OVERIDENTIFICATION IN REGULAR MODELS

By

Xiaohong Chen and Andres Santos

April 2015
Revised June 2018

COWLES FOUNDATION DISCUSSION PAPER NO. 1999R

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

Overidentification in Regular Models*

Xiaohong Chen† Yale University xiaohong.chen@yale.edu

Andres Santos‡ UCLA andres@econ.ucla.edu

Initial Draft: October 2014; 1st Revision: July 2017; Final Revision: June 2018

Abstract

In the unconditional moment restriction model of Hansen (1982), specification tests and more efficient estimators are both available whenever the number of moment restrictions exceeds the number of parameters of interest. We show a similar relationship between potential refutability of a model and existence of more efficient estimators is present in much broader settings. Specifically, a condition we name *local overidentification* is shown to be equivalent to both the existence of specification tests with nontrivial local power and the existence of more efficient estimators of some "smooth" parameters in general semi/nonparametric models. Under our notion of local overidentification, various locally nontrivial specification tests such as Hausman tests, incremental Sargan tests (or optimally weighted quasi likelihood ratio tests) naturally extend to general semi/nonparametric settings. We further obtain simple characterizations of local overidentification for general models of nonparametric conditional moment restrictions with possibly different conditioning sets. The results are applied to determining when semi/nonparametric models with endogeneity are locally testable, and when nonparametric plug-in and semiparametric two-step GMM estimators are semiparametrically efficient. Examples of empirically relevant semi/nonparametric structural models are presented.

Keywords: Overidentification, semiparametric efficiency, specification testing, nonparametric conditional moment restrictions, semiparametric two step.

*We thank Lars Peter Hansen and four anonymous referees for very constructive and detailed suggestions that greatly improved the paper. We also thank Manuel Arellano, Stephane Bonhomme, Phil Haile, Hide Ichimura, Yuichi Kitamura, Pat Kline, Whitney Newey, Peter C Phillips, James Powell, Richard J Smith, and other participants at the August 1-2, 2015 International Conference on "Frontiers of Theoretical Econometrics" in celebration of Donald Andrews' 60th birthday, Konstanz, and at various econometrics seminars for helpful discussions. Any errors are the responsibility of the authors.

†Research supported by Cowles Foundation for Research in Economics.

‡Research supported by NSF Grant SES-1426882.

1 Introduction

In work originating with Anderson and Rubin (1949) and Sargan (1958), and culminating in Hansen (1982), overidentification in the generalized method of moments (GMM) framework was equated with the number of unconditional moment restrictions exceeding the number of parameters of interest. Under mild regularity conditions, such a surplus of moment restrictions was shown to enable the construction of both more efficient estimators and model specification tests. It is hard to overstate the importance of this result, which has granted practitioners with an intuitive condition characterizing the existence of both efficiency gains and specification tests, and has thus intrinsically linked both phenomena to the notion of overidentification.

Unfortunately, the existence of an analogous simple condition in general semi/non-parametric models is to the best of our knowledge unknown. Yet, such a result stands to be particularly valuable for these more flexible models, as their richer structure renders their potential refutability harder to evaluate while simultaneously generating a broader set of parameters for which efficiency considerations are of concern.

In this paper we show that, just as in GMM, efficiency and testability considerations are linked by a single condition we name *local overidentification.* In order to be applicable to general semi/non-parametric models, however, we must abstract from "counting" parameters and moment restrictions as in GMM when defining local overidentification. Instead we employ the *tangent set* $T(P)$ which, given a (data) distribution P and a candidate model $\mathbf{P}$, consists of the set of scores corresponding to all parametric submodels of $\mathbf{P}$ that contain P (Bickel et al., 1993). Heuristically, $T(P)$ consists of all the paths from which P may be approached from within $\mathbf{P}$. In particular, whenever the *closure* of $T(P)$ in the mean squared norm equals the set of all possible scores, the model $\mathbf{P}$ is locally consistent with any parametric specification and hence we say P is *locally just identified* by $\mathbf{P}$. In contrast, whenever there exist scores that do not belong to the *closure* of $T(P)$, the model $\mathbf{P}$ is locally inconsistent with some parametric specification and hence we say P is *locally overidentified* by $\mathbf{P}$. While these definitions can be generally applied, we mainly focus on models that are regular – in the sense that $T(P)$ is linear – due to the importance of this condition in semiparametric efficiency analysis (van der Vaart, 1989).[1] When specialized to GMM, our notion of local overidentification is equivalent to the standard condition that the number of unconditional moment restrictions exceed the number of parameters of interest.

Our definition of local overidentification arises naturally from embedding estimators of "smooth" (i.e. regular or root-n estimable for n the sample size) parameters and specifications tests in a common limiting experiment of LeCam (1986). This enables us to establish several equivalent characterizations of local overidentification. In particular, we show that if P is locally just identified by $\mathbf{P}$, then: (i) *All* asymptotically linear regular estimators of *any* common "smooth" parameter must be first order equivalent; and (ii) The local asymptotic power of *any* local asymptotic level α specification test cannot exceed α along *all* paths approaching P from outside $\mathbf{P}$. Moreover, we establish that the local overidentification of P by a regular model $\mathbf{P}$ is equivalent to *both*: (i) The existence of asymptotically distinct linear regular estimators for any "smooth" parameter

[1] See Section 5 for a partial extension of our results for regular models to non-regular models in which $T(P)$ is a convex cone.

that admits one such estimator; and (ii) The existence of a locally unbiased asymptotic level α specification test with non-trivial power against some path approaching P from outside $\mathbf{P}$.

Our equivalent characterizations of local overidentification are very useful. They offer researchers seemingly different yet equivalent ways to verify whether a data distribution P is locally overidentified by a complicated semi/nonparametric regular model $\mathbf{P}$. One obvious way is to directly verify the definition by first computing the closure of the tangent set $T(P)$ and then checking whether it is a strict subset of the space of all possible scores. An equivalent but sometimes simpler approach is to examine whether it is possible to obtain two asymptotically distinct regular estimators of a common "smooth" function of $P \in \mathbf{P}$,[2] such as the cumulative distribution function or a mean parameter $\int f dP$ for a known bounded function f. Given some structure on $\mathbf{P}$, it is often easy to compute two root-n consistent asymptotically normal estimators of a simple common "smooth" parameter, say as approximate optimizers of weighted criterion functions with different weights, and then verify whether their asymptotic variances differ.

The local overidentification condition by itself, however, may not lead to feasible efficient estimators of parameters of interest nor to feasible tests with nontrivial local power in general regular models. Indeed, in parallel to GMM, additional regularity conditions are required to accomplish the latter two objectives. In our general setting, these regularity conditions are imposed by assuming the existence of a score statistic (a stochastic process) $\hat{\mathbb{G}}_n$ whose marginals are first order equivalent to sample means of scores orthogonal to the tangent set $T(P)$. We show that such a score statistic $\hat{\mathbb{G}}_n$ can be constructed from two asymptotically distinct regular estimators of a common "smooth" parameter of $P \in \mathbf{P}$ [3] – a result that can be exploited to provide low level sufficient conditions for the availability of $\hat{\mathbb{G}}_n$ given additional structure on $\mathbf{P}$. In addition, we show that $\hat{\mathbb{G}}_n$ can be used to obtain locally unbiased nontrivial specification tests. The constructed tests encompass, among others, Hausman (1978) type test, and criterion-based tests such as the J test of Sargan (1958) and Hansen (1982) as special cases. In particular, proceeding in analogy to an incremental J test proposed in Eichenbaum et al. (1988) for GMM models, we demonstrate, for general regular models $\mathbf{P}$ and $\mathbf{M}$ satisfying $\mathbf{P} \subset \mathbf{M}$, how to build specification tests that aim their power at deviations from $\mathbf{P}$ that satisfy the maintained larger model $\mathbf{M}$.

We deduce from the described results that the equivalence between efficiency gains and nontrivial specification tests found in Hansen (1982) is not coincidental, but rather the reflection of a deeper principle applicable to all regular models. Our results on local overidentification in general regular models should be widely applicable. For example, our equivalent characterizations immediately imply that semi/nonparametric models of conditional moment restrictions (with a common conditioning set) containing unknown functions of potentially endogenous variables are locally overidentified because they allow for both inefficient and efficient estimators (Ai and Chen, 2003; Chen and Pouzo, 2009). Hence locally unbiased nontrivial specification tests of these models exist. Our results further show that the optimally weighted sieve quasi likelihood ratio tests of Chen and Pouzo (2009, 2015) direct the power at deviations of $\mathbf{P}$ that remain within a larger model $\mathbf{M}$. We also show that Hausman (1978) type tests that compare estimators efficient under

[2]We stress that a "smooth" parameter of $P \in \mathbf{P}$ always exists and does not need to be any structural parameter associated with the model $\mathbf{P}$; see Remark 3.1.

[3]We also establish a converse, that is, the availability of such a score statistic $\hat{\mathbb{G}}_n$ yields asymptotically distinct regular estimators of any common "smooth" parameter of $P \in \mathbf{P}$.

$\mathbf{P}$ to those efficient under a larger model $\mathbf{M}$ aim the power at violations of $\mathbf{P}$ that remain within $\mathbf{M}$. Therefore, both kinds of tests could be understood as generalized incremental J tests.

In this paper, we focus on a new application to nonparametric conditional moment restriction models with possibly different conditioning sets and potential endogeneity. We derive simple equivalent characterizations of local just identification for this very large class of models so that other researchers do not need to compute the closure of the tangent set $T(P)$ case by case. When specialized to nonparametric conditional moment restrictions with possibly different conditioning sets but without endogeneity, such as nonparametric conditional mean or quantile regressions, our characterization of P being locally just identified reduces to the condition of the nonparametric functions being "exactly identified" in Ackerberg et al. (2014) for such models. When specialized to semi/nonparametric models using a control function approach for endogeneity (Heckman, 1990; Olley and Pakes, 1996; Newey et al., 1999; Blundell and Powell, 2003), our characterization implies that P is typically locally overidentified by such models. When specialized to the semi/nonparametric models of sequential moment restrictions containing unknown functions of potentially endogenous variables, our characterization implies that P is typically locally overidentified, which is consistent with the semiparametric efficiency bound calculation in Ai and Chen (2012) for such models. In Section 4, our results are applied further to determining when nonparametric plug-in and semiparametric two-step GMM estimators are semiparametrically efficient. Empirically relevant examples of semi/nonparametric structural models are also presented.

The rest of the paper is organized as follows. Section 2 formally defines local overidentification while Section 3 establishes its connections to efficient estimation and testing in regular models. Section 4 applies the general theoretical results to characterize local overidentification in nonparametric conditional moment restriction models with possibly different information sets and potential endogeneity. Section 5 provides a partial extension of the main theoretical results in Section 3 for regular models to non-regular models in which $T(P)$ is a convex cone. Section 6 briefly concludes. Appendix A provides a short discussion of limiting experiments. Appendix B contains the proofs for Sections 2 and 3 while Appendix C contains the proofs for Section 5. The Online Appendix contains additional technical Lemmas, examples, and the proofs for Section 4.

2 Local Overidentification

2.1 Main Definition

We let $\mathcal{M}$ denote the collection of all probability measures on a measurable space $(\mathbf{X}, \mathcal{B})$. A model $\mathbf{P}$ is a (not necessarily strict) subset of $\mathcal{M}$. Typically, a model $\mathbf{P}$ is indexed by (model) parameters that consist of parameters of interest and perhaps additional nuisance parameters. We say a model $\mathbf{P}$ is semiparametric if the parameters of interest are finite dimensional but the nuisance parameters are infinite dimensional (such as the GMM model of Hansen (1982)); semi-nonparametric if the parameters of interest contain both finite and infinite dimensional parts; nonparametric if all the parameters are infinite dimensional. We call a model $\mathbf{P}$ fully unrestricted if $\mathbf{P} = \mathcal{M}$.

Throughout, the data $\{X_i\}_{i=1}^n$ is assumed to be an i.i.d. sample from a distribution $P \in \mathbf{P}$ of $X \in \mathbf{X}$. We call P the data distribution, which is always identified from the data, although its associated model parameters might not be. Our analysis is local in nature and hence we introduce suitable perturbations to P. Following the literature on limiting experiments (LeCam, 1986), we consider arbitrary smooth parametric likelihoods, which are defined by:

Definition 2.1. *A "path" $t \mapsto P_{t,g}$ is a function defined on $[0,1)$ such that $P_{t,g}$ is a probability measure on $(\mathbf{X}, \mathcal{B})$ for every $t \in [0,1)$, $P_{0,g} = P$, and*

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

The scalar measurable function $g : \mathbf{X} \to \mathbf{R}$ is referred to as the "score" of the path $t \mapsto P_{t,g}$.

For any σ-finite positive measure μ_t dominating $(P_t + P)$, the integral in (1) is understood as

$$\int \left[\frac{1}{t}\left(\left(\frac{dP_{t,g}}{d\mu_t}\right)^{1/2} - \left(\frac{dP}{d\mu_t}\right)^{1/2}\right) - \frac{1}{2} g \left(\frac{dP}{d\mu_t}\right)^{1/2}\right]^2 d\mu_t$$

(the choice of μ_t does not affect the value of the integral). Heuristically, a path is a parametric model that passes through P and is smooth in the sense of satisfying (1) or, equivalently, being differentiable in quadratic mean. We note Definition 2.1 implies any score must have mean zero and be square integrable with respect to P, and therefore belong to the space $L_0^2(P)$ given by

$$L_0^2(P) \equiv \left\{g : \mathbf{X} \to \mathbf{R}, \quad \int g dP = 0 \text{ and } \|g\|_{P,2} < \infty\right\}, \qquad \|g\|_{P,2}^2 \equiv \int g^2 dP. \tag{2}$$

The restriction $g \in L_0^2(P)$ is solely the result of $P_{t,g} \in \mathcal{M}$ for all t in a neighborhood of zero. If we in addition demand that $P_{t,g} \in \mathbf{P}$, then the set of feasible scores reduces to

$$T(P) \equiv \{g \in L_0^2(P) : \text{ (1) holds for some } t \mapsto P_{t,g} \in \mathbf{P}\}, \tag{3}$$

which is called the *tangent set* at P. Finally, we let $\bar{T}(P)$ denote the *closure* of $T(P)$ under $\|\cdot\|_{P,2}$. By definition, $\bar{T}(P)$ is a (not necessarily strict) subset of $L_0^2(P)$. For instance, if $\mathbf{P} = \mathcal{M}$, then $\bar{T}(P) = T(P) = L_0^2(P)$ for any $P \in \mathbf{P}$.

Given the introduced notation, we can now formally define local overidentification.

Definition 2.2. *If $\bar{T}(P) = L_0^2(P)$, then we say P is locally just identified by $\mathbf{P}$. Conversely, if $\bar{T}(P) \neq L_0^2(P)$, then we say P is locally overidentified by $\mathbf{P}$.*

Intuitively, P is locally overidentified by a model $\mathbf{P}$ if $\mathbf{P}$ yields meaningful restrictions on the scores that can be generated by parametric submodels. Conversely, P is locally just identified by $\mathbf{P}$ when the sole imposed restriction is that the scores have mean zero and a finite second moment – a quality common to the scores of all paths regardless of whether they belong to $\mathbf{P}$ or not. It is clear that Definition 2.2 is inherently local in that it concerns only the "shape" of $\mathbf{P}$ at the point P rather than $\mathbf{P}$ in its entirety as would be appropriate for a global notion of overidentification.

