A Misuse of Specification Tests

Naoya Sueishi*

Kobe University†

January 22, 2023

Abstract

Empirical researchers often perform model specification tests, such as the Hausman test and the overidentifying restrictions test, to confirm the validity of estimators rather than the validity of models. This paper examines the effectiveness of specification pretests in finding invalid estimators. We study the local asymptotic properties of test statistics and estimators and show that locally unbiased specification tests cannot determine whether asymptotically efficient estimators are asymptotically biased. The main message of the paper is that correct specification and valid estimation are different issues. Correct specification is neither necessary nor sufficient for asymptotically unbiased estimation under local overidentification.

*This research was supported by JSPS KAKENHI Grant Number 21K01427.

†Graduate School of Economics, 2-1 Rokkodai-cho, Nada-ku, Kobe, Hyogo 657-8501, Japan. Email: sueishi@econ.kobe-u.ac.jp. Phone: +81-78-803-6827.

1 Introduction

Model specification tests such as the Hausman test (Hausman 1978) and the overidentifying restrictions test (Sargan 1958, Hansen 1982) are commonly used in empirical studies, even though finding a true model is rarely the focus of the studies. Empirical researchers often conduct these tests to find an appropriate estimator rather than to assess the correctness of the model. For example, the Durbin-Wu-Hausman test is used to determine whether to adopt the ordinary least squares (OLS) estimator or the two-stage least squares (2SLS) estimator. Acceptance (or rejection) of the null hypothesis is interpreted as evidence of the validity of the OLS (or 2SLS) estimator.

This paper examines the effectiveness of specification pretests in finding invalid estimators. It is commonly recognized that an estimator based on a null model is valid (or invalid) if a specification test accepts (or rejects) the null hypothesis. Thus, we investigate whether specification tests have nontrivial power against a local alternative when estimators are asymptotically biased. We show that even though specification tests have nontrivial power for local misspecification, they are not useful for detecting asymptotically biased estimators.

The analytical framework is as follows. The parameter of interest θ_0 is given by $\theta_0 = \psi(P)$ for a distribution P and a functional ψ. The distribution P may be suggested by an underlying economic theory. An estimator $\hat{\theta}_n$ for θ_0 is asymptotically normally distributed and efficient if the data generating process (DGP) is P, but a sample of size n may be drawn from a local deviation from P. The deviation is due to data contamination, measurement error, model misspecification, or other reasons. If the deviation disappears at the rate $n^{-1/2}$, then $\hat{\theta}_n$ is consistent for θ_0 but may be asymptotically biased in the sense that the asymptotic distribution of $\sqrt{n}(\hat{\theta}_n - \theta_0)$ has a nonzero mean. Alternatively, a specification test may have nontrivial power against the local deviation. The asymptotic bias of the estimator and the local power of the test depend on the direction of the local deviation from P. If the directions that cause the asymptotic bias are different from those that can be detected by the test, then we cannot avoid the asymptotically biased estimator.

The main contribution of the paper is to show that locally unbiased specification tests cannot detect asymptotic bias in asymptotically efficient estimators. To show this, we borrow the framework of Chen and Santos (2018). We describe the deviation from P by a path that goes through P. We then define the direction of the deviation by the score function of the path, which can be decomposed into two orthogonal parts. We show that the asymptotic bias of asymptotically efficient estimators depends only on one part of the decomposition, while the local power of locally unbiased tests depends only on the other part of the decomposition. Since two parts are orthogonal to each other, specification tests provide no information about the existence of asymptotic bias.

Our orthogonality result is a generalization of Section 5.1.3 of Hall (2005), which examined the connection between the efficient GMM estimator and the J test. Hall (2005) gave an orthogonal decomposition of moment restrictions into identifying restrictions and overidentifying restrictions, and showed that the asymptotic bias of the GMM estimator is only affected by a local violation of identifying restrictions, while the local power of the J test is only affected by a local violation of overidentifying restrictions. Thus, the J test does not detect asymptotic bias in the GMM estimator. As discussed in Section 4, the decomposition of moment restrictions is essentially the same as the decomposition of a score function.

Another contribution of the paper is to point out a common misconception about the Hasuman test. The Hausman test compares asymptotically efficient and inefficient estimators. There it is assumed that the inefficient estimator is consistent for the parameter of interest under both null and alternative hypotheses. This assumption is inappropriate because different hypotheses correspond to different DGPs. No estimator can be consistent for the same parameter under two different DGPs in general. Although we can assume that a maintained model is correctly specified under both hypotheses, we cannot assume that an estimator is consistent under both hypotheses. We show that the Hausman test may reject the null hypothesis even when the efficient estimator is asymptotically unbiased and the inefficient estimator is asymptotically biased.

This paper is inspired by Chen and Santos (2018), who introduced the notion of local identification for general semiparametric models and showed that local overidentification is equivalent to both the existence of specification tests with nontrivial local power and the existence of asymptotically efficient estimators. In fact, main results of this paper are obtained by a simple application of their results. However, given the current state of how specification tests are used in practice, this study has its own importance.

There are many studies that investigate the impact of pretest or model selection on subsequent inference. Examples include Judge and Bock (1978), Pötscher (1991), Kabaila (1995), Leeb and Pötscher (2005, 2006), and Andrews and Guggenberger (2009a,b). Guggenberger (2010a,b), Guggenberger and Kumar (2012), and Doko Tchatoka and Wang (2021) investigated the impact of specification tests on subsequent inference. These studies show that inference that ignores the effect of pretest can be highly misleading.

The study of the local power properties of specification tests can be traced back to Newey (1985a,b). These papers pointed out that specification tests may fail to detect local deviations that make estimators asymptotically biased, although they did not investigate the orthogonality property considered in this paper.

The local asymptotic framework is also used in the context of robust estimation and inference. Kitamura et al. (2013) proposed a robust point estimator for θ_0 when a sample is obtained from a local deviation from P. Armstrong and Kolesár (2021) proposed confidence intervals that takes into account the potential bias resulting from a local deviation. See also Andrews et al. (2017) and Bonhomme and Weidner (2022) for a related issue. Given the result of this paper, it would be preferable to use robust inference methods rather than perform specification pretests to check the validity of conventional inference methods.

The organization of the rest of the paper is as follows. Section 2 introduces the setting of our analysis. Section 3 gives the main results of the paper. Sections 4 and 5 show that the results of Section 3 hold in popular models. Section 4 investigates the connection between the efficient GMM estimator and the J test statistic. Section 5 investigates the properties of the Durbin-Wu-Hausman test to test exogeneity. Section 6 concludes. The Appendix gives an auxiliary result to Section 5.

2 Preliminaries

Let P be a probability distribution defined on a set $\mathcal{X}$ and let $\mathcal{M}$ be the set of all distributions on $\mathcal{X}$. The distribution P may be suggested by an underlying economic theory, or it may be an ideal distribution from which researchers hope to draw a sample. We assume that P is an element of a semiparametric model $\mathbf{P} \subset \mathcal{M}$. The finite-dimensional parameter of interest $\theta_0 \in \Theta$ is given by $\theta_0 = \psi(P)$ for some functional $\psi : \mathbf{P} \to \Theta$.

