^2 = \mathrm{Var}(\hat{\theta}_\Omega)$, the set of GMM estimates for θ with variance at most $(1+\tau^2)\sigma_{\Sigma^{-1}}^2$ is (almost surely) an interval centered at $\hat{\theta}_{\Sigma^{-1}}$ with radius $\sigma_{\Sigma^{-1}}\tau\sqrt{J}$:*

$$\left\{h'\hat{\phi}_\Omega : \sigma_\Omega^2 \leq (1+\tau^2)\sigma_{\Sigma^{-1}}^2\right\} = \left[\hat{\theta}_{\Sigma^{-1}} \pm \sigma_{\Sigma^{-1}}\tau\sqrt{J}\right].$$

Proof. Consider the reparametrization of (5) in (8). Proposition A.5 verifies that

(1) GMM estimators can be written as $\hat{\theta}_\Omega = h'\hat{\phi}_\Omega = h'\hat{\phi}_{\Sigma^{-1}} + v_\Omega' Z$, where any v_Ω corresponds to some choice of the weighting matrix

$$\{v_\Omega : \Omega \text{ is positive definite}\} = \mathbb{R}^{k-p}.$$

(2) $\sigma_\Omega^2 = \sigma_{\Sigma^{-1}}^2 + v_\Omega'\Sigma_Z^* v_\Omega$

(3) The J-statistic is the norm of Z: $J = Z'\Sigma_Z^{*\,-1} Z$.

As a result, for $v_\Omega = c{\Sigma_Z^*}^{-1}Z$, we have that $\hat{\theta}_\Omega = \hat{\theta}_{\Sigma^{-1}} + cJ$ and

$$\sigma_\Omega^2 = \sigma_{\Sigma^{-1}}^2 + c^2 J.$$

As we vary $c^2 \leq \frac{\tau^2\sigma_{\Sigma^{-1}}^2}{J}$, we obtain the range of estimates $\left[\hat{\theta}_{\Sigma^{-1}} \pm \sigma_{\Sigma^{-1}}\tau\sqrt{J}\right]$. This proves

$$\left\{h'\hat{\phi}_\Omega : \sigma_\Omega^2 \leq (1+\tau^2)\sigma_{\Sigma^{-1}}^2\right\} \subset \left[\hat{\theta}_{\Sigma^{-1}} \pm \sigma_{\Sigma^{-1}}\tau\sqrt{J}\right].$$

On the other hand, note that

$$\sup_{v_\Omega'\Sigma_Z^* v_\Omega \leq \tau^2\sigma_{\Sigma^{-1}}^2} v_\Omega' Z = \tau\sigma_{\Sigma^{-1}}\|{\Sigma_Z^*}^{-1/2}Z\| = \tau\sigma_{\Sigma^{-1}}\sqrt{J}$$

Thus, anything outside of $\left[\hat{\theta}_{\Sigma^{-1}} \pm \sigma_{\Sigma^{-1}}\tau\sqrt{J}\right]$ is not attainable by any choice of v_Ω. This completes the proof. □

Corollary A.4. *In the setup of Proposition A.2, for a critical value c, consider confidence sets centered around the GMM estimator $\hat{\theta}_\Omega$, $\hat{C}_\Omega(c) = [\hat{\theta}_\Omega \pm c\sigma_\Omega]$. Then $\sqrt{J}$ is the smallest critical value such that all confidence sets intersect:*

$$\sqrt{J} = \inf\left\{c \geq 0 : \bigcap_\Omega C_\Omega(c) \neq \varnothing\right\}. \tag{7}$$

The intersection point is the efficient GMM estimator: $\bigcap_\Omega C_\Omega(\sqrt{J}) = \{\hat{\theta}_{\Sigma^{-1}}\}$.

Proof. Suppose $c < \sqrt{J}$ and let $\tau > 0$. By the proof of Proposition A.2, there is some estimator $\hat{\theta}_\Omega$ where

$$\hat{\theta}_\Omega - \hat{\theta}_{\Sigma^{-1}} = \sigma_{\Sigma^{-1}}\tau\sqrt{J} \quad \sigma_\Omega = \sqrt{1+\tau^2}\sigma_{\Sigma^{-1}}$$

The confidence set $C_\Omega(c)$ has radius $c\sqrt{1+\tau^2}\sigma_{\Sigma^{-1}} < \tau\sqrt{J}\sigma_{\Sigma^{-1}}$ for sufficiently large τ^2. Thus, for sufficiently large τ^2, the confidence set $C_\Omega(c)$ would fail to include $\hat{\theta}_{\Sigma^{-1}}$. On the other hand, for such a τ, there is another estimator that is on the other side of $\hat{\theta}_{\Sigma^{-1}}$: $\hat{\theta}_{\Omega'} = \hat{\theta}_{\Sigma^{-1}} - \sigma_{\Sigma^{-1}}\tau\sqrt{J}$. Hence $C_\Omega(c) \cap C_{\Omega'}(c) = \varnothing$ since neither extends to $\hat{\theta}_{\Sigma^{-1}}$. For $c = \sqrt{J}$,

$$c\sqrt{1+\tau^2}\sigma_{\Sigma^{-1}} \geq \tau\sqrt{J}\sigma_{\Sigma^{-1}}$$

for all $\tau \geq 0$. Hence all confidence sets include $\hat{\theta}_{\Sigma^{-1}}$. This proves (7).