Remark 2.1. Koopmans and Riersol (1950) refer to a model $\mathbf{P}$ as overidentified whenever there is a possibility that P does not belong to $\mathbf{P}$. Thus, $\mathbf{P}$ is deemed globally overidentified if $\mathbf{P} \neq \mathcal{M}$ (i.e. $\mathbf{P}$ is a strict subset of $\mathcal{M}$), and globally just identified if $\mathbf{P} = \mathcal{M}$ (i.e. $\mathbf{P}$ is fully unrestricted).

Clearly, global just identification implies local just identification, while local overidentification implies global overidentification. Although more demanding, local overidentification will provide a stronger connection to both the testability of $\mathbf{P}$ and the performance of regular estimators. ■

It is worth emphasizing that local overidentification concerns solely a relationship between the data distribution P and a model $\mathbf{P}$. Hence, it is possible for P to be locally overidentified despite underlying (structural) parameters of the model $\mathbf{P}$ being partially identified – an observation that simply reflects the fact that partially identified models may still be refuted by the data. See, e.g., Koopmans and Riersol (1950); Hansen and Jagannathan (1997); Manski (2003); Haile and Tamer (2003); Hansen (2014) and references therein.

2.2 Equivalent Definitions in Regular Models

In many applications, the following condition holds and simplifies our analysis.

Assumption 2.1. *(i) $\{X_i\}_{i=1}^n$ is an i.i.d. sequence with $X_i \in \mathbf{X}$ distributed according to $P \in \mathbf{P}$; (ii) $T(P)$ is linear – i.e. if $g, f \in T(P)$, $a, b \in \mathbf{R}$, then $ag + bf \in T(P)$.*

The i.i.d. requirement in Assumption 2.1(i) may be relaxed but is imposed to streamline exposition. Assumption 2.1(ii) requires the model $\mathbf{P}$ to be *regular* at P in the sense that its tangent set be linear. This is satisfied by numerous models (such as the GMM model), and is either implicitly or explicitly imposed whenever semiparametric efficiency bounds and efficient estimators are considered (Hájek, 1970; Hansen, 1985; Chamberlain, 1986; Newey, 1990; Bickel et al., 1993; Ai and Chen, 2003). We stress, however, that a model $\mathbf{P}$ being regular does not imply that all the parameters underlying the model are regular (i.e., "smooth" or root-n estimable). In fact, some parameters of a regular model $\mathbf{P}$ may only be slower-than root-n estimable or not even be identified. Nonetheless, Assumption 2.1(ii) does rule out models in which the tangent set $T(P)$ is not linear but a convex cone instead. See Section 5 for a partial extension of the main results for regular models to non-regular models where $T(P)$ is a convex cone.

In the literature, the closed linear span of $T(P)$ under $\|\cdot\|_{P,2}$ is called the *tangent space* at $P \in \mathbf{P}$ (see, e.g., Definition 3.2.2 in Bickel et al. (1993)). Under Assumption 2.1(ii), $\bar{T}(P)$ becomes the tangent space at P, and hence a vector subspace of $L_0^2(P)$. We also define

$$\bar{T}(P)^{\perp} \equiv \left\{ g \in L_0^2(P) : \int g f dP = 0 \text{ for all } f \in \bar{T}(P) \right\}, \tag{4}$$

which is the orthogonal complement of $\bar{T}(P)$. The vector spaces $\bar{T}(P)$ and $\bar{T}(P)^{\perp}$ then form an orthogonal decomposition of $L_0^2(P)$ (the space of all possible scores)

$$L_0^2(P) = \bar{T}(P) \oplus \bar{T}(P)^{\perp}, \tag{5}$$

and we let $\Pi_T(\cdot)$ and $\Pi_{T^{\perp}}(\cdot)$ denote the orthogonal projections under $\|\cdot\|_{P,2}$ onto $\bar{T}(P)$ and $\bar{T}(P)^{\perp}$ respectively. Every $g \in L_0^2(P)$ then satisfies $g = \Pi_T(g) + \Pi_{T^{\perp}}(g)$ and $Var(g) = Var(\Pi_T(g)) + Var(\Pi_{T^{\perp}}(g))$. Intuitively, $\Pi_T(g) \in \bar{T}(P)$ is the component of g that is in accord with model $\mathbf{P}$, while $\Pi_{T^{\perp}}(g) \in \bar{T}(P)^{\perp}$ is the component orthogonal to $\mathbf{P}$.

The decomposition in (5) implies equivalent characterizations of local overidentification that we summarize in the following simple yet useful Lemma.

Lemma 2.1. *Under Assumption 2.1, the following are equivalent to Definition 2.2:*

(i) *P is locally just identified by $\mathbf{P}$ if and only if $\bar{T}(P)^{\perp} = \{0\}$, or equivalently, $Var(\Pi_{T^{\perp}}(g)) = 0$ for all $g \in L_0^2(P)$.*

(ii) *P is locally overidentified by $\mathbf{P}$ if and only if $\bar{T}(P)^{\perp} \neq \{0\}$, or equivalently, $Var(\Pi_{T^{\perp}}(g)) > 0$ for some $g \in L_0^2(P)$.*

We next illustrate the introduced concepts in the GMM model.[4]

GMM Illustration. Let $\Gamma \subseteq \mathbf{R}^{d_\gamma}$ with $d_\gamma < \infty$ be the parameter space and $\rho : \mathbf{X} \times \mathbf{R}^{d_\gamma} \to \mathbf{R}^{d_\rho}$ be a known moment function with $d_\rho \geq d_\gamma$. The GMM model $\mathbf{P}$ is

$$\mathbf{P} \equiv \left\{ P \in \mathcal{M} : \int \rho(\cdot, \gamma) dP = 0 \text{ for some } \gamma \in \Gamma \right\}, \tag{6}$$

and for any $P \in \mathbf{P}$ we let $\gamma(P)$ solve $\int \rho(\cdot, \gamma(P)) dP = 0$. For simplicity, let ρ be differentiable in γ, and set $D(P) \equiv \int \nabla_\gamma \rho(\cdot, \gamma(P)) dP$. For any path $t \mapsto P_{t,g} \in \mathbf{P}$, we then obtain

$$0 = \frac{d}{dt} \int \rho(\cdot, \gamma(P_{t,g})) dP_{t,g} \Big|_{t=0} = \int \rho(\cdot, \gamma(P)) g dP + D(P) \dot{\gamma}(g), \tag{7}$$

where $\dot{\gamma}(g)$ is the derivative of $\gamma(P_{t,g})$ at $t = 0$. If $\int \rho(\cdot, \gamma(P)) \rho(\cdot, \gamma(P))' dP$ is full rank, then the linear functional $g \mapsto \int \rho(\cdot, \gamma(P)) g dP$ maps $L_0^2(P)$ onto $\mathbf{R}^{d_\rho}$. On the other hand, $D(P)$ maps $\mathbf{R}^{d_\gamma}$ onto a linear subspace of $\mathbf{R}^{d_\rho}$ whose dimension equals the rank of $D(P)$. Therefore, (7) imposes restrictions on the possible set of scores g only when the rank of $D(P)$ is smaller than d_ρ. When $D(P)$ is full rank, we thus obtain that P is locally just identified by $\mathbf{P}$ if and only if the "standard" GMM just identification condition that $d_\rho = d_\gamma$ is satisfied. ■

Our definition of local overidentification extends that in GMM models to general infinite dimensional models. This will be very useful for nonparametric conditional moment restriction models, where both the number of parameters (of interest) and the number of (unconditional) moments are infinite. Moreover, for general regular models, we will show Definition 2.2 retains the fundamental link to the properties of regular estimators and specification tests present in Hansen (1982). For instance, just as all regular estimators of $\gamma(P)$ in the GMM model are asymptotically equivalent whenever $d_\rho = d_\gamma$, Theorem 2.1 in Newey (1994) has shown that the asymptotic variance of root-n consistent plug-in estimators is invariant to the choice of first-stage nonparametric estimators whenever $L_0^2(P) = \bar{T}(P)$.

3 General Results for Regular Models

In this Section we show that, in general regular models, local overidentification is intrinsically linked to the importance of efficiency considerations and the potential refutability of a model.

[4]We thank Lars Peter Hansen for sharing with us his notes on GMM and for helping us with the GMM example.

3.1 The Setup

3.1.1 The Setup: Estimation

Since the data distribution P is always identified, many known functions of P are identified and consistently estimable even if some underlying (structural) parameters of a model $\mathbf{P}$ are not identified. For general regular models, we therefore represent an identifiable parameter as a known mapping $\theta : \mathbf{P} \to \mathbf{B}$ and the "true" parameter value as $\theta(P) \in \mathbf{B}$, where $\mathbf{B}$ is a Banach space with norm $\|\cdot\|_{\mathbf{B}}$. We further denote the dual space of $\mathbf{B}$ by $\mathbf{B}^* \equiv \{b^* : \mathbf{B} \to \mathbf{R} : b^* \text{ is linear, } \|b^*\|_{\mathbf{B}^*} < \infty\}$, which is the space of continuous linear functionals with norm $\|b^*\|_{\mathbf{B}^*} \equiv \sup_{\|b\|_{\mathbf{B}} \leq 1} |b^*(b)|$.

An estimator $\hat{\theta}_n : \{X_i\}_{i=1}^n \to \mathbf{B}$ for $\theta(P) \in \mathbf{B}$ is a map from the data into the space $\mathbf{B}$. To address the question of whether $\theta(P)$ admits asymptotically distinct estimators (i.e. efficiency "matters") we focus on *asymptotically linear regular* estimators. In what follows, for any path $t \mapsto P_{t,g} \in \mathcal{M}$ we use the notation $\overset{L_{n,g}}{\to}$ to represent convergence in law under $P^n_{1/\sqrt{n},g} \equiv \bigotimes_{i=1}^n P_{1/\sqrt{n},g}$, and $\overset{L}{\to}$ for convergence in law under $P^n \equiv \bigotimes_{i=1}^n P$.

Definition 3.1. $\hat{\theta}_n : \{X_i\}_{i=1}^n \to \mathbf{B}$ *is a regular estimator of* $\theta(P)$ *if there is a tight random variable* $\mathbb{D}$ *such that* $\sqrt{n}\{\hat{\theta}_n - \theta(P_{1/\sqrt{n},g})\} \overset{L_{n,g}}{\to} \mathbb{D}$ *for any path* $t \mapsto P_{t,g} \in \mathbf{P}$.

Definition 3.2. $\hat{\theta}_n : \{X_i\}_{i=1}^n \to \mathbf{B}$ *is an asymptotically linear estimator of* $\theta(P)$ *if*

$$\sqrt{n}\{\hat{\theta}_n - \theta(P)\} = \frac{1}{\sqrt{n}} \sum_{i=1}^n \nu(X_i) + o_p(1) \quad \text{under } P^n, \tag{8}$$

for some $\nu : \mathbf{X} \to \mathbf{B}$ *satisfying* $b^*(\nu) \in L^2_0(P)$ *for any* $b^* \in \mathbf{B}^*$. *Here* ν *is called the influence function of the estimator* $\hat{\theta}_n$.

By restricting attention to regular estimators, we focus on root-n consistent estimators whose asymptotic distribution is invariant to local perturbations to P within the model $\mathbf{P}$. While most commonly employed estimators are regular and asymptotically linear, their existence does impose restrictions on the map $\theta : \mathbf{P} \to \mathbf{B}$. In fact, the existence of an asymptotically linear regular estimator of $\theta(P)$ in regular models implies $\theta : \mathbf{P} \to \mathbf{B}$ must be "pathwise differentiable" (or "smooth") relative to $T(P)$ (van der Vaart, 1991b).

Remark 3.1. Regardless of a model $\mathbf{P}$ being regular or non-regular, there always exists a "smooth" map $\theta : \mathbf{P} \to \mathbf{B}$ and an asymptotically linear regular estimator $\hat{\theta}_n : \{X_i\}_{i=1}^n \to \mathbf{B}$ of $\theta(P)$ under i.i.d. data. For example, for any bounded function $f : \mathbf{X} \to \mathbf{R}$, the sample mean, $n^{-1} \sum_{i=1}^n f(X_i)$, is an asymptotically linear regular estimator of $\theta(P) \equiv \int f dP$ along *any* path $t \mapsto P_{t,g} \in \mathcal{M}$. Thus, we emphasize that $\theta(P)$ should not be solely thought of as an intrinsic parameter of the model $\mathbf{P}$, but rather as any "smooth" map of $P \in \mathbf{P}$. ■

3.1.2 The Setup: Testing

A specification test for a general model $\mathbf{P}$ is a test of the null hypothesis that P belongs to $\mathbf{P}$ against the alternative that it does not – i.e. it is a test of the hypotheses

$$H_0 : P \in \mathbf{P} \qquad \text{vs} \qquad H_1 : P \in \mathcal{M} \setminus \mathbf{P} \,. \tag{9}$$

We denote an arbitrary (possibly randomized) test of (9) by $\phi_n : \{X_i\}_{i=1}^n \to [0,1]$, which is a function specifying for each realization of the data a corresponding probability of rejecting the null hypothesis. In our analysis, we restrict attention to specification tests ϕ_n that have *local asymptotic level* α and possess an *asymptotic local power function.*

Definition 3.3. *A specification test $\phi_n : \{X_i\}_{i=1}^n \to [0,1]$ for a model $\mathbf{P}$ has local asymptotic level α if for any path $t \mapsto P_{t,g} \in \mathbf{P}$ it follows*

$$\limsup_{n\to\infty} \int \phi_n dP^n_{1/\sqrt{n},g} \leq \alpha. \tag{10}$$

Definition 3.4. *A specification test $\phi_n : \{X_i\}_{i=1}^n \to [0,1]$ for a model $\mathbf{P}$ has a local asymptotic power function $\pi : L_0^2(P) \to [0,1]$ if for any path $t \mapsto P_{t,g} \in \mathcal{M}$, it follows*

$$\lim_{n\to\infty} \int \phi_n dP^n_{1/\sqrt{n},g} = \pi(g). \tag{11}$$

Finally, a test ϕ_n for (9) with a local asymptotic power function π is said to be locally *unbiased* if it satisfies: $\pi(g) \leq \alpha$ for all $t \mapsto P_{t,g} \in \mathbf{P}$ and $\pi(g) \geq \alpha$ for all $t \mapsto P_{t,g} \in \mathcal{M} \setminus \mathbf{P}$.

Note that a local asymptotic power function only depends on the score $g \in L_0^2(P)$ and is independent of any other characteristics of the path $t \mapsto P_{t,g} \in \mathcal{M}$. This is because the product measures of any two local paths that share the same score must converge in the Total Variation metric (see Lemma D.1 in the Online Appendix). Intuitively, a test possesses a local asymptotic power function if the limiting rejection probability of the test is well defined along *any* local perturbation to P. The existence of a local asymptotic power function is a mild requirement that is typically satisfied; see Remark 3.2.