A random sample $\{X_1, \ldots, X_n\}$ is drawn from a distribution μ. Let $\hat{\theta}_n = \hat{\theta}_n(X_1, \ldots, X_n)$ be an asymptotically efficient estimator for θ_0 if $\mu = P$, that is, it attains the minimum asymptotic variance among all regular estimators. The estimator may perform poorly if μ deviates from P. To detect the deviation, we conduct a specification test $\phi_n : \{X_i\}_{i=1}^n \to [0, 1]$ for the following null and alternative hypotheses:

$$H_0 : \mu \in \mathbf{P} \quad \text{vs.} \quad H_1 : \mu \in \mathcal{M} \setminus \mathbf{P}. \tag{2.1}$$

The test may reject the null hypothesis if $\mu \neq P$.

We investigate the relationship between $\hat{\theta}_n$ and ϕ_n when μ is a local deviation from P. Our concern is whether the test can detect local deviations that make the estimator asymptotically biased. Throughout the paper, we say that $\hat{\theta}_n$ is asymptotically biased if the asymptotic mean of $\sqrt{n}(\hat{\theta}_n - \theta_0)$ is nonzero. As we shall see shortly, not all local deviations cause asymptotic bias in $\hat{\theta}_n$. Thus, if the test is used to check the validity of the estimator, it should have nontrivial local power if and only if the estimator is asymptotically biased. The asymptotic behavior of $\hat{\theta}_n$ and ϕ_n depends on the direction of local deviation. We borrow the setting of Chen and Santos (2018) to define the direction.

The direction of the deviation from P is defined in terms of the score function of a path. A path $t \to P_t$ is a function defined on $[0, \epsilon)$ for some $\epsilon > 0$ that satisfies $P_t \in \mathcal{M}$ for all $t \in [0, \epsilon)$ and $P_0 = P$. We consider a path $t \to P_{t,g} \in \mathcal{M}$ that satisfies

$$\lim_{t \to 0} \int \left(\frac{dP_{t,g}^{1/2} - dP^{1/2}}{t} - \frac{1}{2} g dP^{1/2} \right)^2 = 0 \tag{2.2}$$

for some function $g : \mathcal{X} \mapsto \mathbb{R}$. We say that the path is Hellinger differentiable or differentiable in quadratic mean at $t = 0$ if (2.2) is satisfied for some g. The function g is referred to as the score function because it is usually given by

$$g(x) = \frac{\partial}{\partial t} \log dP_{t,g}(x) \bigg|_{t=0}.$$

If (2.2) holds for some g, then g must satisfy $\mathbb{E}[g(X)] = 0$ and $\mathbb{E}[g^2(X)] < \infty$, where $\mathbb{E}$ denotes the expectation with respect to P. Thus, the set of all possible score functions is given by

$$L_0^2(P) \equiv \left\{ g : \mathcal{X} \to \mathbb{R} : \mathbb{E}[g(X)] = 0 \text{ and } \mathbb{E}[g^2(X)] < \infty \right\}.$$

The directions consistent with the null hypothesis are specified by the tangent set. We define

$$T(P) = \left\{ g \in L_0^2(P) : (2.2) \text{ holds for some } t \mapsto P_{t,g} \in \mathbf{P} \right\}.$$

The set $T(P)$ is called the tangent set of model $\mathbf{P}$ at P. The tangent set is the set of directions along which a path P_t can deviate from P when $P_t \in \mathbf{P}$ is imposed. We consider the case where $T(P)$ is a linear space, which is the typical case in many semiparametric models.

As stated in Chen and Santos (2018), any score $g \in L_0^2(P)$ can be decomposed into two orthogonal parts. Let $\bar{T}(P)$ be the closure of $T(P)$ and let

$$\bar{T}(P)^\perp = \left\{ f \in L_0^2(P) : \mathbb{E}[f(X)g(X)] = 0 \text{ for all } g \in \bar{T}(P) \right\},$$

which is the orthogonal complement of $\bar{T}(P)$. Then, we obtain $L_0^2(P) = \bar{T}(P) \oplus \bar{T}(P)^\perp$. Thus, for any $g \in L_0^2(P)$, we have $g = \Pi_T(g) + \Pi_{T^\perp}(g)$ and $\mathrm{Var}[g(X)] = \mathrm{Var}[\Pi_T(g)(X)] + \mathrm{Var}[\Pi_{T^\perp}(g)(X)]$, where Π_T and $\Pi_{T^\perp}$ denote the projection onto $\bar{T}(P)$ and $\bar{T}(P)^\perp$, respectively, and Var denotes the variance with respect to P.

Chen and Santos (2018) showed that local overidentification is necessarily and sufficient for the existence of an asymptotically efficient estimator for θ_0 and the existence of a locally unbiased test for (2.1). They define that P is locally overidentified by $\mathbf{P}$ if $\bar{T}(P) \neq L_0^2(P)$. This definition generalizes the classical definition of overidentification defined by the number of moments and the number of parameters. Our aim is to investigate the relationship between the asymptotically efficient estimator and the locally unbiased test when μ is a local deviation from P.

Example 2.1 (J test)

The parameter of interest $\theta_0 \in \Theta \subset \mathbb{R}^p$ is a unique vector that satisfies

$$0 = \mathbb{E}[m_{\theta_0}(X)] = \int m_{\theta_0} dP, \tag{2.3}$$

where $m : \mathcal{X} \times \Theta \to \mathbb{R}^l$ is a known vector-valued function with $l > p$. If a random sample is obtained from μ, then the null hypothesis of the J test is

$$H_0 : \int m_\theta d\mu = 0 \text{ for some } \theta \in \Theta.$$

The null hypothesis can be alternatively written as $\mu \in \mathbf{P}$ where

$$\mathbf{P} = \left\{ Q \in \mathcal{M} : \int m_\theta dQ = 0 \text{ for some } \theta \in \Theta \right\}.$$

See Chen et al. (2007) and Chen and Santos (2018).

Suppose that μ is a local deviation from P. Then, the efficient GMM estimator for θ_0 may be asymptotically biased, while the J test may have nontrivial local power to the deviation. Our concern is whether the J test has nontrivial local power when the GMM estimator is asymptotically biased.

Remark 2.1

Overidentifying restrictions tests are often used to test the following null hypothesis:

$$H_0 : \mathbb{E}[m_{\theta_0}(X)] = 0 \tag{2.4}$$

when $X \sim P$ is known. A problem of testing (2.4) *has been pointed out by some studies (see Deaton 2010 and Parente and Santos Silva 2012 among others). The point is that* (2.3) *is the identifying assumption for θ_0 that must be made before the analysis. Overidentifying restrictions tests do not check whether* (2.4) *holds because they do not have power when $\mathbb{E}[m_\theta(X)] = 0$ for some $\theta \neq \theta_0$.*

3 Main Results

This section shows that the directions of local deviation that can be detected by locally unbiased specification tests are orthogonal to those that induce asymptotic bias in asymptotically efficient estimators. This means that it is impossible to know whether efficient estimators are asymptotically biased by using any locally unbiased specification tests. This section also discusses some problems with the Hausman test. We clarify what the null and alternative hypotheses of the Hausman test are and show that the Hausman test cannot be used to confirm the validity of estimators.