Given $\hat{\theta} \neq \hat{\theta}_{\Sigma^{-1}}$, without loss, suppose $\hat{\theta} < \hat{\theta}_{\Sigma^{-1}}$. Then for $\hat{\theta}_\Omega = \hat{\theta}_{\Sigma^{-1}} + \sigma_{\Sigma^{-1}} \tau \sqrt{J}$,

$$|\hat{\theta}_\Omega - \hat{\theta}| = \sigma_{\Sigma^{-1}} \tau \sqrt{J} + \hat{\theta}_{\Sigma^{-1}} - \hat{\theta} > \sqrt{J}\sqrt{1+\tau^2}\sigma_{\Sigma^{-1}}$$

for sufficiently large τ, since

$$\sqrt{1+\tau^2} - \tau = O(\tau^{-1})$$

as $\tau \to \infty$. Thus the confidence set around $\hat{\theta}_\Omega$ would exclude $\hat{\theta}$. This proves that $\bigcap_\Omega C_\Omega(\sqrt{J}) = \{\hat{\theta}_{\Sigma^{-1}}\}$. □

Corollary A.3. *In the setup of Proposition A.2, consider the t-statistic for $H_0 : \theta = \theta_0$ of the GMM estimator with weighting matrix Ω:*

$$t_\Omega(\theta_0) = \frac{h'\hat{\phi}_\Omega - \theta_0}{\sqrt{\mathrm{Var}(h'\phi_\Omega)}}.$$

Then

$$\inf_{\theta_0} \sup_{\Omega} |t_\Omega(\theta_0)| = \sqrt{J}.$$

Proof. This is immediate from Corollary A.4. In particular, note that by Corollary A.4, $t_\Omega(\hat{\theta}_{\Sigma^{-1}}) \leq \sqrt{J}$ for all Ω, while for any θ there exists Ω with $t_\Omega(\theta) \geq \sqrt{J}$. □

Proposition A.5. *Suppose Γ is full-rank and Σ is positive definite. Then (5) is equivalent to (8) in the sense that there is some known and invertible matrix Q whereby*

$$Y^* = QY \quad \eta^* = Q\eta \quad \Sigma^* = Q\Sigma Q'.$$

Moreover, recall that $\hat{\theta}_\Omega = -h'(\Gamma'_P \Omega_P \Gamma)^{-1}\Gamma'_P \Omega_P Y$, $\sigma^2_\Omega = \mathrm{Var}(\hat{\theta}_\Omega)$, $\hat{\theta}_{\Sigma^{-1}} = \hat{\theta}_{\Sigma^{-1}}$, $\sigma^2_{\Sigma^{-1}} = \sigma^2_{\Sigma^{-1}}$. Then,

(1) GMM estimators can be written as $\hat{\theta}_\Omega = h'\hat{\phi}_\Omega = h'\hat{\phi}_{\Sigma^{-1}} + v'_\Omega Z$, where any v_Ω corresponds to some choice of the weighting matrix

$$\{v_\Omega : \Omega \text{ is positive definite}\} = \mathbb{R}^{k-p}.$$

(2) $\sigma^2_\Omega = \sigma^2_{\Sigma^{-1}} + v_\Omega' \tilde{\Sigma}^*_Z v_\Omega$

(3) The J-statistic is the norm of Z*:* $J = Z' \left(\Sigma^*_Z\right)^{-1} Z$*.*

Proof. Let $\Lambda = -(\Gamma'\Sigma^{-1}\Gamma)^{-1}\Gamma'\Sigma^{-1} \in \mathbb{R}^{p\times k}$ and let $M = \tilde{M}(I + \Gamma\Lambda)$ where $\tilde{M}$ is any matrix so that M is full-rank. $\tilde{M}$ exists since $-\Gamma\Lambda$ is idempotent and of rank p. Thus $I + \Gamma\Lambda$ is of rank $k - p$. Let

$$Q = \begin{bmatrix} \Lambda \\ M \end{bmatrix}.$$

By multiplication, it is easy to verify that

$$Q^{-1} = [-\Gamma, \Sigma M'(M\Sigma M')^{-1}].$$

It is also easy to verify that $M\Gamma = 0 = \Lambda\Sigma M'$. With this choice of Q,

$$\Sigma^* = Q\Sigma Q' = \begin{bmatrix} \Lambda\Sigma \\ M\Sigma \end{bmatrix} [\Lambda', M'] = \begin{bmatrix} \Lambda\Sigma\Lambda' & \\ & M\Sigma M' \end{bmatrix}$$

is diagonal, as desired. Note that

$$QY = \begin{bmatrix} \hat{\phi}_{\Sigma^{-1}} \\ MY \end{bmatrix} \quad M\Gamma\phi = 0$$

is of the desired form as well.

For the first claim after "moreover," write

$$\begin{aligned} -(\Gamma'\Omega\Gamma)^{-1}\Gamma'\Omega Y &= -(\Gamma'\Omega\Gamma)^{-1}\Gamma'\Omega \begin{bmatrix} \Lambda \\ M \end{bmatrix}^{-1} \begin{bmatrix} \hat{\phi}_{\Sigma^{-1}} \\ Z \end{bmatrix} \\ &= \hat{\phi}_{\Sigma^{-1}} - (\Gamma'\Omega\Gamma)^{-1}\Gamma'\Omega\Sigma M'(M\Sigma M')^{-1} Z. \end{aligned}$$

Thus,

$$v_\Omega = -(M\Sigma M')^{-1} M\Sigma\Omega\Gamma(\Gamma'\Omega\Gamma)^{-1} h.$$