Remark 3.2. Tests ϕ_n are often constructed by comparing a test statistic $\hat{T}_n$ to an estimate of the $(1-\alpha)$ quantile of its asymptotic distribution. By LeCam's 3rd Lemma and the Portmanteau Theorem, such tests have a local asymptotic power function provided that: (i) $(\hat{T}_n, \frac{1}{\sqrt{n}} \sum_{i=1}^n g(X_i)) \in \mathbf{R}^2$ converges jointly in distribution under P^n for any $g \in L_0^2(P)$, and (ii) the limiting distribution of $\hat{T}_n$ under P^n is continuous. See Theorem 6.6 in van der Vaart (1998). ■

3.2 Equivalent Characterizations of Local Overidentification

In Hansen (1982)'s GMM framework, overidentifying restrictions are necessary for both the existence of efficiency gains in estimation and the testability of the model. We now extend this conclusion to general regular models.

Theorem 3.1. *Let Assumption 2.1 hold and P be locally just identified by $\mathbf{P}$.*

(i) Let $\hat{\theta}_n : \{X_i\}_{i=1}^n \to \mathbf{B}$ and $\tilde{\theta}_n : \{X_i\}_{i=1}^n \to \mathbf{B}$ be any asymptotically linear regular estimators of any parameter $\theta(P) \in \mathbf{B}$. Then: $\sqrt{n}\{\hat{\theta}_n - \tilde{\theta}_n\} = o_p(1)$ in $\mathbf{B}$.

(ii) Let ϕ_n be any specification test for (9) with local asymptotic level α and a local asymptotic power function π. Then: $\pi(g) \leq \alpha$ for all paths $t \mapsto P_{t,g} \in \mathcal{M}$.

Theorem 3.1 establishes that the local overidentification of P is a necessary condition for the existence of efficiency gains and nontrivial specification tests. Specifically, Theorem 3.1(i) shows that if P is locally just identified, then *all* asymptotically linear regular estimators of *any* "smooth" parameter $\theta(P)$ must be first order equivalent. This conclusion is a generalization of Newey (1990) who showed scalar (i.e. $\mathbf{B} = \mathbf{R}$) asymptotically linear and regular estimators are first order equivalent when $\bar{T}(P) = L_0^2(P)$. Theorem 3.1(ii) establishes that if P is locally just identified by $\mathbf{P}$, then the local asymptotic power of *any* local asymptotic level α specification test cannot exceed α along *any* path, including all paths approaching P from outside $\mathbf{P}$. Heuristically, under local just identification, the set of scores corresponding to paths $t \mapsto P_{t,g} \in \mathbf{P}$ is dense in the set of all possible scores and hence every path is locally on the "boundary" of the null hypothesis.

In order to discern how the local overidentification of P can facilitate the existence of efficiency gains and the testability of the model, we consider the asymptotic behavior of sample means of scores. For any $0 \neq \tilde{f} \in L_0^2(P)$, if X_i is distributed according to $P_{1/\sqrt{n},g}$ for a path $t \mapsto P_{t,g} \in \mathcal{M}$, then

$$\mathbb{G}_n(\tilde{f}) \equiv \frac{1}{\sqrt{n}} \sum_{i=1}^n \tilde{f}(X_i) \overset{L_{n,g}}{\to} N\Big(\int \tilde{f} g dP, \int \tilde{f}^2 dP\Big) \tag{12}$$

by LeCam's 3rd Lemma. Recall that for regular models local overidentification is equivalent to the existence of at least one score $0 \neq \tilde{f} \in \bar{T}(P)^\perp$. For any such $0 \neq \tilde{f} \in \bar{T}(P)^\perp$ and *all* path $t \mapsto P_{t,g} \in \mathbf{P}$, we have $\int \tilde{f} g dP = 0$ and hence $\mathbb{G}_n(\tilde{f})$ converges to a centered Gaussian random variable – i.e. $\mathbb{G}_n(\tilde{f})$ behaves as "noise" that can alter the efficiency of estimators. On the other hand, for any $0 \neq \tilde{f} \in \bar{T}(P)^\perp$, there is a path $t \mapsto P_{t,g} \in \mathcal{M} \setminus \mathbf{P}$ such that $\int \tilde{f} g dP \neq 0$, and hence $\mathbb{G}_n(\tilde{f})$ can be employed to construct an asymptotically locally nontrivial specification test – i.e. $\mathbb{G}_n(\tilde{f})$ is a "signal" that enables the detection of violations of the model $\mathbf{P}$. Our next result builds on this intuition by using the score statistics $\mathbb{G}_n(\tilde{f})$ to establish a converse to Theorem 3.1.

Theorem 3.2. *Let Assumption 2.1 hold. Then the following statements are equivalent:*

(i) P is locally overidentified by $\mathbf{P}$.

(ii) If a parameter $\theta(P) \in \mathbf{B}$ admits an asymptotically linear regular estimator $\hat{\theta}_n$, then: there exists another asymptotically linear regular estimator $\tilde{\theta}_n$ of $\theta(P)$ such that $\sqrt{n}\{\hat{\theta}_n - \tilde{\theta}_n\} \overset{L}{\to} \Delta \neq 0$ in $\mathbf{B}$.

(iii) There exists a locally unbiased asymptotic level α test ϕ_n for (9) with a local asymptotic power function π such that: $\pi(g) > \alpha$ for some path $t \mapsto P_{t,g} \in \mathcal{M} \setminus \mathbf{P}$.

Theorems 3.1 and 3.2 establish that the local overidentification of P is equivalent to the availability of efficiency gains and also to the existence of locally nontrivial specification tests. In addition, Theorems 3.1(i) and 3.2(i)-(ii) imply the following equivalent characterization of local just identification.

Corollary 3.1. *Let Assumption 2.1 hold and $\mathcal{D}$ be a set of bounded functions that is dense in $(L^2(P), \|\cdot\|_{P,2})$. For any $f \in \mathcal{D}$ let Ω_f^* be the semiparametric efficient variance bound for estimating $\int f dP$ under $\mathbf{P}$. Then: $\Omega_f^* = Var\{f(X)\}$ for all $f \in \mathcal{D}$ if and only if P is locally just identified by $\mathbf{P}$.*

This Corollary is very useful in assessing whether P is locally overidentified by a complicated model $\mathbf{P}$. For example, in Subsection 4.1.1 we employ Corollary 3.1 and the semiparametric efficiency bound analysis in Ai and Chen (2012) to characterize local overidentification in nonparametric models defined by sequential moment restrictions.

3.3 Feasible Estimators and Tests

The intuition for Theorem 3.2 suggests that *any* statistics asymptotically equivalent to the score statistics $\mathbb{G}_n(\tilde{f})$ with some $0 \neq \tilde{f} \in \bar{T}(P)^{\perp}$ (see (12)) may be employed to obtain distinct regular estimators for *arbitrary* "smooth" parameters $\theta(P)$ and specification tests with nontrivial local power. To elaborate on this intuition, for any set A we let $\ell^{\infty}(A) \equiv \{f : A \to \mathbf{R} \text{ s.t. } \|f\|_{\infty} < \infty\}$ where $\|f\|_{\infty} = \sup_{a \in A} |f(a)|$, and impose the following condition:

Assumption 3.1. *For some set $\mathbf{T}$ there is a statistic $\hat{\mathbb{G}}_n : \{X_i\}_{i=1}^n \to \ell^{\infty}(\mathbf{T})$ satisfying:*
(i) $\hat{\mathbb{G}}_n(\tau) = \frac{1}{\sqrt{n}} \sum_{i=1}^n s_\tau(X_i) + o_p(1)$ uniformly in $\tau \in \mathbf{T}$, where $0 \neq s_\tau \in \bar{T}(P)^{\perp}$ for all $\tau \in \mathbf{T}$;
(ii) for some tight nondegenerate centered Gaussian measure $\mathbb{G}_0$, $\hat{\mathbb{G}}_n \xrightarrow{L} \mathbb{G}_0$ in $\ell^{\infty}(\mathbf{T})$.

Assumption 3.1 requires the availability of a statistic $n^{-1/2}\hat{\mathbb{G}}_n(\tau)$ that is first order equivalent to the sample mean of some score (or influence function) $s_\tau \in \bar{T}(P)^{\perp}$. We further let

$$S(P) \equiv \left\{ s_\tau \in \bar{T}(P)^{\perp} : \tau \in \mathbf{T} \right\} \tag{13}$$

denote the collection of such scores, which will play an important role in our analysis. As we argue below, statistics $\hat{\mathbb{G}}_n$ satisfying Assumption 3.1 are implicitly constructed by various specification tests, such as Hausman tests and criterion based tests; see Remark 3.5. In order to establish a connection to Hausman tests in particular, we introduce the following Assumption:

Assumption 3.2. *For some parameter $\theta(P) \in \mathbf{B}$ there are asymptotically linear regular estimators $\hat{\theta}_n$ and $\tilde{\theta}_n$ with influence functions ν and $\tilde{\nu}$ such that $\sqrt{n}\{\hat{\theta}_n - \tilde{\theta}_n\} \xrightarrow{L} \Delta \neq 0$.*

Assumption 3.2 simply requires the existence of two distinct estimators of some "smooth" function of $P \in \mathbf{P}$, which needs not be structural parameter of the model $\mathbf{P}$; see Remark 3.1.

Lemma 3.1. *Let Assumption 2.1 hold.*

(i) Let Assumption 3.1 hold. Then: For any parameter $\theta(P) \in \mathbf{B}$ that admits an asymptotically linear regular estimator $\hat{\theta}_n$, Assumption 3.2 is satisfied with $\tilde{\theta}_n = \hat{\theta}_n + \tilde{b} \times n^{-1/2}\hat{\mathbb{G}}_n(\tau^*)$ and $\Delta = -\tilde{b} \times \mathbb{G}_0(\tau^*)$ for some $0 \neq \tilde{b} \in \mathbf{B}$ and some $\tau^* \in \mathbf{T}$.

(ii) Let Assumption 3.2 hold. Then: Assumption 3.1 is satisfied with $\mathbf{T} = \{b^* \in \mathbf{B}^* : \|b^*\|_{\mathbf{B}^*} \leq 1\}$, and $\mathbb{G}_0$, $\hat{\mathbb{G}}_n \in \ell^{\infty}(\mathbf{T})$ given by $\mathbb{G}_0(b^*) = b^*(\Delta)$, $\hat{\mathbb{G}}_n(b^*) = b^*(\sqrt{n}\{\hat{\theta}_n - \tilde{\theta}_n\})$ where $s_{b^*} = b^*(\nu - \tilde{\nu})$.

Lemma 3.1 establishes that Assumptions 3.1 and 3.2 are equivalent to each other. In particular, Lemma 3.1(ii) shows that the difference of any two asymptotically distinct linear regular estimators of *any* common parameter $\theta(P)$ may be employed to construct $\hat{\mathbb{G}}_n$ – i.e. Assumption 3.2 implies Assumption 3.1. As a result, given the specific structure of a regular model $\mathbf{P}$, it is straightforward to obtain lower level sufficient conditions for Assumption 3.1. Specifically, we need only ensure the existence of two asymptotically distinct linear regular estimators of some "smooth" parameter, which could be a simple identified reduced form parameter if the structural parameters are not identified. In a large class of semiparametric and nonparametric models, asymptotically distinct estimators may be found as the optimizers of weighted criterion functions with alternative choices of weights. See, e.g., Shen (1997) for efficient estimation based on sieve or penalized maximum likelihood in semi/nonparametric likelihood models, and Ai and Chen (2003) for efficient estimation based on optimally weighted sieve minimum distance of semi/nonparametric conditional moment restrictions models. In the Online Appendix we apply Lemma 3.1(ii) to verify Assumption 3.1 in nonparametric conditional moment restriction models.

We next employ the fact that $\hat{\mathbb{G}}_n$ behaves as a "signal" from a testing perspective (i.e. Theorem 3.2(iii)) to construct nontrivial local specification tests. Let $\bar{S}(P) \equiv \overline{\text{lin}}\{S(P)\}$ be the closed linear span of $S(P)$ in $L^2_0(P)$, and $\Pi_S(g)$ be the metric projection of $g \in L^2_0(P)$ onto $\bar{S}(P)$. We note Assumption 3.1(i) (or (12)) implies that $\hat{\mathbb{G}}_n$ exhibits a non-zero asymptotic drift along a path $t \mapsto P_{t,g} \in \mathcal{M}$ if and only of $\Pi_S(g) \neq 0$. Intuitively $\bar{S}(P)$ therefore represents the alternatives for which specification tests based on $\hat{\mathbb{G}}_n$ have nontrival local asymptotic power. To obtain such a test, we employ a map $\Psi : \ell^\infty(\mathbf{T}) \to \mathbf{R}_+$ to reduce $\hat{\mathbb{G}}_n$ to a scalar test statistic $\hat{T}_n = \Psi(\hat{\mathbb{G}}_n)$.

Assumption 3.3. *(i)* $\Psi : \ell^\infty(\mathbf{T}) \to \mathbf{R}_+$ *is continuous, convex and nonconstant; (ii)* $\Psi(0) = 0$, $\Psi(b) = \Psi(-b)$ *for all* $b \in \ell^\infty(\mathbf{T})$*; (iii)* $\{b \in \ell^\infty(\mathbf{T}) : \Psi(b) \leq c\}$ *is bounded for all* $c > 0$.

Finally, we let $c_{1-\alpha} > 0$ be the $(1-\alpha)$ quantile of $\Psi(\mathbb{G}_0)$ and define the specification test

$$\phi_n \equiv 1\{\Psi(\hat{\mathbb{G}}_n) > c_{1-\alpha}\}; \tag{14}$$

i.e. we reject proper model specification for large values of $\Psi(\hat{\mathbb{G}}_n)$. Multiple specification tests in the literature are in fact asymptotically equivalent to (14) with different choices of Ψ; see Theorem 3.3 Part (ii) and Remark 3.5 below.

Theorem 3.3. *Let Assumption 2.1 hold.*

(i) Let Assumptions 3.1 and 3.3 hold. Then: ϕ_n *defined in* (14) *with* $c_{1-\alpha} > 0$ *is a locally unbiased asymptotic level* α *specification test for* (9) *with a local asymptotic power function* π*. Moreover, for any path* $t \mapsto P_{t,g} \in \mathcal{M}$ *with* $\Pi_S(g) \neq 0$ *it follows*

$$\pi(g) \equiv \lim_{n\to\infty} P_{1/\sqrt{n},g}\big(\Psi(\hat{\mathbb{G}}_n) > c_{1-\alpha}\big) > \alpha. \tag{15}$$

(ii) Let Assumption 3.2 hold. Then: Assumption 3.3 holds with $\Psi = \|\cdot\|_\infty$*, and Part (i) holds with* $\Psi(\hat{\mathbb{G}}_n) = \sqrt{n}\|\hat{\theta}_n - \tilde{\theta}_n\|_{\mathbf{B}}$*, and* $S(P) = \{b^*(\nu - \tilde{\nu}) : b^* \in \mathbf{B}^*, \|b^*\|_{\mathbf{B}^*} \leq 1\}$.