3.1 Orthogonality

We introduce some more definitions given by Chen and Santos (2018). The test ϕ_n has local asymptotic level α if

$$\limsup_{n\to\infty} \int \phi_n dP^n_{1/\sqrt{n},g} \leq \alpha$$

for any path $t \mapsto P_{t,g} \in \mathbf{P}$. The test has a local asymptotic power function $\pi : L^2_0(P) \to [0,1]$ if

$$\lim_{n\to\infty} \int \phi_n dP^n_{1/\sqrt{n},g} = \pi(g)$$

for any path $t \to P_{t,g} \in \mathcal{M}$. Finally, the test is locally unbiased if it satisfies $\pi(g) \leq \alpha$ for all $t \to P_{t,g} \in \mathbf{P}$ and $\pi(g) \geq \alpha$ for all $t \mapsto P_{t,g} \in \mathcal{M} \setminus \mathbf{P}$.

It is clear from above definitions that locally unbiased tests do not have nontrivial local power if $g \in \bar{T}(P)$ because the local deviation is consistent with the null hypothesis. Thus, locally unbiased tests cannot distinguish between P and $P_{1/\sqrt{n},g}$ if $g \in \bar{T}(P)$. The power of locally unbiased tests depends only on $\Pi_{T^\perp}(g)$.

Local unbiasedness of a test implies that its test statics is asymptotically composed of elements of $\bar{T}(P)^\perp$. For instance, if the test statistic is asymptotically chi-squared distributed under P, then it satisfies

$$T_n = \sum_{j=1}^{K} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} f(X_i) \right)^2 + o_P(1)$$

where $f_1, \ldots, f_K$ are orthonormal and K determines the degrees of freedom. Moreover, by the Hellinger differentiability of the path, $P_{t,g}$ satisfies

$$\log \prod_{i=1}^{n} \frac{dP_{1/\sqrt{n},g}}{dP}(X_i) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} g(X_i) - \frac{1}{2}\mathbb{E}[g^2(X)] + o_P(1)$$

for all $g \in L_0^2(P)$. Thus, by the LeCam's third lemma, we obtain

$$T_n \overset{g}{\rightsquigarrow} \chi_K^2\left(\sum_{j=1}^{K} \mathbb{E}[f_j(X)g(X)]^2\right),$$

where $\overset{g}{\rightsquigarrow}$ denotes the weak convergence under $P_{1/\sqrt{n},g}$ and $\chi_k^2(a)$ denotes the noncentral chi-squared distribution with degrees of freedom k and the noncentrality parameter a. Because the locally unbiased test has nontrivial local power if and only if $\mathrm{Var}[\Pi_{T^\perp}(g)(X)] \neq 0$, it must be the case that $f_j \in \bar{T}(P)^\perp$ for all $j = 1, \ldots, K$.

Next, we see that asymptotic bias of asymptotically efficient estimators depends only on $\Pi_T(g)$. Note that any asymptotically efficient regular estimator for θ_0 satisfies

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \nu(X_i) + o_P(1),$$

where ν is an efficient influence function. Hence, by the LeCam's third lemma, we obtain

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \overset{g}{\rightsquigarrow} N(\mathbb{E}[\nu(X)g(X)], \mathbb{E}[\nu(X)\nu(X)']).$$

This means that the asymptotic distribution of $\hat{\theta}_n$ is unaffected by the local deviation if $\nu(X)$ is uncorrelated with $g(X)$. It is known in the semiparametric estimation literature that each element of the efficient influence function belongs to $\bar{T}(P)$. Therefore, the asymptotic bias depends only on $\Pi_T(g)$. If $g \in \bar{T}(P)^\perp$, then the asymptotic distribution of $\hat{\theta}_n$ under $P_{1/\sqrt{n},g}$ is the same as that of under P.

Combining these results, we obtain the following proposition.

Proposition 3.1

Suppose that P is locally overidentified by $\mathbf{P}$. Let $\hat{\theta}_n$ be an asymptotically efficient estimator for θ_0 under P. Moreover, let ϕ_n be a locally unbiased asymptotic level α test for (2.1) whose local asymptotic power function is π. Suppose further that a random sample is drawn form $\mu = P_{1/\sqrt{n},g}$. Then, $\hat{\theta}_n$ is asymptotically biased only if $\mathrm{Var}[\Pi_T(g)(X)] \neq 0$. Moreover, ϕ_n satisfies $\pi(g) > \alpha$ only if $\mathrm{Var}[\Pi_{T^\perp}(g)(X)] \neq 0$.

Proposition 3.1 states that the directions of local deviation that can be detected by locally unbiased specification tests are orthogonal to those cause asymptotic bias in asymptotically efficient estimators. Therefore, specification tests provide no information on the validity of estimators. It is true that if both $\mathrm{Var}[\Pi_T(g)(X)] \neq 0$ and $\mathrm{Var}[\Pi_{T^\perp}(g)(X)] \neq 0$ hold, then the test has nontrivial local power and the estimator is asymptotically biased. However, it is impossible to distinguish cases between $\mathrm{Var}[\Pi_T(g)(X)] = 0$ and $\mathrm{Var}[\Pi_T(g)(X)] \neq 0$ by using any locally unbiased specification test. Thus, the detection of the bias is merely coincidental.

Example 3.1 (J test)

Suppose that a random sample is drawn from $P_{1/\sqrt{n},g}$. Then, the efficient GMM estimator for θ_0 is asymptotically biased only if $\mathrm{Var}[\Pi_T(g)(X)] \neq 0$. In contrast, the J test statistic converges to a noncentral chi-squared distribution only if $\mathrm{Var}[\Pi_{T^\perp}(g)(X)] \neq 0$. Since $\Pi_T(g)$

and $\Pi_{T^\perp}(g)$ are orthogonal to each other, the J test cannot be used to detect the asymptotic bias of the GMM estimator. This result is essentially the same as the one shown by Hall (2005). We further investigate this issue in Section 4.

The same orthogonality as in Proposition 3.1 holds even when a maintained hypothesis exists. If a maintained hypothesis exists, we consider the following null and alternative hypotheses:

$$H_0 : \mu \in \mathbf{P} \quad \text{vs.} \quad H_1 : \mu \in \mathbf{M} \setminus \mathbf{P}, \tag{3.1}$$

where $\mathbf{M}$ is another semiparametric model corresponding to the maintained hypothesis and satisfies $\mathbf{P} \subset \mathbf{M} \subset \mathcal{M}$. That is, $\mu \in \mathbf{M}$ is assumed to be true under both the null and alternative hypotheses.

We can decompose $L_0^2(P)$ using the tangent set of model $\mathbf{M}$ at P. Let

$$M(P) = \left\{g \in L_0^2(P) : (2.2) \text{ holds for some } t \mapsto P_{t,g} \in \mathbf{M}\right\}$$

and let $\bar{M}(P)$ be the closure of $M(P)$. Since $\bar{T}(P) \subset \bar{M}(P)$, we have

$$L_0^2(P) = \bar{T}(P) \oplus \left\{\bar{T}(P)^\perp \cap \bar{M}(P)\right\} \oplus \bar{M}(P)^\perp.$$

Thus, $g \in L_0^2(P)$ can be decomposed as $g = \Pi_T(g) + \Pi_{T^\perp \cap M}(g) + \Pi_{M^\perp}(g)$.

It is clear that locally unbiased tests have nontrivial local power for the deviation $P_{1/\sqrt{n},g}$ only if $\mathrm{Var}[\Pi_{T^\perp \cap M}(g)(X)] \neq 0$. In contrast, asymptotically efficient estimators for θ_0 are asymptotically biased only if $\mathrm{Var}[\Pi_T(g)(X)] \neq 0$. Thus, again, specification tests cannot detect local deviations that cause asymptotic bias in efficient estimators.