We first show that

$$\begin{aligned}\left\{\Omega\Gamma(\Gamma'\Omega\Gamma)^{-1}h : \Omega\right\} &= \left\{q \in \mathbb{R}^k : \Gamma' q = h\right\} \\ &= \left\{q \in \mathbb{R}^k : q = q_0 + v, \Gamma' q_0 = h, v \in N(\Gamma')\right\}\end{aligned} \tag{9}$$

The $\subset$ direction in (9) is immediate. For the $\supset$ direction, fix some q where $\Gamma' q = h$. We can uniquely write

$$q = u + w$$

where $u \in R(\Gamma)$ and $w \in N(\Gamma')$. Note that we can write u as its own projection onto $R(\Gamma)$, $u = \Gamma(\Gamma'\Gamma)^{-1}\Gamma' u$. Since $\Gamma' q = \Gamma' u = h$, we have $u = \Gamma(\Gamma'\Gamma)^{-1}h$. We can decompose

$$\mathbb{R}^k = R(\Gamma) \oplus \{cw : c \in \mathbb{R}\} \oplus V$$

for subspace V the orthocomplement to $R(\Gamma) \oplus \{cw : c \in \mathbb{R}\}$. For some $a > 0$, take

$$\Omega = \Gamma\Gamma' + a(uw' + wu') + aww' + P_V$$

for P_V the projection matrix onto V. Note that by construction, any vector t can be written as

$$t = \Gamma t_1 + cw + t_2 \quad t_1 \in \mathbb{R}^p \quad t_2 \in V$$

Choose $a = \frac{1}{h'(\Gamma'\Gamma)^{-2}h} > 0$. Observe that Ω is positive definite:

$$\begin{aligned} t'\Omega t &= t_1(\Gamma'\Gamma)^2 t_1 + ac^2\|w\|^4 + \|t_2\|^2 + 2ac\|w\|^2 t_1' h \\ &\geq a(t_1' h)^2 + ac^2\|w\|^4 + 2ac\|w\|^2 t_1' h + \|t_2\|^2 \\ &\qquad \text{(By Cauchy–Schwarz, } (t_1' h)^2 \leq (t_1'(\Gamma'\Gamma)^2 t_1)(h'(\Gamma'\Gamma)^{-2}h)\text{)} \\ &= a(t_1' h + c\|w\|^2)^2 + \|t_2\|^2 \geq 0. \end{aligned}$$

Now,

$$\Gamma'\Omega\Gamma = (\Gamma'\Gamma)^2 \qquad \Omega\Gamma = \Gamma(\Gamma'\Gamma) + awu'\Gamma = \Gamma(\Gamma'\Gamma) + awh'$$

and thus

$$\Omega\Gamma(\Gamma'\Omega\Gamma)^{-1} = \Gamma(\Gamma'\Gamma)^{-1} + awh'(\Gamma'\Gamma)^{-2} \implies \Omega\Gamma(\Gamma'\Omega\Gamma)^{-1}h = \Gamma(\Gamma'\Gamma)^{-1}h + w = u + w = q$$

This proves (9).

Thus, the set of v_Ω, as Ω ranges over all weighting matrices, is the set

$$\left\{-(M\Sigma M')^{-1}M\Sigma q_0 - (M\Sigma M')^{-1}M\Sigma v : v \in N(\Gamma')\right\}$$

Thus, for (1), it suffices to show that $\{M\Sigma v : v \in N(\Gamma')\} = \mathbb{R}^{k-p}$. It further suffices to show that

$$\left\{\Sigma^{1/2}v : v \in N(\Gamma')\right\} = N(M\Sigma^{1/2})^\perp.$$

To that end, note that

$$\begin{aligned}
\left\{\Sigma^{1/2}v : v \in N(\Gamma')\right\} &= \left\{u : \Gamma'\Sigma^{-1/2}u = 0\right\} = N(\Gamma'\Sigma^{-1/2}) \\
N(M\Sigma^{1/2}) &= \left\{v : M\Sigma^{1/2}v = 0\right\} = \left\{v : \Sigma^{1/2}v \in R(\Gamma)\right\} \qquad (N(M) = R(\Gamma)) \\
&= \left\{\Sigma^{-1/2}\Gamma u : u \in \mathbb{R}^p\right\} \\
&= R(\Sigma^{-1/2}\Gamma) = N(\Gamma'\Sigma^{-1/2})^\perp.
\end{aligned}$$

This verifies (1).

(2) follows immediately from the observation that Σ^* is block-diagonal.

For (3), we may compute that

$$\begin{aligned}
J &= \|Y + \Gamma\hat{\phi}_{\Sigma^{-1}}\|^2_{\Sigma^{-1}} \\
&= \|Q^{-1}(QY + Q\Gamma\hat{\phi}_{\Sigma^{-1}})\|^2_{\Sigma^{-1}} \\
&= \|QY + Q\Gamma\hat{\phi}_{\Sigma^{-1}}\|^2_{(Q'\Sigma Q)^{-1}} \\
QY + Q\Gamma\hat{\phi}^* &= \begin{bmatrix}\hat{\phi}_{\Sigma^{-1}} \\ Z\end{bmatrix} - \begin{bmatrix}\hat{\phi}_{\Sigma^{-1}} \\ 0\end{bmatrix} = \begin{bmatrix}0 \\ Z\end{bmatrix} \\
Q'\Sigma Q &= \Sigma^*.
\end{aligned}$$