Theorems 3.1, 3.2, 3.3 and Lemma 3.1 link local overidentification to the existence of asymptotically distinct estimators and locally nontrivial specification tests. The latter two concepts were also intrinsically linked by the seminal work of Hausman (1978), who proposed comparing estimators of a common parameter to perform specification tests. Theorem 3.3(ii) shows Hausman tests are a special case of (14) in general regular models.

Remark 3.3. Whenever $\bar{S}(P) = \bar{T}(P)^{\perp}$, result (15) holds for any path $t \mapsto P_{t,g}$ with

$$\liminf_{n\to\infty} \inf_{Q\in\mathbf{P}} n \int \left[dQ^{1/2} - dP^{1/2}_{1/\sqrt{n},g}\right]^2 > 0; \tag{16}$$

i.e., the proposed test has nontrivial local power against *any* path that does not approach $\mathbf{P}$ "too fast". If condition (16) fails, then there is a sequence $Q_n \in \mathbf{P}$ for which

$$\limsup_{n\to\infty} \left| \int \phi_n (dQ_n^n - dP^n_{1/\sqrt{n},g}) \right| \leq \limsup_{n\to\infty} \|Q_n^n - P^n_{1/\sqrt{n},g}\|_{TV} = 0 \tag{17}$$

where $\|\cdot\|_{TV}$ denotes the total variation distance; see, e.g., Theorem 13.1.3 in Lehmann and Romano (2005). Therefore, a violation of (16) implies $P_{1/\sqrt{n},g}$ approaches $\mathbf{P}$ "too fast" in the sense that it is not possible to discriminate the induced distribution on the data $\{X_i\}_{i=1}^n$ from a distribution that is in accord with $\mathbf{P}$. ■

Remark 3.4. Theorem 3.3(ii) states a Hausman test has nontrivial local power against any path $t \mapsto P_{t,g} \in \mathcal{M}$ whose score g is correlated with $b^*(\nu - \tilde{\nu})$ for some $b^* \in \mathbf{B}^*$. When $\hat{\theta}_n$ is a semiparametric efficient estimator, it follows $\bar{S}(P) = \overline{\text{lin}}\{\Pi_{T^{\perp}}(b^*(\tilde{\nu})) : b^* \in \mathbf{B}^*\}$ (see Proposition 3.3.1 in Bickel et al. (1993)). In particular, $L_0^2(P) = \overline{\text{lin}}\{b^*(\tilde{\nu}) : b^* \in \mathbf{B}^*\}$ implies $\bar{S}(P) = \bar{T}(P)^{\perp}$, and hence the corresponding Hausman test has nontrivial power against all local alternatives. ■

Remark 3.5. In addition to Hausman tests, multiple specification tests for (9) are also asymptotically equivalent to test (14). For example, optimally weighed criterion based tests employ statistics $\hat{T}_n$ that have a chi-squared asymptotic distribution and satisfy

$$\hat{T}_n = \sum_{k=1}^{K} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} f_k(X_i) \right)^2 + o_p(1) \tag{18}$$

where K corresponds to the degrees of freedom and $\{f_k\}_{k=1}^K \subset L_0^2(P)$ are orthonormal. A test with this property can only have nominal and asymptotical local level α if $f_k \in \bar{T}(P)^{\perp}$ for all k. Otherwise, there is a path $t \mapsto P_{t,g} \in \mathbf{P}$ such that $\int g\{\Pi_T(f_k)\}dP \neq 0$ for at least one k, which by (12) leads to a null rejection probability exceeding α. As a result, the structure in test (14) is also present in the J test of Hansen (1982), the semiparametric LR statistic of Murphy and van der Vaart (1997), the sieve QLR statistic in Chen and Pouzo (2009), and the generalized emipirical likelihood test in Parente and Smith (2011) among many others. ■

3.4 Incremental J Tests

In applications, specification tests are often informed by a concern with a particular violation of the model. For instance, in GMM we may question the validity of a subset of the moment conditions but have confidence in the remaining ones; see, e.g., Eichenbaum et al. (1988). In such circumstances, a J test, which entertains the possibility of any moment being violated, can be less revealing than the so-called incremental J (Sargan-Hansen) test, which focuses on the specific moments that are of concern (Arellano, 2003).

The tests in Theorem 3.3 can similarly direct their power at specific violations of the model. To this end, we introduce a set $\mathbf{M}$ satisfying $\mathbf{P} \subseteq \mathbf{M} \subseteq \mathcal{M}$, which represents the characteristics of the model we believe P satisfies even when $P \notin \mathbf{P}$, and consider

$$H_0 : P \in \mathbf{P} \qquad \text{vs} \qquad H_1 : P \in \mathbf{M} \setminus \mathbf{P}. \tag{19}$$

The "maintained" model $\mathbf{M}$ generates its own tangent set, which we denote by

$$M(P) \equiv \big\{g \in L_0^2(P) : \ (1) \text{ holds for some } t \mapsto P_{t,g} \in \mathbf{M}\big\}, \tag{20}$$

with $\bar{M}(P)$ being the closure of $M(P)$ in $(L_0^2(P), \|\cdot\|_{P,2})$. If $M(P)$ is linear, then $\bar{M}(P) = \bar{T}(P) \oplus \{\bar{T}(P)^\perp \cap \bar{M}(P)\}$ and the space $L_0^2(P)$ of all possible scores satisfies

$$L_0^2(P) = \bar{M}(P) \oplus \bar{M}(P)^\perp = \bar{T}(P) \oplus \{\bar{T}(P)^\perp \cap \bar{M}(P)\} \oplus \bar{M}(P)^\perp; \tag{21}$$

i.e. any score consists of a component that agrees with $\mathbf{P}$ (in $\bar{T}(P)$), a component that disagrees with $\mathbf{P}$ but still agrees with $\mathbf{M}$ (in $\bar{T}(P)^\perp \cap \bar{M}(P)$), and a component that disagrees with $\mathbf{M}$ (in $\bar{M}(P)^\perp$). Intuitively, when testing for the validity of $\mathbf{P}$ while remaining confident on the correct specification of $\mathbf{M}$ we should employ tests that direct their power towards the subspace $\bar{T}(P)^\perp \cap \bar{M}(P)$ rather than all of $\bar{T}(P)^\perp$.

In the following, recall that $\Pi_{T^\perp}(\cdot)$ denotes the orthogonal projection under $\|\cdot\|_{P,2}$ onto $\bar{T}(P)^\perp$.

Lemma 3.2. *Let Assumption 2.1 hold, $\mathbf{P} \subseteq \mathbf{M}$, and $M(P)$ be linear.*

(i) Let Assumptions 3.1 and 3.3 hold with $S(P) \subseteq \bar{T}(P)^\perp \cap \bar{M}(P)$. Then: Theorem 3.3(i) remains valid for testing (19) for any path $t \mapsto P_{t,g} \in \mathbf{M}$ with $\Pi_S(g) \neq 0$.

(ii) Let Assumption 3.1 hold with $\mathbf{T} = \{1, \ldots, d\}$, $d < \infty$, and $\{s_\tau\}_{\tau=1}^d$ be an orthonormal basis for $\bar{S}(P) = \bar{T}(P)^\perp \cap \bar{M}(P)$. Then: For any asymptotic level α specification test ϕ_n for (19) with an asymptotic local power function, it follows

$$\inf_{g \in \mathcal{G}(\kappa)} \lim_{n \to \infty} \int \phi_n dP^n_{1/\sqrt{n},g} \leq \inf_{g \in \mathcal{G}(\kappa)} \lim_{n \to \infty} P_{1/\sqrt{n},g}\big(\|\hat{\mathbb{G}}_n\|^2 > c_{1-\alpha}\big), \tag{22}$$

where $\mathcal{G}(\kappa) \equiv \{g \in \bar{M}(P) : \|\Pi_{T^\perp}(g)\|_{P,2} \geq \kappa\}$, and $c_{1-\alpha}$ is the $(1-\alpha)$ quantile of a chi-square distribution with d degrees of freedom.

(iii) Let Assumption 3.2 hold with $\hat{\theta}_n$ and $\tilde{\theta}_n$ being efficient estimators of $\theta(P) \in \mathbf{B}$ under $\mathbf{P}$ and $\mathbf{M}$ respectively. Then: Theorem 3.3(ii) remains valid for testing (19) with $b^*(\nu - \tilde{\nu}) \in \bar{T}(P)^\perp \cap \bar{M}(P)$ for all $b^* \in \mathbf{B}^*$.

Lemma 3.2(i) revisits the tests examined in Theorem 3.3(i) under the additional requirement that the test focus its power on detecting deviation from $\mathbf{P}$ that remain within $\mathbf{M}$ (i.e. $\bar{S}(P) \subseteq \bar{T}(P)^\perp \cap \bar{M}(P)$) rather than arbitrary deviations from $\mathbf{P}$ (i.e. $\bar{S}(P) \subseteq \bar{T}(P)^\perp$). In order for the resulting test to be able to detect any local deviation of $\mathbf{P}$ that remains within $\mathbf{M}$, $\hat{\mathbb{G}}_n$ must be chosen so that $\bar{S}(P) = \bar{T}(P)^\perp \cap \bar{M}(P)$. When $\bar{T}(P)^\perp \cap \bar{M}(P)$ is finite dimensional, Lemma 3.2(ii) additionally provides a characterization of the optimal specification test in the sense of maximizing local minimum power against alternatives in $\mathbf{M} \setminus \mathbf{P}$ that are a "local distance" of κ away from $\mathbf{P}$. Specifically, the optimal test corresponds to a quadratic form in $\hat{\mathbb{G}}_n$ where $\hat{\mathbb{G}}_n$ must be chosen so that it weights every possible local deviation in $\mathbf{M} \setminus \mathbf{P}$ "equally" – i.e. $S(P) = \{s_\tau : \tau \in \mathbf{T}\}$ should be an orthonormal basis for $\bar{T}(P)^\perp \cap \bar{M}(P)$.

In parallel to our results in Section 3.3, multiple tests for (19) implicitly possess the structure of the tests described in Lemma 3.2(i)-(ii); see our GMM discussion below. Lemma 3.2(iii), for example, shows that a process $\hat{\mathbb{G}}_n$ satisfying the conditions of Lemma 3.2(i) may be obtained by comparing an estimator $\hat{\theta}_n$ that is efficient under $\mathbf{P}$ to an estimator $\tilde{\theta}_n$ that is efficient under the larger model $\mathbf{M}$. It is again helpful to note that $\hat{\theta}_n$ and $\tilde{\theta}_n$ can be regular estimators of any "smooth" function of $P \in \mathbf{P}$ and need not be of any structural parameter of the model $\mathbf{P}$. The resulting Hausman type test then satisfies the optimality claim in Lemma 3.2(ii) provided the influence function of $\sqrt{n}\{\hat{\theta}_n - \tilde{\theta}_n\}$ spans $\bar{T}(P)^\perp \cap \bar{M}(P)$; see Remark 3.4. Finally, we emphasize, as in Remark 3.5, that many alternatives to a Hausman type test also satisfy the conditions of Lemmas 3.2(i)-(ii). In fact, the sieve likelihood ratio test of Shen and Shi (2005) and Chen and Liao (2014) for semi/nonparametric likelihood models, and the sieve optimally weighed quasi likelihood ratio test of Chen and Pouzo (2009, 2015) for semi/nonparametric conditional moment restriction models can be regarded as versions of incremental J tests for (19). These incremental J tests are also applicable to testing hypotheses on structural parameters of a model $\mathbf{M}$, in which case $\mathbf{P}$ corresponds to the subset of distributions in $\mathbf{M}$ that satisfy the conjectured null hypothesis on the structural parameters.

GMM Illustration (cont.) For $\mathbf{P}$ as defined in (6), we now let $\rho(x,\gamma) = (\rho_1(x,\gamma)', \rho_2(x,\gamma)')'$ where $\rho_j : \mathbf{X} \times \mathbf{R}^{d_\gamma} \to \mathbf{R}^{d_{\rho_j}}$ with $d_{\rho_1} \geq d_\gamma$, and let

$$\mathbf{M} \equiv \Big\{P \in \mathcal{M} : \int \rho_1(\cdot,\gamma)dP = 0 \text{ for some } \gamma \in \Gamma\Big\}. \tag{23}$$

Eichenbaum et al. (1988) propose testing $\mathbf{P}$ with $\mathbf{M}$ as a maintained hypothesis by employing an incremental J statistic $J_n(\rho) - J_n(\rho_1)$, where $J_n(\rho)$ and $J_n(\rho_1)$ are the J statistics based on the moments ρ (for $\mathbf{P}$) and ρ_1 (for $\mathbf{M}$) respectively. As in Remark 3.5, it can be shown that for $\{p_k\}_{k=1}^{d_\rho - d_\gamma}$ and $\{m_k\}_{k=1}^{d_{\rho_1}-d_\gamma}$ orthonormal bases for $\bar{T}(P)^\perp$ and $\bar{M}(P)^\perp$ we have

$$\begin{aligned} J_n(\rho) - J_n(\rho_1) &= \sum_{k=1}^{d_\rho - d_\gamma}\Big(\frac{1}{\sqrt{n}}\sum_{i=1}^n p_k(X_i)\Big)^2 - \sum_{k=1}^{d_{\rho_1}-d_\gamma}\Big(\frac{1}{\sqrt{n}}\sum_{i=1}^n m_k(X_i)\Big)^2 + o_p(1) \\ &= \sum_{k=1}^{d_{\rho_2}}\Big(\frac{1}{\sqrt{n}}\sum_{i=1}^n f_k(X_i)\Big)^2 + o_p(1) \end{aligned} \tag{24}$$

where the second equality holds for $\{f_k\}_{k=1}^{d_{\rho_2}}$ an orthonormal basis for $\bar{T}(P)^\perp \cap \bar{M}(P)$ since $\bar{M}(P)^\perp \subseteq \bar{T}(P)^\perp$. Therefore, an incremental J test corresponds to a special case of the test discussed in Lemma 3.2(i) for which $\bar{S}(P) = \bar{T}(P)^\perp \cap \bar{M}(P)$. Moreover, by Lemma 3.2(ii), the resulting test is locally maximin optimal. Instead of the statistic $J_n(\rho) - J_n(\rho_1)$, an alternative approach employs the ρ_1 moments for efficient estimation of $\gamma(P)$ (under $\mathbf{M}$) and the remaining ρ_2 moments for testing; see, e.g., Christiano and Eichenbaum (1992), Hansen and Heckman (1996), and Hansen (2010). Such a test corresponds to the Hausman type test in Lemma 3.2(iii). Specifically, $\hat{\theta}_n = 0$ is an efficient estimator of $\theta(P) = \int \rho_2(\cdot, \gamma(P))dP$ under $\mathbf{P}$, while an efficient estimator $\tilde{\theta}_n$ of $\theta(P)$ under $\mathbf{M}$ equals

$$\frac{1}{n}\sum_{i=1}^{n}\{\rho_2(X_i, \hat{\gamma}_n) - \hat{B}_n' \rho_1(X_i, \hat{\gamma}_n)\} \tag{25}$$

for $\hat{\gamma}_n$ an efficient estimator of $\gamma(P)$ using ρ_1 moments (under $\mathbf{M}$), and $\hat{B}_n$ the OLS coefficients from regressing $\{\rho_2(X_i, \hat{\gamma}_n)\}_{i=1}^n$ on $\{\rho_1(X_i, \hat{\gamma}_n)\}_{i=1}^n$. By Lemma 3.2(ii), an (orthogonalized) quadratic form in (25) leads to a locally maximin optimal test that is asymptotically equivalent to (24). ■

4 General Nonparametric Conditional Moment Models

In this section we apply our previous results to a rich class of models defined by nonparametric conditional moment restrictions with possibly different conditioning sets and potential endogeneity.