Remark 3.1
Since $\hat{\theta}_n$ is a best regular estimator, it satisfies

$$\sqrt{n}(\hat{\theta}_n - \psi(P_{1/\sqrt{n},g})) \overset{g}{\rightsquigarrow} N(0, \mathbb{E}[\nu(X)\nu(X)']) \tag{3.2}$$

for any $g \in T(P)$. The limiting distribution does not depend on g. Regularity is a desirable property because (3.2) *implies that a small change in the DGP does not change the distribution of the estimator. The estimator is not asymptotically biased if the parameter of interest is $\psi(P_{1/\sqrt{n},g})$ rather than $\psi(P)$. It follows from* (3.2) *that the asymptotic bias of $\hat{\theta}_n$ for estimating θ_0 is given by*

$$\lim_{t \to 0} \frac{\psi(P_{t,g}) - \psi(P)}{t}.$$

The above derivative coincides with $\mathbb{E}[\nu(X)g(X)]$ (see van der Vaart 1998).

3.2 Hausman test

The Hausman test statistic is constructed by comparing two estimators. Let $\hat{\theta}_n$ and $\tilde{\theta}_n$ be estimators for θ_0. Then the test statistic is given by

$$T_n = n(\hat{\theta}_n - \tilde{\theta}_n)' \hat{V}^{-1}_{\tilde{\theta}-\hat{\theta}} (\hat{\theta}_n - \tilde{\theta}_n),$$

where $\hat{V}_{\hat{\theta}-\tilde{\theta}}$ is a consistent estimator for the asymptotic variance of $\hat{\theta}_n - \tilde{\theta}_n$. If the inverse matrix does not exist, then it is replaced with the Moore-Penrose generalized inverse.

We assume that $\hat{\theta}_n$ is asymptotically efficient in model $\mathbf{P}$ and that $\tilde{\theta}_n$ is asymptotically efficient in model $\mathbf{M}$. Moreover, we assume that $\hat{\theta}_n$ is asymptotically more efficient than $\tilde{\theta}_n$, which implies $\mathbf{P} \subset \mathbf{M}$. Let $T(P)$ and $M(P)$ be the tangent set at P for $\mathbf{P}$ and $\mathbf{M}$, respectively. Since $\hat{\theta}_n$ and $\tilde{\theta}_n$ are asymptotically efficient under corresponding models, they can be written as

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \nu(X_i) + o_P(1)$$
$$\sqrt{n}(\tilde{\theta}_n - \theta_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \tau(X_i) + o_P(1)$$

for some $\nu \in \bar{T}(P)$ and $\tau \in \bar{M}(P)$, where $\nu \in \bar{T}(P)$ means that all elements of ν belong to $\bar{T}(P)$ and the same for $\tau \in \bar{M}(P)$.

Under the above conditions, the Hausman test is a locally unbiased test for (3.1). The reason is as follows. It is known in the semiparametric literature that the efficient influence function is obtained by projecting any other influence function onto the tangent space (see van der Vaart 1998). Since $\hat{\theta}_n$ is more efficient than $\tilde{\theta}_n$, we have $\nu = \Pi_T(\tau)$ and $\tau - \nu \in \bar{T}(P)^\perp \cap \bar{M}(P)$. This implies that the well-known fact that the asymptotic variance of $\hat{\theta}_n - \tilde{\theta}_n$ is the same as the difference of the asymptotic variances of $\tilde{\theta}_n$ and $\hat{\theta}_n$ when $\hat{\theta}_n$ is asymptocially efficient. Moreover, the test statistic can be written as

$$T_n = \sum_{j=1}^{K} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} f_j(X_i) \right)^2 + o_P(1)$$

for some K, where $f_1, \dots, f_K \in \bar{T}(P)^\perp \cap \bar{M}(P)$ are orthonormal. Under $P_{1/\sqrt{n}}$, T_n converges weakly to $\chi^2_K(\sum_{j=1}^{K} \mathbb{E}[f_j(X)g(X)]^2)$. Therefore, the test has nontrivial local power only if $\mathrm{Var}[\Pi_{T^\perp \cap M}(g)(X)] \neq 0$.

Next, we investigate the asymptotic properties of two estimators. By the LeCam's third lemma, we have

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \overset{g}{\rightsquigarrow} N(\mathbb{E}[\nu(X)g(X)], \mathbb{E}[\nu(X)\nu(X)'])$$
$$\sqrt{n}(\tilde{\theta}_n - \theta_0) \overset{g}{\rightsquigarrow} N(\mathbb{E}[\tau(X)g(X)], \mathbb{E}[\tau(X)\tau(X)']).$$

Since $\nu \in \bar{T}(P)$, the asymptotic bias of $\hat{\theta}_n$ depends only on $\Pi_T(g)$. Thus, the Hausman test cannot detect asymptotic bias in $\hat{\theta}_n$. Furthermore, if $g \in \bar{T}(P)^\perp \cap \bar{M}(P)$, then only $\tilde{\theta}_n$ can be asymptotically biased. Thus, the test may reject the null hypothesis due to the bias of the inefficient estimator rather than that of the efficient estimator. The Hasuman test has a risk of selecting the biased inefficient estimator even when the efficient estimator is asymptotically unbiased.

The above result is incompatible with a common setting of the Hausman test, which assumes that the inefficient estimator is consistent under both null and alternative hypotheses.

In fact, this common setting is inappropriate. Because the DGP differs between null and alternative hypotheses, no estimator can be assumed to have the same asymptotic distribution under two different hypotheses. What we can assume is that $\mu \in \mathbf{M}$ is true under both null and alternative hypotheses.

Another important feature of the Hausman test is that the estimators that construct the test statistic determine the null and alternative (maintained) hypotheses. For example, consider a linear model

$$Y = X'\beta_0 + e = X_1'\beta_{01} + X_2'\beta_{02} + e,$$

where X_1 is possibly endogenous and X_2 is exogenous. Let Z_1 be a vector of instrumental variables for X_1 and let $Z = (Z_1', X_2')'$. Suppose that we test the exogeneity of X_1 by comparing the OLS and 2SLS estimators. Since the 2SLS estimator is asymptotically efficient in homoskedastic linear instrumental variable model, the maintained hypothesis is that

$$\mathbb{E}_\mu[Z(Y - X'\beta)] = 0 \quad \text{and} \quad \mathbb{E}_\mu[(Y - X'\beta)^2|Z] = \sigma^2 \tag{3.3}$$

hold for some (β, σ^2), where $\mathbb{E}_\mu[\cdot]$ and $\mathbb{E}_\mu[\cdot|\cdot]$ denote the unconditional and conditional expectations with respect to μ. On the other hand, since the OLS estimator is asymptotically efficient in homoskedastic linear regression model, the null hypothesis is that

$$\mathbb{E}_\mu[Y - X'\beta|X_1, Z] = 0 \quad \mathbb{E}_\mu[(Y - X'\beta)^2|X_1, Z] = \sigma^2 \tag{3.4}$$

hold for (β, σ^2) that satisfies (3.3). Notice that the conditioning variables in (3.4) are (X_1, Z) rather than X. We need this condition so that Z satisfies an exclusion restriction. If we only impose $\mathbb{E}[e|X] = 0$ and $\mathbb{E}[e^2|X] = \sigma_0^2$ (constant), then the OLS estimator for β_0 may not be asymptotically efficient because the use of Z_1 may improve efficiency.