Finally, observe that

$$\begin{bmatrix} 0 \\ Z \end{bmatrix}' \Sigma^{*-1} \begin{bmatrix} 0 \\ Z \end{bmatrix} = Z' \Sigma_Z^{*\,-1} Z.$$

This completes the proof. □

Appendix B. Proofs for Results in Main Text

Proposition 5.2. *Under Assumption 5.1, and for*

$$d_H(A,B) = \max\left\{ \sup_{a\in A} \inf_{b\in B} |a-b|, \sup_{b\in B} \inf_{a\in A} |a-b| \right\}$$

the Hausdorff distance between the sets A and B, for any $\varepsilon > 0$,

$$\lim_{\kappa\to\infty} \limsup_{n\to\infty} \mathrm{P}\left(\sqrt{n} \cdot d_H\left(\hat{\Theta}\left(\kappa,\tau\right), \hat{\Theta}(\tau) \right) > \varepsilon \right) = 0.$$

Proof. We have assumed that for all fixed $\kappa \geq 1$, and $\phi_\Omega^* = -\sqrt{n} \cdot (\Gamma'\Omega\Gamma)^{-1}\Gamma'\Omega g_n(\psi)$

$$\sqrt{n}(\hat{\psi}_\Omega - \psi) - \phi_\Omega^* \xrightarrow{p} 0,\ \hat{\Gamma}_\Omega \xrightarrow{p} \Gamma,\ \hat{\Sigma}_\Omega \xrightarrow{p} \Sigma$$

uniformly over $\Omega \in \mathcal{W}_\kappa$. Since we have also assumed ϑ is continuously differentiable and Γ, Σ are full rank, it follows immediately that for $\hat{\theta}_\Omega^* = h'\phi_\Omega^*$

$$\sqrt{n}(\hat{\theta}_\Omega - \vartheta(\psi)) - \hat{\theta}_\Omega^* \xrightarrow{p} 0,\ \sqrt{n}\hat{\sigma}_\Omega \xrightarrow{p} \sigma_\Omega = \left(h'\left(\Gamma'\Omega\Gamma\right)^{-1}\Gamma'\Omega\Sigma\Omega\Gamma\left(\Gamma'\Omega\Gamma\right)^{-1} h \right)^{1/2}, \quad (10)$$

again uniformly over $\mathcal{W}_\kappa$. Since $\Sigma^{-1} \in \mathcal{W}_\kappa$ for κ sufficiently large (which we assume for the remainder of the proof), it further follows that

$$\sqrt{n} g_n\left(\hat{\psi}_{\Sigma^{-1}}\right) - \sqrt{n} g_n\left(\psi\right) - \Gamma\phi_{\Sigma^{-1}}^* \xrightarrow{p} 0,$$

and that for

$$J = n \cdot g_n\left(\hat{\psi}_{\Sigma^{-1}}\right)' \hat{\Sigma} g_n\left(\hat{\psi}_{\Sigma^{-1}}\right)$$

and

$$J^* = \min_u (\sqrt{n} g_n(\psi) + \Gamma u)\Sigma^{-1}(\sqrt{n} g_n(\psi) + \Gamma u),$$

where the right hand side is minimized at $\phi^*_{\Sigma^{-1}}$, $J - J^* \xrightarrow{p} 0$.

Let us define analogs of $\hat{\Theta}(\kappa, \tau)$ and $\hat{\Theta}(\tau)$ based on $\hat{\phi}^*_\Omega$,

$$\hat{\Theta}^*(\kappa, \tau) = \left\{ h'\hat{\phi}^*_\Omega : \Omega \in \mathcal{W}_\kappa, \sigma^2_\Omega \leq (1+\tau^2)\sigma^2_{\Sigma^{-1}} \right\}$$

$$\hat{\Theta}^*(\tau) = \left\{ h'\hat{\phi}^*_{\Sigma^{-1}} \pm \tau\sqrt{J^*}\sigma_{\Sigma^{-1}} \right\}.$$

Note that by the triangle inequality and the definition of Hausdorff distance,

$$\begin{aligned}
&\sqrt{n} \cdot d_H\left(\hat{\Theta}(\kappa,\tau), \hat{\Theta}(\tau)\right) = d_H\left(\sqrt{n}\cdot\left(\hat{\Theta}(\kappa,\tau) - \vartheta(\psi)\right), \sqrt{n}\cdot\left(\hat{\Theta}(\tau) - \vartheta(\psi)\right)\right) \\
&\leq d_H\left(\sqrt{n}\cdot\left(\hat{\Theta}(\kappa,\tau) - \vartheta(\psi)\right), \hat{\Theta}^*(\kappa,\tau)\right) + d_H\left(\hat{\Theta}^*(\kappa,\tau), \hat{\Theta}^*(\tau)\right) \\
&\quad + d_H\left(\sqrt{n}\cdot\left(\hat{\Theta}(\tau) - \vartheta(\psi)\right), \hat{\Theta}^*(\tau)\right).
\end{aligned}$$

We will show that the first and third terms on the right hand side converge in probability to zero as $n \to \infty$ for fixed κ, while the second term converges in probability to zero as $n \to \infty$ and $\kappa \to \infty$, in that order.

Let us begin with the third term. Since

$$\sqrt{n}\cdot\left(\hat{\Theta}(\tau) - \vartheta(\psi)\right) = \left[\sqrt{n}(\hat{\theta}_{\Sigma^{-1}} - \vartheta(\psi)) \pm \tau\sqrt{J}\sqrt{n}\sigma_{\Sigma^{-1}}\right],$$

(10) immediately implies that $\text{plim}\left(d_H\left(\sqrt{n}\cdot\left(\hat{\Theta}(\tau) - \vartheta(\psi)\right), \hat{\Theta}^*(\tau)\right)\right) = 0$.