4.1 Models and Characterizations

The data distribution P of $X = (Z, W) \in \mathbf{X}$ is assumed to satisfy the following nonparametric conditional moment restrictions

$$E[\rho_j(Z, h_P)|W_j] = 0 \text{ for all } 1 \leq j \leq J \quad \text{for some} \quad h_P \in \mathbf{H}, \tag{26}$$

for some known measurable mappings $\rho_j : \mathbf{Z} \times \mathbf{H} \to \mathbf{R}$, where $\mathbf{H}$ is some Banach space (with norm $\|\cdot\|_{\mathbf{H}}$) of measurable functions of $X = (Z, W)$. Here $Z \in \mathbf{Z}$ denotes potentially endogenous random variables, and $W \in \mathbf{W}$ denotes the union of distinct random elements of the conditioning variables (or instruments) $(W_1, \ldots, W_J)$. Note that there are no restrictions imposed on how the conditioning variables relate – e.g. W_j and $W_{j'}$ may have all, some, or no elements in common, and some of the W_j could be constants (indicating unconditional moment restrictions).

Model (26) encompasses a very wide array of semiparametric and nonparametric models. It was first studied in Ai and Chen (2007) for root-n consistent estimation of a particular "smooth" linear functional of h_P when the generalized residual functions ρ_j are pointwise differentiable (in h_P) for all $j = 1, ..., J$. Since Ai and Chen (2007) focused on possibly globally misspecified models, in that P may fail to satisfy (26), they did not characterize the tangent space.

In this section we characterize the tangent space for model (26) without assuming the differentiability of $\rho_j(Z, \cdot) : \mathbf{H} \to L^2(P)$ for all j. We assume instead that

$$m_j(W_j, h) \equiv E[\rho_j(Z, h)|W_j] \tag{27}$$

is "smooth" (at h_P) when viewed as a map from $\mathbf{H}$ into $L^2(W_j)$, where $L^2(W_j)$ is the subset of functions $f \in L^2(P)$ depending only on W_j. Specifically, we require Fréchet differentiability of each $m_j(W_j, \cdot) : \mathbf{H} \to L^2(W_j)$ (at h_P) and denote its derivative by $\nabla m_j(W_j, h_P)$, which could be computed as $\nabla m_j(W_j, h_P)[h] \equiv \frac{\partial}{\partial\tau} m_j(W_j, h_P + \tau h)|_{\tau=0}$ for any $h \in \mathbf{H}$. Employing these derivatives, we may then define the linear map $\nabla m(W, h_P) : \mathbf{H} \to \bigotimes_{j=1}^J L^2(W_j)$ to be given by

$$\nabla m(W, h_P)[h] \equiv \big(\nabla m_1(W_1, h_P)[h], \ldots, \nabla m_J(W_J, h_P)[h]\big)'. \tag{28}$$

Note that $\bigotimes_{j=1}^J L^2(W_j)$ is itself a Hilbert space when endowed with an inner product (and induced norm) equal to $\langle f, \tilde{f}\rangle \equiv \sum_{j=1}^J E[f_j(W_j)\tilde{f}_j(W_j)]$ for any $f = (f_1, \ldots, f_J)$ and $\tilde{f} = (\tilde{f}_1, \ldots, \tilde{f}_J)$. The range space $\mathcal{R}$ of the linear map $\nabla m(W, h_P)$ is then defined as

$$\mathcal{R} \equiv \Big\{ f \in \bigotimes_{j=1}^J L^2(W_j) : f = \nabla m(W, h_P)[h] \text{ for some } h \in \mathbf{H} \Big\}, \tag{29}$$

and we let $\bar{\mathcal{R}}$ be its closure (in $\bigotimes_{j=1}^J L^2(W_j)$), which is a vector space and plays an important role in this section. Finally, we set $\bar{\mathcal{R}}^\perp$ to be the orthocomplement of $\bar{\mathcal{R}}$ (in $\bigotimes_{j=1}^J L^2(W_j)$).

In order to be explicit about the local perturbations we consider, we next introduce a set of conditions on the paths $t \mapsto P_{t,g}$ employed to construct the tangent set $T(P)$.

Condition A. *(i) Under $X \sim P_{t,g}$, $E[\rho_j(Z, h_t)|W_j] = 0$ for some $h_t \in \mathbf{H}$ and all $1 \leq j \leq J$; (ii) $\|t^{-1}(h_t - h_P) - \Delta\|_{\mathbf{H}} = o(1)$ as $t \downarrow 0$ for some $\Delta \in \mathbf{H}$; (iii) $|\rho_j(z, h_t)| \leq F(z)$ for all $1 \leq j \leq J$ and $F \in L^2(P)$ satisfying $\int F^2 dP_{t,g} = O(1)$ as $t \downarrow 0$.*

Thus, a path $t \mapsto P_{t,g}$ satisfies Condition A if whenever X is distributed according to $P_{t,g}$, the conditional moment restrictions in (26) hold for some $h_t \in \mathbf{H}$ and the map $t \mapsto h_t$ is "smooth" in t. These requirements are satisfied, for instance, by the paths considered in semiparametric efficiency calculations, in which distributions are parametrized by h_t and a complementary infinite dimensional parameter describing aspects of the distribution not characterized by h_t; see, e.g., Begun et al. (1983), Hansen (1985), Chamberlain (1986, 1992), Newey (1990), and Ai and Chen (2012). We also introduce a vector space $\mathcal{V}$ given by

$$\mathcal{V} \equiv \Big\{ g = \sum_{j=1}^J \rho_j(Z, h_P)\psi_j(W_j) : (\psi_1, \ldots, \psi_J) \in \bigotimes_{j=1}^J L^2(W_j) \Big\}, \tag{30}$$

which is a subset of $L_0^2(P)$ provided P satisfies (26) and $E[\{\rho_j(Z, h_P)\}^2|W_j]$ is bounded P-a.s. for $1 \leq j \leq J$. Let $\bar{\mathcal{V}}$ be the closure of $\mathcal{V}$, and $\bar{\mathcal{V}}^\perp$ be the orthocomplement of $\bar{\mathcal{V}}$ (in $L_0^2(P)$).

Finally, we impose the following regularity conditions on the distribution P.

Assumption 4.1. *(i) P satisfies model (26); (ii) $m_j(W_j, \cdot) : \mathbf{H} \to L^2(W_j)$ is Fréchet differentiable at h_P for $1 \leq j \leq J$; (iii) $\rho_j(Z, \cdot) : \mathbf{H} \to L^2(P)$ is continuous at h_P for $1 \leq j \leq J$; (iv) There is a $\mathcal{D} \subseteq \mathbf{H}$ such that $\overline{lin}\{\mathcal{D}\} = \mathbf{H}$ and for every $h \in \mathcal{D}$ there is a $t \mapsto P_{t,g}$ satisfying Condition A with $\Delta = h$; (v) $\mathcal{V}^\perp$ has a dense subset of bounded functions.*

Assumption 4.2. *(i) $\sum_{j=1}^J E[\rho_j^2(Z, h_P)|W_j]$ is bounded P-a.s.; (ii) There is $C_0 < \infty$ such that $\sum_{j=1}^J \|\psi_j\|_{P,2} \leq C_0 \|\sum_{j=1}^J \rho_j(\cdot, h_P)\psi_j\|_{P,2}$ for all $(\psi_1, \ldots, \psi_J) \in \bigotimes_{j=1}^J L^2(W_j)$.*

Assumptions 4.1(i)(ii)(iii) and 4.2(i) are standard. Assumptions 4.1(iv)(v) and 4.2(ii) are sufficient conditions for the simple characterization of the tangent space obtained in Theorem 4.1

below. Assumption 4.1(iv) assumes that $\mathbf{H}$ is the local parameter space for h_P, while Assumption 4.1(v) assumes that any function $g \in \mathcal{V}^{\perp}$ can be approximated by sequences of bounded functions in $\mathcal{V}^{\perp}$ – low level sufficient conditions for this requirement are often readily available in specific applications. Assumption 4.2(ii) imposes a linear independence restriction on $\{\rho_j(Z, h_P)\}_{j=1}^J$.

Our next result provides a simple characterization for local overidentification.

Theorem 4.1. *Let P satisfy Assumptions 4.1 and 4.2. Then: $\bar{T}(P)^{\perp}$ satisfies*

$$\bar{T}(P)^{\perp} = \left\{ g \in L_0^2(P) : g = \sum_{j=1}^J \rho_j(Z, h_P)\psi_j(W_j) \text{ for some } (\psi_1, \ldots, \psi_J) \in \bar{\mathcal{R}}^{\perp} \right\},$$

and moreover $\bar{T}(P)^{\perp} = \{0\}$ if and only if $\bar{\mathcal{R}} = \bigotimes_{j=1}^J L^2(W_j)$.

In view of Lemma 2.1, Theorem 4.1 implies that P is locally just identified by a regular model (26) if and only if $\bar{\mathcal{R}} = \bigotimes_{j=1}^J L^2(W_j)$. Theorem 4.1 also has a useful dual representation.

Lemma 4.1. *Let Assumption 4.1(ii) hold. Let $\mathbf{H}^*$ be the dual space of a Banach space $\mathbf{H}$, and $\nabla m_j(W_j, h_P)^* : L^2(W_j) \to \mathbf{H}^*$ be the adjoint of $\nabla m_j(W_j, h_P) : \mathbf{H} \to L^2(W_j)$ for $j = 1, ..., J$. Then: $\bar{\mathcal{R}} = \bigotimes_{j=1}^J L^2(W_j)$ if and only if*

$$\left\{ f = (f_1, \ldots, f_J) \in \bigotimes_{j=1}^J L^2(W_j) : \sum_{j=1}^J \nabla m_j(W_j, h_P)^*[f_j] = 0 \right\} = \{0\}.$$

Theorem 4.1 and Lemma 4.1 together imply that P is locally just identified by model (26) if and only if the adjoint operator $\nabla m(W, h_P)^* : \bigotimes_{j=1}^J L^2(W_j) \to \mathbf{H}^*$ is injective. Interestingly, this resembles a necessary condition, the injectivity of $\nabla m(W, h_P) : \mathbf{H} \to \bigotimes_{j=1}^J L^2(W_j)$, for local identification of the unknown function h_P by model (26) in Chen et al. (2014) – when (26) is linear in h_P this condition is also sufficient (Newey and Powell, 2003).

4.1.1 Sequential moment restrictions

While the proof of Theorem 4.1 directly computes $\bar{T}(P)^{\perp}$, we note that in special cases of model (26) for which semiparametric efficiency bounds are known, one could also employ Corollary 3.1 to characterize local just identification. We next follow such an approach by employing the efficiency bound results in Ai and Chen (2012) to characterize the local just identification of P by models defined by sequential moment restrictions.

The data distribution P of $X = (Z, W) \in \mathbf{X}$ is now assumed to satisfy the following nonparametric sequential moment restrictions

$$\text{model (26) holds with } \sigma(\{W_j\}) \subseteq \sigma(\{W_{j'}\}) \text{ for all } 1 \leq j \leq j' \leq J, \tag{31}$$

where $\sigma(\{W_j\})$ denotes the σ-field generated by W_j for $j = 1, ..., J$. Note now $W = W_J$, which is assumed to be a non-degenerate random variable.

We will restrict attention to distributions P for which the conditional moments in (31) are suitably linearly independent. To this end, we define

$$s_j^2(W_J) \equiv \inf_{\{a_k\}_{k=j+1}^J} E[\{\rho_j(Z,h_P) - \sum_{k=j+1}^J a_k \rho_k(Z,h_P)\}^2 | W_J] \text{ for } j = 1, ..., J-1, \quad (32)$$

and $s_J^2(W_J) \equiv E[\{\rho_J(Z,h_P)\}^2|W_J]$. Since W_J is the most informative conditioning variable, we may interpret $s_j^2(W_J)$ as the residual variance obtained by projecting $\rho_j(Z,h_P)$ on $\{\rho_{j'}(Z,h_P)\}_{j'>j}$ conditionally on all instruments.

The following assumption imposes the basic condition on the distribution P.

Assumption 4.3. *(i) P satisfies (31); (ii) Assumption 4.1(ii) holds; (iii) $\max_j E[\{\rho_j(Z,h)\}^2] < \infty$ for any $h \in \mathbf{H}$ (a Banach space); (iv) $P(\eta \le E[s_j^2(W_J)|W_j]) = 1$ for some $\eta > 0$ and all $1 \le j \le J$; (v) $P(|E[\rho_k(Z,h_P)\rho_j(Z,h_P)|W_j]| \le M) = 1$ for some $M < \infty$ and all $1 \le k \le j \le J$; (vi) $L^2(W_J)$ is infinite dimensional.*

Assumptions 4.3(i)(ii)(iii) are standard. Assumption 4.3(iv) restricts the conditional dependence across moments, while Assumption 4.3(v) imposes an almost sure upper bound in the conditional covariance across residuals. When the same instrument is used in all conditioning equations, so that $W_j = W_J$ for all j, Assumptions 4.3(iv)(v) are equivalent to the covariance matrix of the residuals conditional on W_J being nonsingular and finite uniformly in the support of W_J. Finally, Assumption 4.3(vi) ensures that model (31) implies an infinite number of unconditional moment restrictions. If $L^2(W_J)$ is finite dimensional, then model (31) consists of a finite number of unconditional moment restrictions thus reducing to the well understood GMM setting.

Theorem 4.2. *Let Assumption 4.3 hold. Then: P is locally just identified by model (31) if and only if $\bar{\mathcal{R}} = \bigotimes_{j=1}^J L^2(W_j)$.*

It is interesting that the characterization of local just identification in nonparametric sequential moment restrictions (31) coincides with that for the more general model (26) derived in Theorem 4.1. Nevertheless, the additional structure afforded by sequential moment restrictions does allow for the semiparametric efficiency bound calculation in Ai and Chen (2012) and enables us to obtain the local just identification characterization under lower level conditions.