4 J test

This section shows that the orthogonality result of Section 3 holds for the overidentified moment restriction model.

We first rewrite the model by using the method of Sueishi (2022). This formulation is useful for obtaining the tangent set in a conventional way in the semiparametric literature (see, e.g., Section 25.4 of van der Vaart 1998). Since (2.3) involves two unknown parameters θ_0 and P, we write the model as a set of distributions indexed by the finite-dimensional parameter $\theta \in \Theta$ and the infinite-dimensional nuisance parameter $\eta \in \mathcal{M}$. Specifically, for given $\theta \in \Theta$ and $\eta \in \mathcal{M}$, we define $P_{\theta,\eta}$ as the solution to

$$\min_{Q \in \mathbf{P}_\theta} \int \log \frac{dQ}{d\eta} dQ, \tag{4.1}$$

where $\mathbf{P}_\theta = \{Q \in \mathcal{M} : \int m_\theta dQ = 0\}$. That is, $P_{\theta,\eta}$ is the projection of η onto $\mathbf{P}_\theta$ in terms of the Kullback–Leibler divergence (I- divergence). By a duality theorem, $P_{\theta,\eta}$ satisfies

$$\frac{dP_{\theta,\eta}}{d\eta} = \frac{\exp(\lambda_{\theta,\eta}' m_\theta)}{\int \exp(\lambda_{\theta,\eta}' m_\theta) d\eta}$$

where $\lambda_{\theta,\eta} = \arg\min_{\lambda \in \mathbb{R}^l} \int \exp(\lambda' m_\theta) d\eta$. See Borwein and Lewis (1991) and Komunjer and Ragusa (2016) for details.

We specify the tangent set of $\mathbf{P} = \{P_{\theta,\eta} : \theta \in \Theta, \eta \in \mathcal{M}\}$ at P. To do this, we consider a path of the form $P_t = P_{\theta_0 + th, \eta_t}$, where $h \in \mathbb{R}^p$ is a p-dimensional vector and η_t is a perturbation from P that coincides with P at $t = 0$. Under certain conditions, P_t satisifies

$$\lim_{t \to 0} \int \left(\frac{dP_t^{1/2} - dP^{1/2}}{t} - \frac{1}{2}(h' \dot{\ell}_{\theta_0,\eta_0} + \dot{l}) dP^{1/2} \right)^2 = 0$$

where

$$\dot{\ell}_{\theta_0,\eta_0} = -\mathbb{E}[\nabla m_{\theta_0}(X)]' \Sigma^{-1} m_{\theta_0}$$

with $\nabla m_\theta = \partial m_\theta / \partial \theta'$ and $\Sigma = \mathbb{E}[m_{\theta_0}(X) m_{\theta_0}(X)']$. Moreover, $\dot{l} : \mathcal{X} \to \mathbb{R}$ is an element of the set $\dot{\mathbf{P}}_\eta = \{\dot{l} \in L_0^2(P) : \mathbb{E}[m_{\theta_0}(X)\dot{l}(X)] = 0\}$. See Sueishi (2022) for details. The function $\dot{\ell}_{\theta_0,\eta_0}$ is interpreted as the score function for θ_0 when η_0 is fixed while $\dot{l}$ is interpreted as the score function for η_0 when θ_0 is fixed. The tangent set is given by $T(P) = \{\text{lin } \dot{\ell}_{\theta_0,\eta_0} + \dot{\mathbf{P}}_\eta\}$, where lin denotes the linear span.

Notice that $\dot{\ell}_{\theta_0,\eta_0}$ is orthogonal to the all elements of $\dot{\mathbf{P}}_\eta$. Thus, $\dot{\ell}_{\theta_0,\eta_0}$ is the efficient score function for estimating θ_0. Because the efficient information matrix is given by

$$I_{\theta_0,\eta_0} = \mathbb{E}[\dot{\ell}_{\theta_0,\eta_0}(X) \dot{\ell}_{\theta_0,\eta_0}(X)'] = \mathbb{E}\left[\nabla m_{\theta_0}(X)\right]' \Sigma^{-1} \mathbb{E}\left[\nabla m_{\theta_0}(X)\right],$$

the efficient influence function is $I_{\theta_0,\eta_0}^{-1} \dot{\ell}_{\theta_0,\eta_0}$.

Now, we investigate the local asymptotic property of the GMM estimator. The efficient GMM estimator can be expressed as

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^n I_{\theta_0,\eta_0}^{-1} \dot{\ell}_{\theta_0,\eta_0}(X_i) + o_P(1).$$

Thus, it follows from the LeCam's third lemma that

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \overset{g}{\rightsquigarrow} N(I_{\theta_0,\eta_0}^{-1} \mathbb{E}[\dot{\ell}_{\theta_0,\eta_0}(X)(\Pi_T(g)(X) + \Pi_{T^\perp}(g)(X))], I_{\theta_0,\eta_0}^{-1})$$

for any $g \in L_0^2(P)$. Here, the efficient influence function clearly belongs to $\bar{T}(P)$ and is therefore orthogonal to $\Pi_{T^\perp}(g)$. Thus, the asymptotic bias depends only on $\Pi_T(g)$.

The GMM estimator can be asymptotically unbiased even under local deviation because it utilizes only a part of moment restrictions. The expected value of $m_{\theta_0}(X)$ under $P_{1/\sqrt{n},g}$ is approximately given by $\mathbb{E}[m_{\theta_0}(X)g(X)]/\sqrt{n}$, and the GMM estimator is asymptotically unbiased if

$$\mathbb{E}[\nabla m_{\theta_0}(X)]' \Sigma^{-1} \mathbb{E}[m_{\theta_0}(X) g(X)] = 0.$$

Although $\mathbb{E}[m_{\theta_0}(X)g(X)] \neq 0$, the left-hand side of the above equation can be 0 because the rank of $E[\nabla m_{\theta_0}(X)]'\Sigma^{-1}$ is p. Thus, the GMM estimator can be asymptotically unbiased even when moment restrictions are locally violated.

Next, we investigate the local asymptotic property of the J test statistic. The test statistic satisfies

$$J_n = \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n m_{\hat{\theta}_n}(X_i) \right)' \Sigma^{-1} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n m_{\hat{\theta}_n}(X_i) \right) + o_P(1).$$

Some calculation yields

$$\Sigma^{-1/2}\frac{1}{\sqrt{n}}\sum_{i=1}^{n} m_{\hat{\theta}_n}(X_i) \overset{g}{\rightsquigarrow} N((I-P(\theta_0))\Sigma^{-1/2}\mathbb{E}[m_{\theta_0}(X)g(X)], I-P(\theta_0)),$$

where $P(\theta_0) = \Sigma^{-1/2}\mathbb{E}[\nabla m_{\theta_0}(X)]\left(\mathbb{E}[\nabla m_{\theta_0}(X)]'\Sigma^{-1}\mathbb{E}[\nabla m_{\theta_0}(X)]\right)^{-1}\mathbb{E}[\nabla m_{\theta_0}(X)]'\Sigma^{-1/2}$ is a projection matrix. Moreover, for any g, $\Pi_T(g)$ can be written as $\Pi_T(g) = h'\dot{\ell}_{\theta_0,\eta_0} + \dot{l}$ for some $h \in \mathbb{R}^p$ and $\dot{l} \in \dot{\mathbf{P}}_\eta$. Thus, we have

$$(I-P(\theta_0))\Sigma^{-1/2}\mathbb{E}[m_{\theta_0}(X)\Pi_T(g)(X)] = (I-P(\theta_0))\Sigma^{-1/2}\mathbb{E}[\nabla m_{\theta_0}(X)]h = 0, \qquad (4.2)$$

which implies that J_n converges in distribution to the noncentral chi-square distribution with degrees of freedom $l-p$ and noncentrality parameter

$$\mathbb{E}[m_{\theta_0}(X)\Pi_{T^\perp}(g)(X)]'\Sigma^{-1/2}(I-P(\theta_0))\Sigma^{-1/2}\mathbb{E}[m_{\theta_0}(X)\Pi_{T^\perp}(g)(X)].$$

Therefore, the power of the J test depends only on $\Pi_{T^\perp}(g)$. The J test is locally unbiased for testing $\mu \in \mathbf{P}$ against $\mu \in \mathcal{M} \setminus \mathbf{P}$.