Next, let us turn to the first term, and show that

$$\text{plim}\left(d_H\left(\sqrt{n}\cdot\left(\hat{\Theta}(\kappa,\tau) - \vartheta(\psi)\right), \hat{\Theta}^*(\kappa,\tau)\right)\right) = 0. \tag{11}$$

To this end, define

$$\hat{\mathcal{W}}_{\kappa,\tau} = \left\{\Omega \in \mathcal{W}_\kappa : \hat{\sigma}_\Omega \leq \sqrt{(1+\tau^2)}\hat{\sigma}_{\Sigma^{-1}}\right\}$$

and note that since

$$\sqrt{n}\cdot\left(\hat{\Theta}(\kappa,\tau) - \vartheta(\psi)\right) = \left\{\sqrt{n}\left(\vartheta(\hat{\psi}_\Omega) - \vartheta(\psi)\right) : \Omega \in \hat{\mathcal{W}}_{\kappa,\tau}\right\},$$

(10) implies that

$$d_H\left(\sqrt{n}\cdot\left(\hat{\Theta}\left(\kappa,\tau\right)-\vartheta(\psi)\right),\left\{h'\hat{\phi}^*_\Omega:\Omega\in\hat{\mathcal{W}}_{\kappa,\tau}\right\}\right)\xrightarrow{p}0.$$

Hence, to prove the convergence (11) it suffices to show that

$$d_H\left(\left\{h'\hat{\phi}^*_\Omega:\Omega\in\hat{\mathcal{W}}_{\kappa,\tau}\right\},\hat{\Theta}^*\left(\kappa,\tau\right)\right)\xrightarrow{p}0.$$

Note, however, that if we define

$$\mathcal{W}_{\kappa,\tau}=\left\{\Omega\in\mathcal{W}_\kappa:\sigma_\Omega\leq\sqrt{(1+\tau^2)}\sigma_{\Sigma^{-1}}\right\},$$

then $\lim_{\delta\to0}d_H\left(\mathcal{W}_{\kappa,\tau},\mathcal{W}_{\kappa,\tau+\delta}\right)=0$. Consequently, since (10) implies that for $\delta>0$

$$\mathrm{P}\left\{\mathcal{W}_{\kappa,\tau-\delta}\subseteq\hat{\mathcal{W}}_{\kappa,\tau}\subseteq\mathcal{W}_{\kappa,\tau+\delta}\right\}\to1,$$

we see that $d_H\left(\mathcal{W}_{\kappa,\tau},\hat{\mathcal{W}}_{\kappa,\tau}\right)\xrightarrow{p}0$. Note next that $\frac{\phi^*_\Omega}{\left\|\sqrt{n}g_n(\psi)\right\|}$ is a continuous function of $\left(\Omega,\frac{\sqrt{n}g_n(\psi)}{\left\|\sqrt{n}g_n(\psi)\right\|}\right)\in\mathcal{W}_\kappa\times\mathbb{S}^k$ for $\mathbb{S}^k$ the unit sphere in $\mathbb{R}^k$. Since $\mathcal{W}_\kappa\times\mathbb{S}^k$ is compact, it follows that $\frac{\phi^*_\Omega}{\left\|\sqrt{n}g_n(\psi)\right\|}$ is uniformly continuous. Convergence of $\hat{\mathcal{W}}_{\kappa,\tau}$ to $\mathcal{W}_{\kappa,\tau}$ thus implies that

$$\left\|\sqrt{n}g_n\left(\psi\right)\right\|^{-1}d_H\left(\left\{h'\hat{\phi}^*_\Omega:\Omega\in\hat{\mathcal{W}}_{\kappa,\tau}\right\},\hat{\Theta}^*\left(\kappa,\tau\right)\right)\xrightarrow{p}0.$$

However, the convergence of $\sqrt{n}g_n\left(\psi\right)$ to a normal implies that $\left\|\sqrt{n}g_n\left(\psi\right)\right\|^{-1}$ converges in distribution to a random variable with zero mass at zero, and thus that (11) holds, as we wished to show.

Our argument thus far implies that for each $\kappa>0$, and each $\nu>0$,

$$\liminf_{n\to\infty}\mathrm{P}\left(\sqrt{n}\cdot d_H\left(\hat{\Theta}\left(\kappa,\tau\right),\hat{\Theta}(\tau)\right)\leq d_H\left(\hat{\Theta}^*\left(\kappa,\tau\right),\hat{\Theta}^*\left(\tau\right)\right)+\nu\right)=1.$$

To complete the proof, we wish to show that the right hand side converges to zero as $\kappa\to\infty$. To this end, note that $\hat{\Theta}^*\left(\kappa,\tau\right)$ and $\hat{\Theta}^*\left(\tau\right)$ both depend on the data only through $\sqrt{n}g_n(\psi)$, and let

$$\mathcal{G}(\kappa,\zeta)=\left\{\sqrt{n}g_n(\psi):d_H\left(\hat{\Theta}^*\left(\kappa,\tau\right),\hat{\Theta}^*\left(\tau\right)\right)<\zeta\right\}.$$