4.1.2 Models with triangular structures

Numerous nonparametric structural models possess a triangular structure in which the (conditional) moment restrictions depend on a non-decreasing subset of the parameters; see examples in Subsection 4.2. Lemma 4.2 below focuses on such a setting by assuming the parameter space takes the form $\mathbf{H} = \bigotimes_{j=1}^J \mathbf{H}_j$ and imposing that the moment conditions can be ordered in a manner such that the k-th moment condition depends only on the subset $\bigotimes_{j=1}^k \mathbf{H}_j$ of the parameter space. In the Lemma we let $\nabla m_{j,j}(W_j, h_P)^* : L^2(W_j) \to \mathbf{H}_j^*$ be the adjoint of $\nabla m_{j,j}(W_j, h_P) : \mathbf{H}_j \to L^2(W_j)$, and $\bar{\mathcal{R}}_j$ be the closure of $\mathcal{R}_j$ (in $L^2(W_j)$), where $\mathcal{R}_j$ is given by

$$\mathcal{R}_j \equiv \{f \in L^2(W_j) : f = \nabla m_{j,j}(W_j, h_P)[h_j] \text{ for } h_j \in \mathbf{H}_j\}.$$

Lemma 4.2. *Let Assumption 4.1(ii) hold, and* $\mathbf{H} = \bigotimes_{j=1}^{J} \mathbf{H}_j$ *with* $\mathbf{H}_j$ *being Banach spaces for all* j*. Suppose there are linear maps* $\nabla m_{j,k}(W_j, h_P) : \mathbf{H}_k \to L^2(W_j)$ *such that*

$$\nabla m_j(W_j, h_P)[h] = \sum_{k=1}^{J} \nabla m_{j,k}(W_j, h_P)[h_k] \quad \text{for any} \quad h = (h_1, \ldots, h_J) \in \bigotimes_{j=1}^{J} \mathbf{H}_j, \tag{33}$$

where $\nabla m_{j,k}(W_j, h_P)[h_k] = 0$ *for all* $k > j$*, and there is* $0 \leq C < \infty$ *such that*

$$\|\nabla m_{j,k}(\cdot, h_P)[h_k]\|_{P,2} \leq C \|\nabla m_{k,k}(\cdot, h_P)[h_k]\|_{P,2} \quad \text{for all } k \leq j. \tag{34}$$

Then: $\bar{\mathcal{R}} = \bigotimes_{j=1}^{J} L^2(W_j)$ *if and only if* $\bar{\mathcal{R}}_j = L^2(W_j)$ *for all* j*, which also holds if and only if* $\{f \in L^2(W_j) : \nabla m_{j,j}(W_j, h_P)^*[f] = 0\} = \{0\}$ *for all* j*.*

Lemma 4.2 implies that, under the stated requirements on the partial derivative maps, one may assess whether P is locally overidentified by examining each (conditional) moment restriction separately. This lemma simplifies the verification of local just identification in many nonparametric models. For example, it is directly applicable to the following class of models

$$E[\rho_j(Z, h_{P,j})|W_j] = 0 \quad \text{for some} \quad h_{P,j} \in \mathbf{H}_j \quad \text{for all } 1 \leq j \leq J \tag{35}$$

where $h_P = (h_{P,1}, \ldots, h_{P,J}) \in \mathbf{H} = \bigotimes_{j=1}^{J} \mathbf{H}_j$, and the unknown functions $h_{P,j} \in \mathbf{H}_j$ could depend on the endogenous variables Z. Our final result applies Lemma 4.2 to special cases of model (35) in which $\mathbf{H}_j$ contains functions of conditioning variables W_j only; i.e.,

$$E[\rho_j(Z, h_{P,j}(W_j))|W_j] = 0 \quad \text{for some} \quad h_{P,j} \in \mathbf{H}_j \subseteq L^2(W_j) \quad \text{for all } 1 \leq j \leq J. \tag{36}$$

Corollary 4.1. *Let* P *satisfy model (36) with* $h_P = (h_{P,1}, \ldots, h_{P,J}) \in \mathbf{H} = \bigotimes_{j=1}^{J} \mathbf{H}_j$ *and Assumption 4.1(ii) hold. Suppose for each* $1 \leq j \leq J$*, there is* $d_j \in L^2(W_j)$ *that is bounded* P*-a.s. and* $\nabla m_j(W_j, h_P)[h] = d_j(W_j) h_j(W_j)$ *for any* $h = (h_1, \ldots, h_J) \in \mathbf{H}$*. Then:* $\bar{\mathcal{R}} = \bigotimes_{j=1}^{J} L^2(W_j)$ *if and only if for all* $1 \leq j \leq J$*,* $\mathbf{H}_j$ *is dense in* $L^2(W_j)$ *and* $P(d_j(W_j) \neq 0) = 1$*.*

Corollary 4.1 reduces assessing local just identification of P by model (36) to examining two simple conditions for all $j = 1, ..., J$: (i) $\mathbf{H}_j$ must be sufficiently "rich" ($\mathbf{H}_j$ is dense in $L^2(W_j)$), and (ii) The derivative of the moment restrictions must be injective ($d_j(W_j) \neq 0$ P-a.s.). It immediately implies, for example, that nonparametric conditional mean and quantile regression models are locally just identified,[5] and that restricting the parameter space to the space of bounded or differentiable functions is not sufficient for yielding local overidentification as $\mathbf{H}_j$ remains dense in $L^2(W_j)$. On the other hand, Corollary 4.1 does imply that P will be locally overidentified by model (36) as soon as there is one j such that $\mathbf{H}_j$ is not a dense subset of $(L^2(W_j), \|\cdot\|_{P,2})$. Examples in which $\mathbf{H}_j$ is not dense include, among others, the partially linear or additively separable conditional mean specifications of Robinson (1988) and Stone (1985).

Remark 4.1. Semiparametric two-step GMM models are widely used in applied work. Building on the insights of Newey (1994) and Newey and Powell (1999), Ackerberg et al. (2014) show that

[5] For $Z = (Y, W)$ with $Y \in \mathbf{R}$ note that a mean regression model corresponds to $\rho_1(Z, h) = Y - h(W)$ so that $d_1(W) = -1$. Instead, in a quantile regression model $\rho_1(Z, h) = \tau - 1\{Y \leq h(W)\}$, in which case $d_1(W) = -g_{Y|W}(h_P(W)|W)$ for $g_{Y|W}(y|w)$ the conditional density of Y given W.

when the unknown function $h_P = (h_{P,1}, \ldots, h_{P,J})$ is "exactly identified" by model (36) in the first stage, the second stage optimally weighted GMM estimator of γ_P identified by unconditional moment restriction $E[g(X, \gamma_P, h_P)] = 0$ is semiparametrically efficient. Our Corollary 4.1 shows that their requirement of nonparametric "exact identification" of h_P is equivalent to our P being locally just identified by model (36) in the first stage. Our Theorem 4.1, Lemmas 4.1 and 4.2 further imply, however, that the second stage optimally weighted GMM estimator may be inefficient when P is locally overidentified by model (26) in the first stage, such as in the various semiparametric conditional moment restriction models of Ai and Chen (2003, 2012). See Subsection 4.2 for examples of P being locally overidentified by nonparametric models. ■

4.2 Illustrative Examples

This section presents three empirically relevant examples to illustrate the implications of our results; see the Online Appendix for additional results and discussion.

Example 4.1. (**Differentiated Products Markets**) An extensive literature has studied identification of demand and cost functions in differentiated product markets, including the seminal work of Berry et al. (1995). Here, we follow Berry and Haile (2014) who derive multiple identification results by relying on moment restrictions of the form

$$E[Y_{ij} - h_{j,P}(V_i)|W_{ij}] = 0 \text{ for } 1 \leq j \leq J \tag{37}$$

where $1 \leq i \leq n$ denotes a market and $1 \leq j \leq J$ a good. For instance, in their analysis of demand, $h_{j,P}$ corresponds to the inverse demand function for good j, V_i denotes market shares and prices in market i, Y_{ij} is a "demand shifter", and W_{ij} is a vector of price instruments and product/market characteristics for good j. Letting $h_{j,P} \in \mathbf{H}_j \subseteq L^2(V)$ for all j, and $h_P = (h_{1,P}, \ldots, h_{J,P}) \in \mathbf{H} = \bigotimes_{j=1}^J \mathbf{H}_j$, we then note that this model is a special case of model (35). We may therefore apply Lemma 4.2, and to this end we observe that for any $h = (h_1, \ldots, h_J) \in \mathbf{H}$ we have

$$\nabla m_j(W, h_P)[h] = -E[h_j(V)|W_j], \tag{38}$$

let $\bar{\mathbf{H}}_j$ be the closure of $\mathbf{H}_j$ under $\|\cdot\|_{P,2}$, and for any $f \in L^2(W_j)$ we define

$$\Pi_{\bar{\mathbf{H}}_j} f \equiv \arg\min_{h_j \in \bar{\mathbf{H}}_j} \|(-f) - h_j\|_{P,2}. \tag{39}$$

The map $\Pi_{\bar{\mathbf{H}}_j} : L^2(W_j) \to \bar{\mathbf{H}}_j$ is the adjoint of (38), and Lemma 4.2 implies that P is locally just identified if and only if

$$\{f \in L^2(W_j) : \Pi_{\bar{\mathbf{H}}_j} f = 0\} = \{0\} \text{ for all } 1 \leq j \leq J. \tag{40}$$

For instance, if $\bar{\mathbf{H}}_j = L^2(V)$, then (40) is equivalent to the distribution of (V, W_j) being L^2-complete with respect to V for all j (Newey and Powell, 2003), which is an untestable condition under endogeneity (Andrews, 2017; Canay et al., 2013).[6] Hence, plug-in estimation of average

[6]Since there are examples of distributions for which L^2-completeness fails (Santos, 2012), the model may be locally overidentified even when V and W_j are of equal dimension.

derivatives may not be efficient when L^2-completeness fails (Ai and Chen, 2012). Finally, we note that the structure in model (37) is also present in a large literature on consumer demand; see, for example, Blundell et al. (1998, 2003, 2007). Semiparametric restrictions that are consistent with agents' optimization behaviors, however, can render P locally overidentified (Blundell et al., 2007; Chen and Pouzo, 2009). ■

Example 4.2. (**Nonparametric Selection**) This example concerns nonparametric versions of the canonical selection model of Heckman (1979) as studied in, e.g., Heckman (1990). Suppose that for each individual i, there are latent variables $(Y^*_{0,i}, Y^*_{1,i})$ satisfying

$$Y^*_{d,i} = g_{d,P}(V_i) + U_{d,i} \tag{41}$$

where $d \in \{0,1\}$, V_i is a set of regressors, and $g_{d,P}$ are unknown functions. Instead of $(Y^*_{0,i}, Y^*_{1,i})$, we observe $Y_i = Y^*_{0,i} + D_i(Y^*_{1,i} - Y^*_{0,i})$ where $D_i \in \{0,1\}$ indicates selection into "treatment". As in Heckman and Vytlacil (2005), we assume there exists a variable R_i excluded from $g_{d,P}$ and impose the index sufficiency requirement

$$E[U_{d,i}|V_i, R_i, D_i = d] = \lambda_{d,P}(P(D_i = 1|V_i, R_i)) \tag{42}$$

for unknown functions $\lambda_{d,P}$. Assuming $E[U_{d,i}|V_i] = 0$ for $d \in \{0,1\}$, we can then employ equations (41) and (42) to obtain the system of conditional moment restrictions

$$E[D_i - s_P(V_i, R_i)|V_i, R_i] = 0 \tag{43}$$

$$E[Y_i - g_{d,P}(V_i) - \lambda_{d,P}(s_P(V_i, R_i))|V_i, R_i, D_i = d] = 0, \tag{44}$$

which can be used to identify the conditional average treatment effect $g_{1,P}(V_i) - g_{0,P}(V_i)$; see also Newey et al. (1999) and Das et al. (2003) for related models. Hence, in this context $J = 2$, $h_P = (g_{0,P}, g_{1,P}, \lambda_{0,P}, \lambda_{1,P}, s_P)$, $W_{i1} = (V_i, R_i)$, and $W_{i2} = (V_i, R_i, D_i)$.

We examine a general nonparametric version of this model by only requiring $g_{d,P} \in L^2(V)$ and $\lambda_{d,P}$ be continuously differentiable for $d \in \{0,1\}$. For any $(g_0, g_1, \lambda_0, \lambda_1, s) \in \mathbf{H}$, restrictions (43) and (44) then possess a sequential moment structure which simplifies applying Theorems 4.1 or 4.2. In particular, Lemma 4.2 implies P is locally just identified if and only if

$$\mathcal{S}_d \equiv \{f \in L^2((V,R)) : f(V,R) = g_d(V) + \lambda_d(s_P(V,R)) \text{ for some } g_d, \lambda_d\} \tag{45}$$

is dense in $L^2(V,R)$ for $d \in \{0,1\}$. However, identification of the functions $g_{d,P}$ and $\lambda_{d,P}$ requires

$$P(\text{Var}\{s_P(V,R)|V\} > 0) > 0, \tag{46}$$

i.e. the instrument R must be relevant. When (46) holds, $\mathcal{S}_d$ is not dense in $L^2(V,R)$. Thus, the conditions for the identification of $(g_{d,P}, \lambda_{d,P})$ imply that P is locally overidentified by the model. Hence, the model is testable and efficiency considerations matter when estimating smooth parameters such as the average treatment effects. ■

Example 4.3. (**Nonparametric Production**) This example closely follows the firm's production structural models proposed by Olley and Pakes (1996), Ackerberg et al. (2015) and others.

Econometricians observe a random sample $\{X_i\}_{i=1}^n$ of a panel of firms $i = 1, ..., n$ from the distribution of $X = \{Y_t, K_t, L_t, I_t\}_{t=1}^T$ for a fixed finite $T \geq 2$, where Y_t, K_t, L_t, I_t respectively denotes a firm's log output, capital, labor, and investment levels at time t. Suppose that

$$Y_{it} = g_P(K_{it}, L_{it}) + \omega_{it} + U_{it} \qquad E[U_{it}|K_{it}, L_{it}, I_{it}] = 0, \tag{47}$$

where g_P is an unknown function, and ω_{it} is a productivity factor observed by the firm but not the econometrician. Olley and Pakes (1996) provides conditions under which the firm's dynamic optimization problem implies, for some unknown function λ_P, that

$$\omega_{it} = \lambda_P(K_{it}, I_{it}). \tag{48}$$

Let $W = (K_1, L_1, I_1)$, and for simplicity let $T = 2$ and ω_{it} follow an AR(1) process. The literature has employed (47) and (48) to derive the semiparametric conditional moment restrictions

$$E[Y_1 - g_P(K_1, L_1) - \lambda_P(K_1, I_1)|W] = 0 \tag{49}$$

$$E[Y_2 - g_P(K_2, L_2) - \pi_P \lambda_P(K_1, I_1)|W] = 0 \tag{50}$$

where π_P is the coefficient in the AR(1) process for ω_{it}. This model contains multiple overidentifying restrictions that are easily characterized through Theorem 4.1. Specifically, note $h_P = (g_P, \lambda_P, \pi_P)$ and for any $h = (g, \lambda, \pi)$ we have

$$\nabla m_1(W, h_P)[h] = -g(K_1, L_1) - \lambda(K_1, I_1) \tag{51}$$

$$\nabla m_2(W, h_P)[h] = -E[g(K_2, L_2)|W] - \pi_P \lambda(K_1, I_1) - \pi \lambda_P(K_1, I_1). \tag{52}$$

By Theorem 4.1 a necessary condition for local just identification is for the closure of the range of (51) to equal $L^2(W)$. However, this requirement fails since (51) cannot approximate nonseparable functions $f \in L^2((L_1, I_1))$ – a failure reflecting the assumption that labor is not a dynamic variable. Consequently, sequential estimation of average output elasticities, as in Olley and Pakes (1996), can be inefficient. Similarly, we note

$$\nabla m_2(W, h_P)[h] - \pi_P \nabla m_1(W, h_P)[h] = \pi_P g(K_1, L_1) - E[g(K_2, L_2)|W] - \pi \lambda_P(K_1, I_1), \tag{53}$$

and local just identification requires the closure of the range of (53) to equal $L^2(W)$. However, such a condition can fail reflecting the empirical content of assuming constancy of g_P through time and additive separability of ω_{it}. As in Section 3.4, the power of specification tests can be directed at violations of these assumptions. ■

4.3 Numerical Illustration

We provide a brief numerical illustration based on Example 4.3. As a design, we let g_P be a Cobb-Douglas production function and for simplicity suppose

$$K_{it+1} = 0.9K_{it} + I_{it} \qquad I_{it} = \exp\{-0.1 \times \log(K_{it}) + \omega_{it}\}, \tag{54}$$

where ω_{it} follows an AR(1) process with coefficient $\pi_P = 0.5$ and normally distributed innovations with variance $(0.3)^2$. The variables U_{it} in (47) are drawn from a normal distribution with mean zero and variance $(0.1)^2$, and we assume the firm sets L_{it} to maximize expected (over U_{it}) profits when facing wages V_{it} with $\log(V_{it}) \sim N(0, (0.3)^2)$. Our sample is generated by selecting two observations after a "burn in" period of a thousand time periods for each firm. The results reported below are based on one thousand replications of samples of five thousand observations each.