The result of this section gives another interpretation to the result of Hall (2005), who considered a local deviation of the form $\Sigma^{-1/2}\mathbb{E}_n[m_{\theta_0}(X)] = \delta/\sqrt{n}$ for some $\delta \in \mathbb{R}^l$, where $\mathbb{E}_n$ is the expectation with respect to the local deviation. The vector $\Sigma^{-1/2}\mathbb{E}_n[m_{\theta_0}(X)]$ can be decomposed as

$$P(\theta_0)\Sigma^{-1/2}\mathbb{E}_n[m_{\theta_0}(X)] + (I-P(\theta_0))\Sigma^{-1/2}\mathbb{E}_n[m_{\theta_0}(X)]. \qquad (4.3)$$

Since $P(\theta_0)$ is the projection matrix, the two terms are orthogonal to each other. Hall (2005) defines that the identifying restrictions are satisfied if the first term of (4.3) is 0 and the overidentifying restrictions are satisfied if the second term is 0. He showed that the GMM estimator is asymptotically biased only if the identifying restrictions are locally violated, which is the case where the projection of δ onto the space spanned by $\Sigma^{-1/2}\mathbb{E}[m_{\theta_0}(X)]$ is nonzero. In contrast, the J test has nontrivial local power only if the overidentifying restrictions are locally violated. If the expectaion is taken with respect to $P_{1/\sqrt{n},g}$, then $\mathbb{E}_n[m_{\theta_0}(X)]$ is nearly equal to $\mathbb{E}[m_{\theta_0}(X)g(X)]/\sqrt{n}$. Thus, it follows from (4.2) that the first term of (4.3) can be nonzero only if $\mathrm{Var}[\Pi_T(g)(X)] \neq 0$ while the second term can be nonzero only if $\mathrm{Var}[\Pi_{T^\perp}(g)(X)] \neq 0$. Therefore, the decomposition of the score function produces the same asymptotic result with Hall (2005).

5 Durbin-Wu-Hausman test

This section further investigates the Durbin-Wu-Hausman test for exogeneity. The test statistic is given by

$$T_n = (\hat{\beta}_{ols} - \tilde{\beta}_{2sls})'\hat{V}^{-}(\hat{\beta}_{ols} - \tilde{\beta}_{2sls})$$

where $\hat{\beta}_{ols}$ and $\tilde{\beta}_{2lsls}$ denote the OLS and 2SLS estimators, respectively. Also, $\hat{V}^{-}$ denote the generalized inverse of a consistent estimators for the asymptotic variance of $\hat{\beta}_{ols} - \tilde{\beta}_{2sls}$. Then, by the result of Section 3.2, the implied null and alternative hypotheses are

$$H_0 : \mu \in \mathbf{P} \quad \text{vs.} \quad H_1 : \mu \in \mathbf{M} \setminus \mathbf{P}$$

where

$$\mathbf{P} = \left\{ Q \in \mathcal{M} : \mathbb{E}_Q[Y - X'\beta | X_1, Z] = 0 \text{ and } \mathbb{E}_Q[(Y - X'\beta)^2 | X_1, Z] = \sigma^2 \text{ for some } \beta \text{ and } \sigma^2 \right\}$$
$$\mathbf{M} = \left\{ Q \in \mathcal{M} : \mathbb{E}_Q[Z(Y - X'\beta)] = 0 \text{ and } \mathbb{E}_Q[(Y - X'\beta)^2 | Z] = \sigma^2 \text{ for some } \beta \text{ and } \sigma^2 \right\}.$$

The parameter of interest $\theta_0 = (\beta_0', \sigma_0^2)'$ satisfies

$$\mathbb{E}[Y - X'\beta_0 | X_1, Z] = 0 \quad \text{and} \quad \mathbb{E}[(Y - X'\beta_0)^2 | X_1, Z] = \sigma_0^2.$$

We specify the tangent set of $\mathbf{P}$ and $\mathbf{M}$ at P. The tangent set of $\mathbf{M}$ can be obtained by using the result of Section 4 with $m_\theta(X) = Z(Y - X'\beta)$. Let $e = Y - X'\beta_0$. Then, $\Sigma = \mathbb{E}[ZZ'e^2] = \sigma_0 \mathbb{E}[ZZ']$ by the homoskedasticity. Hence we have $M(P) = \{\text{lin } \dot{\ell}^M_{\theta_0,\eta_0} + \dot{\mathbf{M}}_\eta\}$, where $\dot{\ell}^M_{\theta_0,\eta_0}(x_1, y, z) = \mathbb{E}[XZ']\mathbb{E}[ZZ']^{-1} ze/\sigma_0^2$ and

$$\dot{\mathbf{M}}_\eta = \left\{ \dot{l}^M \in L_0^2(P) : \mathbb{E}[Ze\dot{l}^M(X_1, Y, Z)] = 0 \right\}.$$

To obtain the tangent set of $\mathbf{P}$, we introduce a new formulation of the model similar to the one used in Section 4. The details are given in the Appendix. We can show that $T(P) = \{\text{lin } \dot{\ell}^P_{\theta_0,\eta_0} + \dot{\mathbf{P}}_\eta\}$, where $\dot{\ell}^P_{\theta_0,\eta_0}(x_1, y, z) = xe/\sigma_0^2$ and

$$\dot{\mathbf{P}}_\eta = \left\{ \dot{l}^P \in L_0^2(P) : \mathbb{E}[h(X_1, Z)e\dot{l}^P(X_1, Z, Y)] = 0 \text{ for any function } h \text{ of } (X_1, Z) \right\}.$$

Now, we investigate the asymptotic properties of the two estimators. By the LeCam's third lemma, we have

$$\sqrt{n}(\hat{\beta}_{ols} - \beta_0) \overset{g}{\rightsquigarrow} N\left(\mathbb{E}[XX']^{-1}\mathbb{E}[Xeg(X_1, Y, Z)], \sigma_0^2 E[XX']^{-1} \right)$$
$$\sqrt{n}(\tilde{\beta}_{2lsl} - \beta_0) \overset{g}{\rightsquigarrow} N\left(\mathbb{E}[XZ']\mathbb{E}[ZZ']^{-1}\mathbb{E}[Zeg(X_1, Y, Z)], \sigma_0^2 \left(\mathbb{E}[XZ']\mathbb{E}[ZZ']^{-1}E[ZX'] \right)^{-1} \right).$$

If $g \in \bar{T}(P)$, then g satisfies $g = h'\dot{\ell}^P_{\theta_0,\eta_0} + \dot{l}^P$ for some vector $h \in \mathbb{R}^{\dim(\beta)}$ and $\dot{l}^P \in \dot{\mathbf{P}}_\eta$. Thus, we obtain

$$\mathbb{E}[XX']^{-1}\mathbb{E}[Xeg(X_1, Y, Z)] = \mathbb{E}[XX']^{-1}\mathbb{E}[XX'e^2]h/\sigma_0^2 = h.$$

Similarly, we have $\mathbb{E}[XZ']\mathbb{E}[ZZ']^{-1}\mathbb{E}[Zeg(X_1, Y, Z)] = h$. That is, the OLS and 2SLS estimators have the same asymptotic bias. This result implies that the test does not have nontrivial local power if $g \in \bar{T}(P)$.