Note that since $\mathcal{W}_\kappa$ is a compact, connected set and $h'\hat{\phi}^*_\Omega$ is continuous in Ω, $\hat{\Theta}^*(\kappa,\tau)$ is a closed interval. Since $\hat{\Theta}^*(\tau)$ is likewise an interval, $d_H\left(\hat{\Theta}^*(\kappa,\tau),\hat{\Theta}^*(\tau)\right)$ is simply the maximum of the distance between the upper and lower endpoints. Moreover, since the endpoints of both intervals vary continuously in $\sqrt{n}g_n(\psi)$, $d_H\left(\hat{\Theta}^*(\kappa,\tau),\hat{\Theta}^*(\tau)\right)$ is also continuous in $\sqrt{n}g_n(\psi)$, so $\mathcal{G}(\kappa,\zeta)$ is an open set. By the Portmanteau lemma and convergence in distribution of $\sqrt{n}g_n(\psi)$, it follows that

$$\liminf_{n\to\infty} \mathrm{P}\left\{d_H\left(\hat{\Theta}^*(\kappa,\tau),\hat{\Theta}^*(\tau)\right)<\zeta\right\} \geq \mathrm{P}\left\{\mathcal{N}(\eta,\Sigma)\in\mathcal{G}(\kappa,\zeta)\right\},$$

and thus that the asymptotic probability that $d_H\left(\hat{\Theta}^*(\kappa,\tau),\hat{\Theta}^*(\tau)\right)<\zeta$ is lower bounded by the analogous probability in the normal limit problem. The proof of Proposition A.2 implies that for each fixed $\sqrt{n}g_n(\psi)$, however, $d_H\left(\hat{\Theta}^*(\kappa,\tau),\hat{\Theta}^*(\tau)\right)=0$ for κ sufficiently large, so

$$\lim_{\kappa\to\infty} \mathrm{P}\left\{N(\eta,\Sigma)\in\mathcal{G}(\kappa,\zeta)\right\}=1,$$

which implies the result.

□

Corollary 5.3. *Under the assumptions of Proposition 5.2, consider the t-statistic for testing $H_0:\theta=\theta_0$ based on the GMM estimator with weighting matrix Ω:*

$$t_\Omega(\theta_0)=\frac{\hat{\theta}_\Omega-\theta_0}{\hat{\sigma}_\Omega}.$$

Then for all $\varepsilon>0$,

$$\lim_{\kappa\to\infty}\limsup_{n\to\infty}\mathrm{P}\left(\left|\inf_{\theta_0\in\mathbb{R}}\sup_{\Omega\in\mathcal{W}_\kappa}|t_\Omega(\theta_0)|-\sqrt{J}\right|>\varepsilon\right)=0.$$

Proof. Note that

$$t_\Omega(\theta_0)=\frac{\hat{\theta}_\Omega-\theta_0}{\hat{\sigma}_\Omega}=\frac{\sqrt{n}(\hat{\theta}_\Omega-\vartheta(\psi))}{\sqrt{n}\hat{\sigma}_\Omega}+\frac{\sqrt{n}(\vartheta(\psi)-\theta_0)}{\sqrt{n}\hat{\sigma}_\Omega}$$

Hence, by (10), for any $\kappa \geq 1$

$$t_\Omega(\theta_0) \to_p t^*_\Omega(\theta_0) = \frac{\hat{\theta}^*_\Omega}{\sigma_\Omega} + \frac{\vartheta(\psi) - \theta_0}{\sigma_\Omega}$$

uniformly over $\Omega \in \mathcal{W}_\kappa$ and $\theta_0 \in [\vartheta(\psi) \pm \frac{1}{\sqrt{n}}C]$ for any fixed C. This implies, in particular, that

$$\inf_{\theta_0 \in [\vartheta(\psi) \pm \frac{1}{\sqrt{n}}C]} \sup_{\Omega \in \mathcal{W}_\kappa} |t_\Omega(\theta_0)| \xrightarrow{p} \inf_{\theta_0 \in [\vartheta(\psi) \pm \frac{1}{\sqrt{n}}C]} \sup_{\Omega \in \mathcal{W}_\kappa} |t^*_\Omega(\theta_0)| \,. \tag{12}$$

Note that by the proof of Corollary A.3, $\left|t^*_\Omega\left(\vartheta(\psi) + \frac{1}{\sqrt{n}}\hat{\theta}^*_{\Sigma^{-1}}\right)\right| \leq \sqrt{J^*}$ for all Ω. Hence, for any $C \geq \left|\hat{\theta}^*_{\Sigma^{-1}}\right|$, the right-hand side is bounded above by $\sqrt{J^*}$, regardless of κ. Since $\sqrt{n}g_n(\psi)$ converges in distribution to a tight limit, so does $\hat{\theta}^*_{\Sigma^{-1}}$, so since the left hand side of (12) is decreasing in C, and $\sqrt{J} - \sqrt{J^*} \xrightarrow{p} 0$ as shown in the proof of Proposition 5.2, we have that

$$\limsup_{n\to\infty} \mathrm{P}\left\{\inf_{\theta_0} \sup_{\Omega \in \mathcal{W}_k} |t_\Omega(\theta_0)| > \sqrt{J} + \varepsilon\right\} = 0 \tag{13}$$

for all $\kappa \geq 1$, $\varepsilon > 0$.