Since we impose a Cobb-Douglas specification, the moment restrictions become

$$E[Y_{i1} - \alpha_P \log(K_{i1}) - \beta_P \log(L_{i1}) - \lambda_P(K_{i1}, I_{i1})|W_i] = 0 \tag{55}$$

$$E[Y_{i2} - \alpha_P \log(K_{i2}) - \beta_P \log(L_{i2}) - \pi_P \lambda_P(K_{i1}, I_{i1})|W_i] = 0 \tag{56}$$

and we let $\mathbf{P}$ denote the set of distributions for which (55) and (56) hold for some $(\alpha_P, \beta_P, \pi_P, \lambda_P)$. Following Section 3.4, we conduct a specification test that aims its power at deviations from $\mathbf{P}$ for which labor is a dynamic variable and thus affects investment decisions. While it is possible to construct tests against any such deviations, for illustrative purposes it is convenient to focus on deviations satisfying for some unknown $\gamma_P \in \mathbf{R}$ and unknown function $\Phi_P : \mathbf{R}^2 \mapsto \mathbf{R}$,

$$I_{it} = \Phi_P(\omega_{it} - \gamma_P \log(L_{it}), K_{it}), \tag{57}$$

and we let $\mathbf{M}$ denote such set of distributions. Importantly, we note the tangent spaces of P relative to $\mathbf{P}$ and $\mathbf{M}$ differ by one dimension (i.e., $\bar{T}(P)^{\perp} \cap \bar{M}(P)$ has dimension one).

Table 1: Performance of Estimators

		Mean		N×Variance	
Parameter	Value	**P**-Efficient	**M**-Efficient	**P**-Efficient	**M**-Efficient
π	0.5	0.500	0.499	0.854	0.887
α	0.7	0.698	0.698	0.094	0.177
β	0.3	0.307	0.306	0.071	2.115

We estimate $(\alpha_P, \beta_P, \pi_P)$ efficiently under $\mathbf{P}$ and $\mathbf{M}$ employing the estimator in Ai and Chen (2003).[7] Table 1 reports the performance of such estimators and we see, in accord with the theory, that local overidentification of P makes efficiency considerations relevant. Since $\bar{T}(P)^{\perp} \cap \bar{M}(P)$ has dimension one, Lemma 3.2(ii) implies the influence function of the difference between estimators that are efficient under $\mathbf{P}$ and estimators that are efficient under $\mathbf{M}$ does not depend (up to scale) on the parameter being estimated. Indeed, letting $(\hat{\alpha}_{\mathbf{P}}, \hat{\beta}_{\mathbf{P}})$ and $(\hat{\alpha}_{\mathbf{M}}, \hat{\beta}_{\mathbf{M}})$ denote efficient estimators under $\mathbf{P}$ and $\mathbf{M}$, we find the correlation between $\{\hat{\beta}_{\mathbf{P}} - \hat{\beta}_{\mathbf{M}}\}$ and $\{\hat{\alpha}_{\mathbf{P}} - \hat{\alpha}_{\mathbf{M}}\}$ to be -0.998.

The fact that the dimension of $\bar{T}(P)^{\perp} \cap \bar{M}(P)$ equals one, further implies that all specification tests that direct power at this subspace possess the same local power function. In order to examine this prediction, we consider alternatives for which (57) holds with $\Phi_P(a, b) = \exp\{a - 0.1 \log(b)\}$. Table 2 reports the power curves for Hausman tests based on estimates of α and of β that direct

[7] In the implementation, we employ a sieve for λ_P consisting of five terms and built using BSplines of order two. The conditional expectations are estimated via series using nine terms consisting of BSplines of order two and the cross products $\log(L_{i1}) \log(K_{i1})$ and $\log(L_{i1}) \log(I_{I1})$.

Table 2: Power Functions of Specification Tests

Test	Value of Deviation γ								
	-0.12	-0.09	-0.06	-0.03	0.000	0.03	0.06	0.09	0.12
HT α	0.999	0.994	0.830	0.256	0.042	0.270	0.734	0.955	0.999
HT β	1.000	0.994	0.832	0.262	0.044	0.285	0.738	0.956	0.999
QLR	1.000	0.987	0.752	0.190	0.045	0.306	0.755	0.961	0.998

their power at $\bar{T}(P)^{\perp} \cap \bar{M}(P)$. We also employ the sieve quasi likelihood ratio test in Chen and Pouzo (2015) to test whether $\gamma = 0$ in (57), which also directs its power at $\bar{T}(P)^{\perp} \cap \bar{M}(P)$. We find similar power curves, with the different Hausman tests having virtually identical power.

5 Extension to $T(P)$ Being a Convex Cone

Our main theoretical results in Section 3 rely on the requirement that the tangent set $T(P)$ be linear. In this Section, we examine to what extent our main conclusions apply to models in which $T(P)$ is a convex cone – a setting that can arise, for example, in mixture models (van der Vaart, 1989) and models where a parameter is on a boundary (Andrews, 1999). To this end, we replace Assumption 2.1 with the following weaker condition:

Assumption 5.1. *(i) Assumption 2.1(i) holds; (ii) The tangent set $T(P)$ is a convex cone – i.e. if $g, f \in T(P)$, $a, b \in \mathbf{R}$ with $a \geq 0$ and $b \geq 0$, then $ag + bf \in T(P)$.*

We let $\bar{T}(P)$ still denote the closure of $T(P)$ under $\|\cdot\|_{P,2}$ and maintain Definition 2.2. Crucially, Assumption 5.1(ii) implies $\bar{T}(P)$ is a closed convex cone but not necessarily a closed linear subspace of $L_0^2(P)$ as in regular models. Thus, the alternative characterization of local overidentification in terms of the orthogonal complement of $\bar{T}(P)$ is no longer valid (see Lemma 2.1). However, for any closed convex cone $\bar{T}(P)$ in $L_0^2(P)$, we may define its polar cone, denoted $\bar{T}(P)^{-}$, which is given by

$$\bar{T}(P)^{-} \equiv \left\{ g \in L_0^2(P) : \int gf dP \leq 0 \text{ for all } f \in \bar{T}(P) \right\}. \tag{58}$$

Let $\Pi_T(g)$ and $\Pi_{T^-}(g)$ denote the metric projections of a $g \in L_0^2(P)$ onto $\bar{T}(P)$ and $\bar{T}(P)^{-}$ respectively. For any $g \in L_0^2(P)$, the so called "Moreau decomposition" (Moreau, 1962) implies

$$g = \Pi_T(g) + \Pi_{T^-}(g), \qquad \int \{\Pi_T(g)\}\{\Pi_{T^-}(g)\} dP = 0. \tag{59}$$

Unlike the setting in which $\bar{T}(P)$ is a linear subspace, however, there may in fact exist $f \in \bar{T}(P)$ and $g \in \bar{T}(P)^{-}$ such that $\int fg dP < 0$. Nevertheless, the decomposition in (59) immediately implies the following direct generalization of Lemma 2.1.

Lemma 5.1. *Under Assumption 5.1, the following are equivalent to Definition 2.2:*

(i) P is locally just identified by $\mathbf{P}$ if and only if $\bar{T}(P)^{-} = \{0\}$.

(ii) P is locally overidentified by $\mathbf{P}$ if and only if $\bar{T}(P)^{-} \neq \{0\}$.

By definition it is clear that Theorem 3.1 remains valid for the case that P is locally just identified by $\mathbf{P}$ (i.e.,$\bar{T}(P) = L_0^2(P)$). Given Lemma 5.1, it should also be possible to establish results similar to Theorem 3.2 for the locally overidentified case. That is, if P is locally overidentified by $\mathbf{P}$, then the model should be locally testable and "efficiency" should "matter" even when $T(P)$ is a convex cone. To gain some intuition, we can again rely on the sample means of scores $0 \neq \tilde{f} \in L_0^2(P)$. Recall that if $X_i \sim P_{1/\sqrt{n},g}$ for any path $t \mapsto P_{t,g} \in \mathcal{M}$, then $\mathbb{G}_n(\tilde{f}) \equiv \frac{1}{\sqrt{n}} \sum_{i=1}^n \tilde{f}(X_i)$ is asymptotically normally distributed with mean $\int \tilde{f} g dP$ (see equation (12)). By Lemma 5.1, P being locally overidentified by $\mathbf{P}$ is equivalent to the existence of a $0 \neq \tilde{f} \in \bar{T}(P)^-$. For any such $\tilde{f}$, it follows that $\int \tilde{f} g dP \leq 0$ for all $g \in \bar{T}(P)$. Thus, for the purposes of specification testing, observing a large and positive value for $\mathbb{G}_n(\tilde{f})$ may be viewed as a "signal" that the distribution of $X_i \sim P_{1/\sqrt{n},g}$ is approaching P from outside the model $\mathbf{P}$. On the other hand, from an estimation perspective, we should be able to employ the knowledge that $\int \tilde{f} g dP \leq 0$ for all $g \in \bar{T}(P)$ to improve on "inefficient" estimators.

The potential lack of orthogonality between $\bar{T}(P)$ and $\bar{T}(P)^-$, however, presents some important complications. For instance, it is no longer natural to restrict attention to regular estimators. We instead focus on a broader class of estimators for parameter $\theta(P) \in \mathbf{B}$ satisfying

$$\sqrt{n}\{\hat{\theta}_n - \theta(P_{1/\sqrt{n},g})\} \stackrel{L_{n,g}}{\to} \mathbb{Z}_g \tag{60}$$

for some tight random variable $\mathbb{Z}_g \in \mathbf{B}$ along any path $t \mapsto P_{t,g} \in \mathbf{P}$. Note that in contrast to regular estimators, the limit $\mathbb{Z}_g$ may depend on g. Focusing on estimators satisfying (60) enables us to easily characterize the local asymptotic risk along any path $t \mapsto P_{t,g} \in \mathbf{P}$. Concretely, for a loss function $\Psi : \mathbf{B} \to \mathbf{R}_+$, the local asymptotic risk of $\hat{\theta}_n$ is given by

$$\limsup_{n\to\infty} E_{P_{1/\sqrt{n},g}}\big[\Psi\big(\sqrt{n}\{\hat{\theta}_n - \theta(P_{1/\sqrt{n},g})\}\big)\big], \tag{61}$$

which represents the expected loss of employing $\hat{\theta}_n$ to estimate $\theta(P)$ when the data generating process is locally perturbed within $\mathbf{P}$. For simplicity we consider $\Psi - loss$ functions, defined as

Definition 5.1. *Ψ-loss is a map from $\mathbf{B}$ to $\mathbf{R}_+$ such that: (i) $\{b \in \mathbf{B} : \Psi(b) \leq t\}$ is convex for all $t \in \mathbf{R}$; (ii) $\Psi(0) = 0$ and $\Psi(b) = \Psi(-b)$; (iii) Ψ is bounded, continuous, and nonconstant.*

A minimal requirement on an estimator is that its local asymptotic risk not be dominated by that of an alternative estimator – i.e. a sensible estimator should be "asymptotically locally admissible".

Definition 5.2. *$\hat{\theta}_n : \{X_i\}_{i=1}^n \to \mathbf{B}$ is "asymptotically locally admissible" for $\theta(P)$ under Ψ-loss if it satisfies (60) and there is no estimator $\tilde{\theta}_n : \{X_i\}_{i=1}^n \to \mathbf{B}$ satisfying (60) and*

$$\limsup_{n\to\infty} E_{P_{1/\sqrt{n},g}}\big[\Psi\big(\sqrt{n}\{\tilde{\theta}_n - \theta(P_{1/\sqrt{n},g})\}\big)\big] \leq \limsup_{n\to\infty} E_{P_{1/\sqrt{n},g}}\big[\Psi\big(\sqrt{n}\{\hat{\theta}_n - \theta(P_{1/\sqrt{n},g})\}\big)\big]$$

for all paths $t \mapsto P_{t,g} \in \mathbf{P}$, and with the inequality holding strictly for some path $t \mapsto P_{t,g} \in \mathbf{P}$.

Given the introduced concepts, we can document an equivalence result between the local overidentification of P, the importance of "efficiency" in estimation, and the potential refutability of a model.

Theorem 5.1. *Let Assumption 5.1 hold. Then the following statements are equivalent:*

(i) P is locally overidentified by $\mathbf{P}$.

(ii) There exists a bounded function $f : \mathbf{X} \to \mathbf{R}$ such that $\sum_{i=1}^n f(X_i)/n$ is not an asymptotically locally admissible estimator for $\theta(P) = \int f dP$ under any Ψ-loss.

(iii) There exists a local asymptotic level α test ϕ_n for (9) with a local asymptotic power function π satisfying $\pi(g) > \alpha$ for some path $t \mapsto P_{t,g} \in \mathcal{M} \setminus \mathbf{P}$.

Theorem 3.2 and Theorem 5.1 reflect both the similarities and the differences between regular and non-regular models. With regards to estimation, for example, Theorems 3.2(i)(ii) and 5.1(i)(ii) both show that local overidentification of P is equivalent to the availability of "efficiency" gains in estimation. However, since in non-regular models we need to consider a broader class of estimators than just regular estimators, Theorem 5.1(i)(ii) links the availability of "efficiency" gains to the local overidentification of P through the estimation of simple "smooth" maps $\theta(P) = \int f dP$ (population means) for bounded functions f. In particular, while sample means are always locally admissible when P is locally just identified (see Lemma C.1 in Appendix C), Theorem 5.1(ii) shows this fails to be the case when P is locally overidentified.

With regards to specification testing, Theorem 3.2(i)(iii) and Theorem 5.1(i)(iii) both show that local overidentification of P is equivalent to the potential refutability of the model. However, important differences also exist in the properties of local specification tests for regular and non-regular models. Notably, our next result shows that for any non-regular model whose convex cone $\bar{T}(P)$ contains at least two linearly independent elements, any asymptotically locally unbiased specification test for (9) will have local power no larger than its level against any alternative.