If $g \in \bar{T}(P)^\perp \cap \bar{M}(P)$, then g must be orthogonal to $\dot{\ell}^P_{\theta_0,\eta_0}$. Thus, the OLS estimator is asymptotically unbiased. Moreover, g can be written as $g = h'\dot{\ell}^M_{\theta_0,\eta_0} + \dot{l}^M$ for some $h \in \mathbb{R}^{\dim(\beta)}$ and $\dot{l}^M \in \dot{\mathbf{M}}_\eta$. Thus, the 2SLS estimator is asymptotically biased with bias h. Because two estimators have different asymptotic mean, the test have nontrivial local power. Thus, the test is locally unbiased for testing $\mu \in \mathbf{P}$ against $\mu \in \mathbf{M} \setminus \mathbf{P}$. Note that the test may reject the null hypothesis even when the OLS estimator is asymptotically unbiased and the 2SLS estimator is asymptotically biased.

In general, g can be decomposed as $g = \Pi_T(g) + \Pi_{T^\perp \cap M}(g) + \Pi_{M^\perp}(g)$. Because $\Pi_{M^\perp}(g)$ is orthogonal to $\dot{\ell}^P_{\theta_0,\eta_0}$ and $\dot{\ell}^M_{\theta_0,\eta_0}$, it does not have any impact on the asymptotic bias of $\hat{\beta}_{ols}$ and $\tilde{\beta}_{2sls}$. This also implies that the power of the test does not depend on $\Pi_{M^\perp}(g)$. The asymptotic bias of $\hat{\beta}_{ols}$ depends only on $\Pi_T(g)$ whereas the asymptotic bias of $\tilde{\beta}_{2sls}$ depends on $\Pi_T(g)$ and $\Pi_{T^\perp \cap M}(g)$. There is no clear order of magnitude between the bias of $\hat{\beta}_{ols}$ and $\tilde{\beta}_{2sls}$. Since $\Pi_T(g)$ induces the same amount of bias in $\hat{\beta}_{ols}$ and $\tilde{\beta}_{2sls}$, the power of the test depends only on $\Pi_{T^\perp \cap M}(g)$.

Finally, we consider how the OLS estimator can be asymptotically unbiased when $g \in \bar{T}(P)^\perp$. Because the null model is misspecified in this case, there is no $\theta = (\beta', \sigma^2)'$ that satisfies $\mathbb{E}_\mu[Y - X'\beta | X_1, Z] = 0$ and $\mathbb{E}_\mu[(Y - X'\beta)^2 | X_1, Z] = \sigma^2$. However, the case $\mathbb{E}_\mu[X(Y - X'\beta_0)] = 0$ is not excluded, so that the OLS can be asymptotically unbiased. In this case, X_1 is actually exogenous under μ even though the null hypothesis is not true.

6 Conclusion

This paper studies the local asymptotic properties specification tests and asymptotically efficient estimators. Although many studies have examined the properties of specification tests and estimators separately, there have been few studies that examine the connection between specification tests and estimators. We show that the directions of local deviation that can be detected by locally unbiased specification test are orthogonal to those that cause asymptotic bias in asymptotically efficient estimators. This means that locally unbiased specification tests cannot detect bias in asymptotically efficient estimators. Although often used to check the validity of estimators, it is a misuse to use specification tests to examine the validity of estimators.

A Appendix

This appendix derives the tangent set when the model is specified by conditional moment restrictions. The derivation is similar to the case of the unconditional moment restriction model.

Let P be a joint probability distribution of (X, W) whose support is $\mathcal{X} \times \mathcal{W}$, and let $\mathcal{M}$ be the set of all distributions on $\mathcal{X} \times \mathcal{W}$. We write $P = P_{W|X} P_X$, where $P_{W|X}$ is the conditional distribution of W given X and P_X is the marginal distribution of X. Suppose that the parameter of interest $\theta_0 \in \Theta \subset \mathbb{R}^p$ is a unique vector that satisfies

$$0 = \mathbb{E}[m_{\theta_0}(W)|X] = \int m_{\theta_0} dP_{W|X} \ a.s.\ P_X,$$

where $m : \mathcal{W} \times \Theta \to \mathbb{R}^l$ is a known vector function.

We write the model as a set of distribution on $\mathcal{X} \times \mathcal{W}$ that is indexed by the finite-dimensional parameter $\theta \in \Theta$ and the infinite-dimensional nuisance parameter $\eta \in \mathcal{M}$. Let

$$\mathbf{P}_\theta = \left\{ Q = Q_{W|X} Q_X \in \mathcal{M} : \int m_\theta dQ_{W|X} = 0 \ a.s.\ Q_X \right\},$$

which is a set of distribution that satisfies the conditional moment restrictions for a given value of θ. For given $\theta \in \Theta$ and $\eta = \eta_{W|X}\eta_X \in \mathcal{M}$, we define $P_{\theta,\eta}$ as the solution to

$$\min_{Q \in \mathbf{P}_\theta} \int\int \log \frac{dQ}{d\eta} dQ. \tag{A.1}$$

The model is then written as $\mathbf{P} = \{P_{\theta,\eta} : \theta \in \Theta, \eta \in \mathcal{M}\}$.

The objective function of (A.1) can be written as

$$\int\int \log \frac{dQ_{W|X}}{d\eta_{W|X}} dQ_{W|X} dQ_X + \int \log \frac{dQ_X}{d\eta_X} dQ_X.$$

The second term is minimized when $Q_X = \eta_X$. The first term is minimized by solving

$$\min_{Q_{W|X}(\cdot|x)} \int \log \frac{dQ_{W|X}(w|x)}{d\eta_{W|X}(w|x)} dQ_{W|X}(w|x)$$

subject to

$$\int m_\theta(w) dQ_{W|X}(w|x) = 0$$

for each $x \in \mathcal{X}$. By a duality theorem, the solution $Q^*_{\theta,W|X}$ satisfies

$$\frac{dQ^*_{\theta,W|X}}{d\eta_{W|X}}(w|x) = \frac{\exp(\lambda_{\theta,\eta_{W|X}}(x)' m_\theta(w))}{\int \exp(\lambda_{\theta,\eta_{W|X}}(x)' m_\theta(w)) d\eta_{W|X}(w|x)}$$

where

$$\lambda_{\theta,\eta_{W|X}}(x) = \arg\min_{\lambda \in \mathbb{R}^l} \int \exp(\lambda' m_\theta(w)) d\eta_{W|X}(w|x).$$