Conversely, note that for any $\tau > 0$ and any $\varepsilon > 0$, Proposition 5.2 implies that there exist κ and (random) $\Omega_+, \Omega_- \in \mathcal{W}_\kappa$ such that $\hat{\sigma}_{\Omega_+}$, $\hat{\sigma}_{\Omega_-} \leq \sqrt{1+\tau^2}\hat{\sigma}_{\Sigma^{-1}}$ and

$$\limsup_{n\to\infty} \mathrm{P}\left\{\sqrt{n}\left|\hat{\theta}_{\Omega_+} - \hat{\theta}_{\Sigma^{-1}} - \sqrt{J}\hat{\sigma}_{\Sigma^{-1}}\tau\right| > \varepsilon\right\} < \varepsilon$$

$$\limsup_{n\to\infty} \mathrm{P}\left\{\sqrt{n}\left|\hat{\theta}_{\Omega_-} - \hat{\theta}_{\Sigma^{-1}} + \sqrt{J}\hat{\sigma}_{\Sigma^{-1}}\tau\right| > \varepsilon\right\} < \varepsilon$$

Since this holds for all $\varepsilon > 0$, for all $\varepsilon' > 0$ there exists κ and $\Omega_+, \Omega_- \in \mathcal{W}_\kappa$ such that

$$\limsup_{n\to\infty} \mathrm{P}\left\{\left|t_{\Omega_+}(\hat{\theta}_{\Sigma^{-1}})\right| \leq \frac{\tau}{\sqrt{1+\tau^2}}\sqrt{J} - \varepsilon'\right\} < \varepsilon',$$

$$\limsup_{n\to\infty} \mathrm{P}\left\{\left|t_{\Omega_-}(\hat{\theta}_{\Sigma^{-1}})\right| \leq \frac{\tau}{\sqrt{1+\tau^2}}\sqrt{J} - \varepsilon'\right\} < \varepsilon'.$$

Note, however, that for all θ_0,

$$\sup_{\Omega \in \mathcal{W}_\kappa} |t_\Omega(\theta_0)| \geq \max\left\{\left|t_{\Omega_+}(\theta_0)\right|, \left|t_{\Omega_-}(\theta_0)\right|\right\}.$$

However, since $|t_{\Omega_+}(\theta_0)|$ is increasing in θ_0 for $\theta_0 \leq \hat{\theta}_{\Sigma^{-1}}$ while $|t_{\Omega_+}(\theta_0)|$ is increasing in θ_0 for $\theta_0 \geq \theta_{\Sigma^{-1}}$,

$$\max\left\{\left|t_{\Omega_+}(\theta_0)\right|, \left|t_{\Omega_-}(\theta_0)\right|\right\} \geq \min\left\{\left|t_{\Omega_+}(\hat{\theta}_{\Sigma^{-1}})\right|, \left|t_{\Omega_-}(\hat{\theta}_{\Sigma^{-1}}))\right|\right\}.$$

However, we have shows that for κ sufficiently large, we can lower bound the right hand side by $\sqrt{J} - \varepsilon'$ for any $\varepsilon' > 0$. Thus, for all $\varepsilon > 0$ and $\tau > 0$,

$$\lim_{\kappa\to\infty} \limsup_{n\to\infty} \mathrm{P}\left\{\inf_{\theta_0} \sup_{\Omega\in\mathcal{W}_\kappa} |t_\Omega(\theta_0)| < \frac{\tau}{\sqrt{1+\tau^2}}\sqrt{J} - \varepsilon\right\} = 0.$$

Together with (13), this implies the result.

□

Corollary 5.4. *Under the assumptions of Proposition 5.2, for a critical value c, consider confidence sets centered around the GMM estimator $\hat{\theta}_\Omega$, $\hat{C}_\Omega(c) = [\hat{\theta}_\Omega \pm c\hat{\sigma}_\Omega]$. Then for all $\varepsilon > 0$,*

$$\lim_{\kappa\to\infty} \limsup_{n\to\infty} \mathrm{P}\left(\left|\inf\left\{c \geq 0 : \bigcap_{\Omega\in\mathcal{W}_\kappa} C_\Omega(c) \neq \varnothing\right\} - \sqrt{J}\right| > \varepsilon\right) = 0.$$

The intersection point is asymptotically equal to the efficient GMM estimator:

$$\lim_{\nu\to 0} \lim_{\kappa\to\infty} \limsup_{n\to\infty} \mathrm{P}\left(\left|\sqrt{n}\cdot d_H\left(\bigcap_{\Omega\in\mathcal{W}_\kappa} C_\Omega(\sqrt{J}+\nu), \left\{\hat{\theta}_{\Sigma^{-1}}\right\}\right)\right| > \varepsilon\right) = 0.$$

Proof. The first part of the result is immediate from Corollary 5.3. For the second part, note that the proof of Corollary 5.3 shows that $\left|t^*_\Omega\left(\vartheta(\psi) + \frac{1}{\sqrt{n}}\hat{\theta}^*_{\Sigma^{-1}}\right)\right| \leq \sqrt{J^*}$ for all Ω, while by (10), $t^*_\Omega\left(\vartheta(\psi) + \frac{1}{\sqrt{n}}\hat{\theta}^*_{\Sigma^{-1}}\right) - t_\Omega(\hat{\theta}_{\Sigma^{-1}}) \xrightarrow{p} 0$ uniformly over $\Omega \in \mathcal{W}_\kappa$ and $\sqrt{J^*} - \sqrt{J} \xrightarrow{p} 0$. Thus,

$$\sup_{\Omega\in\mathcal{W}_\kappa} \left|t_\Omega\left(\hat{\theta}_{\Sigma^{-1}}\right) - \sqrt{J}\right| \xrightarrow{p} 0,$$

and $\hat{\theta}_{\Sigma^{-1}} \in \bigcap_{\Omega\in\mathcal{W}_\kappa} C_\Omega(\sqrt{J}+\nu)$ with probability going to one for any $\nu > 0$ and any κ.

On the other hand, note that since $C_\Omega(c)$ is an interval for all c, $\bigcap_{\Omega\in\mathcal{W}_\kappa} C_\Omega(\sqrt{J}+\nu)$ is also an interval. As argued in the proof of Corollary 5.3, for any $\tau > 0$ and any $\nu' > 0$ there

exists $\kappa > 0$ and (random) Ω_+ and $\Omega_- \in \mathcal{W}_\kappa$ such that $\hat{\sigma}_{\Omega_+}$, $\hat{\sigma}_{\Omega_-} \leq \sqrt{1+\tau^2}\hat{\sigma}_{\Sigma^{-1}}$ and

$$\limsup_{n\to\infty} \mathrm{P}\left\{\left|t_{\Omega_+}(\hat{\theta}_{\Sigma^{-1}})\right| \leq \frac{\tau}{\sqrt{1+\tau^2}}\sqrt{J} - \nu'\right\} < \nu',$$

$$\limsup_{n\to\infty} \mathrm{P}\left\{\left|t_{\Omega_-}(\hat{\theta}_{\Sigma^{-1}})\right| \leq \frac{\tau}{\sqrt{1+\tau^2}}\sqrt{J} - \nu'\right\} < \nu'.$$

As also argued in the proof of Corollary 5.3, for all θ_0,

$$\sup_{\Omega\in\mathcal{W}_\kappa} |t_\Omega(\theta_0)| \geq \min\left\{\left|t_{\Omega_+}(\theta_0)\right|, \left|t_{\Omega_-}(\theta_0)\right|\right\}.$$

Note, moreover, that $|t_{\Omega_+}(\theta_0)|$ is decreasing in θ_0 for $\theta_0 \leq \hat{\theta}_{\Sigma^{-1}}$, with slope $1/\hat{\sigma}_{\Omega_+}$, while $|t_{\Omega_-}(\theta_0)|$ is increasing in θ_0 for $\theta_0 \geq \hat{\theta}_{\Sigma^{-1}}$, with slope $1/\hat{\sigma}_{\Omega_-}$. Consequently, for all θ_0

$$\sup_{\Omega\in\mathcal{W}_\kappa} |t_\Omega(\theta_0)| \geq \min\left\{t_{\Omega_+}\left(\hat{\theta}_{\Sigma^{-1}}\right), t_{\Omega_-}\left(\hat{\theta}_{\Sigma^{-1}}\right)\right\} + \frac{\left|\hat{\theta}_{\Sigma^{-1}} - \theta_0\right|}{\max\left\{\hat{\sigma}_{\Omega_+}, \hat{\sigma}_{\Omega_-}\right\}} \geq$$

$$\min\left\{t_{\Omega_+}\left(\hat{\theta}_{\Sigma^{-1}}\right), t_{\Omega_-}\left(\hat{\theta}_{\Sigma^{-1}}\right)\right\} + \frac{\left|\hat{\theta}_{\Sigma^{-1}} - \theta_0\right|}{\sqrt{1+\tau^2}\hat{\sigma}_{\Sigma^{-1}}}.$$

Consequently, for any δ,

$$\limsup_{n\to\infty} \mathrm{P}\left\{\sup_{\Omega\in\mathcal{W}_\kappa}\left|t_\Omega\left(\hat{\theta}_{\Sigma^{-1}} + \frac{1}{\sqrt{n}}\delta\right)\right| \leq \frac{\tau}{\sqrt{1+\tau^2}}\sqrt{J} + \frac{|\delta|}{\sqrt{1+\tau^2}\sigma_{\Sigma^{-1}}} - \nu'\right\} < 2\nu'.$$

Since we can repeat this argument for all $\nu' > 0$ by taking κ sufficiently large, it follows that for all ν' and $\tau > 0$,

$$\lim_{\kappa\to\infty}\limsup_{n\to\infty} \mathrm{P}\left\{\sup_{\Omega\in\mathcal{W}_\kappa}\left|t_\Omega\left(\hat{\theta}_{\Sigma^{-1}} + \frac{1}{\sqrt{n}}\delta\right)\right| > \frac{\tau}{\sqrt{1+\tau^2}}\sqrt{J} + \frac{|\delta|}{\sqrt{1+\tau^2}\sigma_{\Sigma^{-1}}} - \nu'\right\} = 1.$$

Since J converges in distribution to a tight random variable, however, for all $\delta \neq 0$

$$\limsup_{\nu\to 0}\sup_{\tau>0}\liminf_{n\to\infty} \mathrm{P}\left\{\sqrt{J} + \nu < \frac{\tau}{\sqrt{1+\tau^2}}\sqrt{J} + \frac{|\delta|}{\sqrt{1+\tau^2}\sigma_{\Sigma^{-1}}}\right\} = 1,$$

and consequently

$$\lim_{\nu\to 0}\lim_{\kappa\to\infty}\limsup_{n\to\infty} \mathrm{P}\left\{\sup_{\Omega\in\mathcal{W}_\kappa}\left|t_\Omega\left(\hat{\theta}_{\Sigma^{-1}} + \frac{1}{\sqrt{n}}\delta\right)\right| > \sqrt{J} + \nu\right\} = 1.$$

Since this holds for all $\delta \neq 0$ and $\bigcap_{\Omega \in \mathcal{W}_\kappa} C_\Omega(\sqrt{J} + \nu)$ is an interval, the second part of the result is immediate. $\square$