See, for instance, Komunjer and Ragusa (2016) for a rigorous argument. Thus, we also have

$$\frac{dP_{\theta,\eta}}{d\eta}(x,w) = \frac{\exp(\lambda_{\theta,\eta_{W|X}}(x)' m_\theta(w))}{\int \exp(\lambda_{\theta,\eta_{W|X}}(x)' m_\theta(w)) d\eta_{W|X}(w|x)}.$$

Let $\lambda_\theta(x) = \lambda_{\theta,P_{W|X}}(x)$. Then, $\lambda_{\theta_0}(X) = 0$ a.s. P_X. Hence we have

$$\frac{\partial}{\partial\theta} \log dP_{\theta,\eta_0}(x,w)\bigg|_{\theta=\theta_0} = \left[\frac{\partial \lambda_\theta(x)}{\partial \theta'}\bigg|_{\theta=\theta_0} \right]' m_{\theta_0}(w).$$

Moreover, by the implicit function theorem, we obtain

$$\frac{\partial \lambda_\theta(x)}{\partial \theta'}\bigg|_{\theta=\theta_0} = -\mathbb{E}\left[m_{\theta_0}(W) m_{\theta_0}(W)' | X = x\right]^{-1} \mathbb{E}[\nabla m_{\theta_0}(W) | X = x].$$

Thus, we have

$$\frac{\partial}{\partial\theta} \log dP_{\theta,\eta_0}(x,w)\bigg|_{\theta=\theta_0} = -\mathbb{E}[\nabla m_{\theta_0}(W)|X = x]' \mathbb{E}[m_{\theta_0}(W) m_{\theta_0}(W)' | X = x]^{-1} m_\theta(w).$$

We consider a path of the form

$$P_t = P_{\theta_0 + ht, \eta_t}$$

for $t \in [0, \epsilon)$, where $h \in \mathbb{R}^p$ and η_t is a perturbation of P that satisfies $\eta_0 = P$. By the law iterated expectations, P_{θ_0,η_t} must satisfy

$$\int\int m_{\theta_0}(w) h(x) dP_{\theta_0,\eta_t}(x,w) = 0$$

for any t and for any function $h : \mathcal{X} \to \mathbb{R}$. So, we obtain

$$\mathbb{E}\left[m_{\theta_0}(W)h(X)\ \frac{\partial}{\partial t}\log dP_{\theta_0,\eta_t}\bigg|_{t=0}\right] = 0.$$

Thus, under certain conditions, the path satisfies

$$\lim_{t\to 0}\int\int\left(\frac{dP_t^{1/2} - dP^{1/2}}{t} - \frac{1}{2}(h'\dot{\ell}_{\theta_0,\eta_0} + \dot{l})dP^{1/2}\right)^2 = 0$$

where $\dot{\ell}_{\theta_0,\eta_0} = \frac{\partial}{\partial\theta}\log dP_{\theta,\eta_0}\big|_{\theta=\theta_0}$ and $\dot{l}$ is an element of the set

$$\dot{\mathbf{P}}_\eta = \left\{\dot{l} \in L_0^2(P) : \mathbb{E}[m_{\theta_0}(W)h(X)\dot{l}(X,W)] = 0 \text{ for any function } h \text{ of } X\right\}.$$

The tangent set of $\mathbf{P}$ at P is $T(P) = \{\text{lin }\dot{\ell}_{\theta_0,\eta_0} + \dot{\mathbf{P}}_\eta\}$.

References

Andrews, D. W. and P. Guggenberger (2009a). Hybrid and size-corrected subsampling methods. *Econometrica 77*(3), 721–762.

Andrews, D. W. and P. Guggenberger (2009b). Incorrect asymptotic size of subsampling procedures based on post-consistent model selection estimators. *Journal of Econometrics 152*(1), 19–27.

Andrews, I., M. Gentzkow, and J. M. Shapiro (2017). Measuring the sensitivity of parameter estimates to estimation moments. *The Quarterly Journal of Economics 132*(4), 1553–1592.

Armstrong, T. B. and M. Kolesár (2021). Sensitivity analysis using approximate moment condition models. *Quantitative Economics 12*(1), 77–108.

Bonhomme, S. and M. Weidner (2022). Minimizing sensitivity to model misspecification. *Quantitative Economics 13*(3), 907–954.

Borwein, J. M. and A. S. Lewis (1991). Duality relationships for entropy-like minimization problems. *SIAM Journal on Control and Optimization 29*(2), 325–338.

Chen, X., H. Hong, and M. Shum (2007). Nonparametric likelihood ratio model selection tests between parametric likelihood and moment condition models. *Journal of Econometrics 141*(1), 109–140.

Chen, X. and A. Santos (2018). Overidentification in regular models. *Econometrica 86*(5), 1771–1817.

Deaton, A. (2010). Instruments, randomization, and learning about development. *Journal of economic literature 48*(2), 424–55.

Doko Tchatoka, F. and W. Wang (2021). Size-corrected bootstrap test after pretesting for exogeneity with heteroskedastic or clustered data. SSRN Working Paper.

Guggenberger, P. (2010a). The impact of a hausman pretest on the asymptotic size of a hypothesis test. *Econometric Theory 26*(2), 369–382.

Guggenberger, P. (2010b). The impact of a hausman pretest on the size of a hypothesis test: The panel data case. *Journal of Econometrics 156*(2), 337–343.

Guggenberger, P. and G. Kumar (2012). On the size distortion of tests after an overidentifying restrictions pretest. *Journal of Applied Econometrics 27*(7), 1138–1160.

Hall, A. R. (2005). *Generalized Method of Moments*. Oxford University Press.

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica: Journal of the econometric society*, 1029–1054.

Hausman, J. A. (1978). Specification tests in econometrics. *Econometrica 46*, 1251–1271.

Judge, G. G. and M. E. Bock (1978). *The Statistical Implications of Pre-test and Stein-Rule Estimators in Econometrics*. North-Holland.

Kabaila, P. (1995). The effect of model selection on confidence regions and prediction regions. *Econometric Theory 11*(3), 537–549.

Kitamura, Y., T. Otsu, and K. Evdokimov (2013). Robustness, infinitesimal neighborhoods, and moment restrictions. *Econometrica 81*(3), 1185–1201.

Komunjer, I. and G. Ragusa (2016). Existence and characterization of conditional density projections. *Econometric Theory 32*(4), 947–987.

Leeb, H. and B. M. Pötscher (2005). Model selection and inference: Facts and fiction. *Econometric Theory 21*(1), 21–59.

Leeb, H. and B. M. Pötscher (2006). Can one estimate the conditional distribution of post-model-selection estimators? *The Annals of Statistics 34*(5), 2554–2591.

Newey, W. K. (1985a). Generalized method of moments specification testing. *Journal of econometrics 29*(3), 229–256.

Newey, W. K. (1985b). Maximum likelihood specification testing and conditional moment tests. *Econometrica*, 1047–1070.

Parente, P. M. and J. Santos Silva (2012). A cautionary note on tests of overidentifying restrictions. *Economics Letters 115*(2), 314–317.

Pötscher, B. M. (1991). Effects of model selection on inference. *Econometric Theory 7*(2), 163–185.

Sargan, J. D. (1958). The estimation of economic relationships using instrumental variables. *Econometrica 26*, 393–415.

Sueishi, N. (2022). Large sample justifications for the Bayesian empirical likelihood. *Econometric Theory* (forthcoming).

van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge University Press.