Theorem 5.2. *Let Assumption 5.1 hold and there be linearly independent $f_1, f_2 \in \bar{T}(P)^-$ with $\lambda f_1, \lambda f_2 \in \bar{T}(P)$ for any $\lambda \leq 0$. Let ϕ_n be any specification test for (9) with a local asymptotic power function π such that $\pi(g) \leq \alpha$ for all $g \in \bar{T}(P)$ and $\pi(g) \geq \alpha$ for all $g \notin \bar{T}(P)$. Then: $\pi(g) = \alpha$ for any path $t \mapsto P_{t,g} \in \mathcal{M} \setminus \mathbf{P}$ with $\lambda \Pi_{T^-}(g) \in \bar{T}(P)$ for any $\lambda \leq 0$.*

Given Theorem 5.1(iii), Theorem 5.2 does not preclude the existence of asymptotically nontrivial specification tests, but rather implies such tests can necessarily be asymptotically locally *biased* for non-regular models. We next examine in more detail the construction of both such specification tests and of "better" estimators than the sample mean. To this end, we impose the following:

Assumption 5.2. *For some set $\mathbf{T}$ there is a statistic $\hat{\mathbb{G}}_n : \{X_i\}_{i=1}^n \to \ell^\infty(\mathbf{T})$ satisfying:*
(i) $\hat{\mathbb{G}}_n(\tau) = \frac{1}{\sqrt{n}} \sum_{i=1}^n s_\tau(X_i) + o_p(1)$ uniformly in $\tau \in \mathbf{T}$, where $0 \neq s_\tau \in \bar{T}(P)^-$ for all $\tau \in \mathbf{T}$;
(ii) Assumption 3.1(ii) holds.

Assumption 5.2(i) is identical to Assumption 3.1(i) except that s_τ is required to belong to $\bar{T}(P)^-$ instead of $\bar{T}(P)^\perp$. As in Theorem 3.3(i), $\hat{\mathbb{G}}_n$ can be employed to construct a specification test for (9). For any $0 \leq \omega \in \ell^\infty(\mathbf{T})$, we define $\hat{\mathbb{G}}_n^\omega(\tau) \equiv \omega(\tau) \times \hat{\mathbb{G}}_n(\tau)$ and $\mathbb{G}_0^\omega(\tau) \equiv \omega(\tau) \times \mathbb{G}_0(\tau)$ for $\tau \in \mathbf{T}$. Let $c_{1-\alpha}^\omega$ be the $1 - \alpha$ quantile of $\|\max\{\mathbb{G}_0^\omega, 0\}\|_\infty$. We then define the test

$$\phi_n^\omega \equiv 1\{\|\max\{\hat{\mathbb{G}}_n^\omega, 0\}\|_\infty > c_{1-\alpha}^\omega\}. \tag{62}$$

Intuitively, $0 \leq \omega \in \ell^\infty(\mathbf{T})$ is a weight function that determines the local alternatives against which ϕ_n^ω has nontrivial power. In parallel to Theorem 3.3(i), the power properties of ϕ_n^ω also depend on the set $C(P) \equiv \{s_\tau \in \bar{T}(P)^- : \tau \in \mathbf{T}\}$ being sufficiently "rich". Let $\bar{C}(P)$ denote the closed convex cone generated by $C(P)$ (in $L_0^2(P)$) – i.e. $\bar{C}(P)$ parallels $\bar{S}(P)$ in Theorem 3.3(i). For any $g \in L_0^2(P)$ we let $\Pi_C(g)$ denote the metric projection of g onto $\bar{C}(P)$.

Our next result shows that for any path $t \mapsto P_{t,g} \in \mathcal{M}$ with $\Pi_C(g) \neq 0$ it is possible to select an $\omega^\star \in \ell^\infty(\mathbf{T})$ such that the corresponding specification test $\phi_n^{\omega^\star}$ has nontrivial local power against that alternative. Given Theorem 5.2, $\phi_n^{\omega^\star}$ can be asymptotically locally *biased*, however.

Theorem 5.3. *Let Assumptions 5.1, 5.2 hold, and $0 \leq \omega \in \ell^\infty(\mathbf{T})$ satisfy $c_{1-\alpha}^\omega > 0$. Then: ϕ_n^ω is a local asymptotic level α test for (9) with a local asymptotic power function. Moreover, for any $t \mapsto P_{t,g} \in \mathcal{M}$ with $\Pi_C(g) \neq 0$ there is $0 \leq \omega^\star \in \ell^\infty(\mathbf{T})$ with $c_{1-\alpha}^{\omega^\star} > 0$ for $\alpha \in (0, \frac{1}{2})$, and*

$$\lim_{n\to\infty} P_{1/\sqrt{n},g}\big(\|\max\{\hat{\mathbb{G}}_n^{\omega^\star}, 0\}\|_\infty > c_{1-\alpha}^{\omega^\star}\big) > \alpha. \tag{63}$$

Turning to estimation, we note that when restricting attention to paths $t \mapsto P_{t,g} \in \mathbf{P}$, knowledge that $\int s_\tau g dP \leq 0$ should be useful. Specifically, for any bounded function $f : \mathbf{X} \to \mathbf{R}$, we define

$$\hat{\mu}_n(f,\tau) \equiv \frac{1}{n}\sum_{i=1}^n f(X_i) - \beta(f,\tau) \times n^{-1/2} \max\{\hat{\mathbb{G}}_n(\tau), 0\} \tag{64}$$

for $\beta(f,\tau) \equiv \max\{\int f s_\tau dP, 0\}/\|s_\tau\|_{P,2}^2$. The function $\beta(f,\tau)s_\tau$ is the projection of f onto the cone generated by $s_\tau \in \bar{T}(P)^-$ in $L_0^2(P)$. Our final Theorem shows that when P is locally overidentified, $\hat{\mu}_n(f,\tau)$ can be viewed as a more "efficient" estimator for $\theta(P) = \int f dP$ than the sample mean. It is analogous to Lemma 3.1(i).

Theorem 5.4. *Let Assumptions 5.1 and 5.2 hold. Then: for any bounded $f : \mathbf{X} \to \mathbf{R}$ satisfying $\int f s_{\tau^\star} dP > 0$ for some $\tau^\star \in \mathbf{T}$, we have: $\hat{\mu}_n(f,\tau^\star)$ defined in (64) satisfies (60) and*

$$\limsup_{n\to\infty} E_{P_{1/\sqrt{n},g}}\big[\Psi\big(\sqrt{n}\{\hat{\mu}_n(f,\tau^\star) - \int f dP_{1/\sqrt{n},g}\}\big)\big] < \limsup_{n\to\infty} E_{P_{1/\sqrt{n},g}}\big[\Psi\big(\frac{1}{\sqrt{n}}\sum_{i=1}^n\{f(X_i) - \int f dP_{1/\sqrt{n},g}\}\big)\big] \tag{65}$$

for any path $t \mapsto P_{t,g} \in \mathbf{P}$ and any Ψ-loss.

Thus, when P is locally overidentified, there is information in the model that can be employed to both render the model testable (Theorem 5.3) and to obtain "efficiency" gains (Theorem 5.4). As a result, the local testability of a model and "efficiency" considerations remain intrinsically linked to P being locally overidentified when $\bar{T}(P)$ is a convex cone. We emphasize that many important issues, such as optimality in estimation and specification testing, and analog of incremental J test for (19), remain open when $\bar{T}(P)$ is a convex cone. We leave these questions for future research.

6 Conclusion

This paper reinterprets the common practice of counting the numbers of restrictions and parameters of interest in GMM to determine overidentification as an approach that examines whether the tangent space is a strict subset of $L_0^2(P)$. This abstraction naturally leads to a notion of local overidentification, which we show is responsible for an intrinsic link between efficiency considerations in estimation and the local testability of a model. While we have relied on an i.i.d. assumption for simplicity, there are ample works deriving efficiency bounds in time series settings (Hansen, 1985, 1993) and characterizing limit experiments under nonstationary, strongly dependent data (Ploberger and Phillips, 2012). We conjecture the results in this paper could similarly be extended to allow for dependence, but leave such extensions for future work.

References

Ackerberg, D., Chen, X., Hahn, J. and Liao, Z. (2014). Asymptotic efficiency of semiparametric two-step gmm. *Review of Economic Studies*, **81** 919–943.

Ackerberg, D. A., Caves, K. and Frazer, G. (2015). Identification properties of recent production function estimators. *Econometrica*, **83** 2411–2451.

Ai, C. and Chen, X. (2003). Efficient estimation of models with conditional moment restrictions containing unknown functions. *Econometrica*, **71** 1795–1844.

Ai, C. and Chen, X. (2007). Estimation of possibly misspecified semiparametric conditional moment restriction models with different conditioning variables. *Journal of Econometrics*, **141** 5–42.

Ai, C. and Chen, X. (2012). The semiparametric efficiency bound for models of sequential moment restrictions containing unknown functions. *Journal of Econometrics*, **170** 442–457.

Aliprantis, C. D. and Border, K. C. (2006). *Infinite Dimensional Analysis – A Hitchhiker's Guide*. Springer-Verlag, Berlin.

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

Andrews, D. W. K. (1999). Estimation When a Parameter is on a Boundary. *Econometrica*, **67** 1341–1383.

Andrews, D. W. K. (2017). Examples of l2-complete and boundedly-complete distributions. *Journal of Econometrics*, **199** 213–220.

Arellano, M. (2003). *Panel data econometrics*. Oxford University Press.

Begun, J. M., Hall, W., Huang, W.-M. and Wellner, J. A. (1983). Information and asymptotic efficiency in parametric-nonparametric models. *The Annals of Statistics*, **11** 432–452.

Berry, S., Levinsohn, J. and Pakes, A. (1995). Automobile prices in market equilibrium. *Econometrica*, **63** 841–890.

Berry, S. T. and Haile, P. A. (2014). Identification in differentiated products markets using market level data. *Econometrica*, **82** 1749–1797.

Bickel, P., Klaassen, C., Ritov, Y. and Wellner, J. (1993). *Efficient and Adaptive Estimation for Semiparametric Models*. Johns Hopkins University Press.

Blundell, R., Browning, M. and Crawford, I. (2003). Nonparametric engel curves and revealed preference. *Econometrica*, **71** 205–240.

Blundell, R., Chen, X. and Kristensen, D. (2007). Semi-nonparametric iv estimation of shape-invariant engel curves. *Econometrica*, **75** 1613–1669.

Blundell, R., Duncan, A. and Pendakur, K. (1998). Semiparametric estimation and consumer demand. *Journal of Applied Econometrics*, **13** 435–461.

Blundell, R. and Powell, J. (2003). Endogeneity in nonparametric and semiparametric regression models. *Advances in Economics and Econometrics: Theory and Applications, Eighth World Congress, ed. M Dewatripont, LP Hansen, SJ Turnovsky* 312–357.

Bogachev, V. I. (1998). *Gaussian Measures.* American Mathematical Society, Providence.

Bogachev, V. I. (2007). *Measure Theory: Volume II.* Springer-Verlag, Berlin.

Canay, I. A., Santos, A. and Shaikh, A. M. (2013). On the testability of identification in some nonparametric models with endogeneity. *Econometrica*, **81** 2535–2559.

Chamberlain, G. (1986). Asymptotic efficiency in semiparametric models with censoring. *Journal of Econometrics*, **32** 189–218.

Chamberlain, G. (1992). Efficiency bounds for semiparametric regression. *Econometrica*, **60** 567–596.

Chen, X., Chernozhukov, V., Lee, S. and Newey, W. K. (2014). Local identification of nonparametric and semiparametric models. *Econometrica*, **82** 785–809.

Chen, X. and Liao, Z. (2014). Sieve m inference on irregular parameters. *Journal of Econometrics*, **182** 70–86.

Chen, X. and Pouzo, D. (2009). Efficient estimation of semiparametric conditional moment models with possibly nonsmooth residuals. *Journal of Econometrics*, **152** 46–60.

Chen, X. and Pouzo, D. (2015). Sieve wald and qlr inferences on semi/nonparametric conditional moment models. *Econometrica*, **83** 1013–1079.

Choi, S., Hall, W. and Schick, A. (1996). Asymptotically uniformly most powerful tests in parametric and semiparametric models. *The Annals of Statistics*, **24** 841–861.

Christiano, L. J. and Eichenbaum, M. (1992). Current real-business-cycle theories and aggregate labor-market fluctuations. *The American Economic Review*, **82** 430–450.

Das, M., Newey, W. K. and Vella, F. (2003). Nonparametric estimation of sample selection models. *The Review of Economic Studies*, **70** 33–58.

Davydov, Y. A., Lifshits, M. A. and Smorodina, N. V. (1998). *Local Properties of Distributions of Stochastic Functionals.* American Mathematical Society, Providence.

Eichenbaum, M. S., Hansen, L. P. and Singleton, K. J. (1988). A time series analysis of representative agent models of consumption and leisure choice under uncertainty. *The Quarterly Journal of Economics*, **103** 51–78.

Haile, P. and Tamer, E. (2003). Inference With an Incomplete Model of English Auctions. *Journal of Political Economy*, **111** 1–51.

Hájek, J. (1970). A characterization of limiting distributions of regular estimates. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, **14** 323–330.

Hansen, L. (2014). Nobel Lecture: Uncertainty Outside and Inside Economic Models. *Journal of Political Economy*, **122** 945–987.

Hansen, L. and Jagannathan, R. (1997). Assessing Specification Errors in Stochastic Discount Factor Models. *Journal of Finance*, **52** 557–590.

Hansen, L. P. (1982). Large Sample Properties of Generalized Method of Moments Estimators. *Econometrica*, **50** 1029–1054.

Hansen, L. P. (1985). A method for calculating bounds on the asymptotic covariance matrices of generalized method of moments estimators. *Journal of Econometrics*, **30** 203–238.

Hansen, L. P. (1993). Semiparametric efficiency bounds for linear time-series models. In *Models, Methods and Applications of Econometrics: Essays in Honor of A. R. Bergstrom* (P. C. B. Phillips, ed.). Blackwell.

Hansen, L. P. (2010). Generalized method of moments estimation. In *Macroeconometrics and Time Series Analysis*. Springer, 105–118.

Hansen, L. P. and Heckman, J. J. (1996). The empirical foundations of calibration. *The Journal of Economic Perspectives*, **10** 87–104.

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

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica*, **47** 153–161.

Heckman, J. J. (1990). Varieties of selection bias. *The American Economic Review*, **80** 313–318.

Heckman, J. J. and Vytlacil, E. (2005). Structural equations, treatment effects, and econometric policy evaluation1. *Econometrica*, **73** 669–738.

Hirano, K. and Porter, J. R. (2009). Asymptotics for statistical treatment rules. *Econometrica*, **77** 1683–1701.

Koopmans, T. C. and Riersol, O. (1950). The identification of structural characteristics. *The Annals of Mathematical Statistics*, **21** 165–181.

LeCam, L. (1986). *Asymptotic Methods in Statistical Decision Theory.* Springer-Verlag, New York.

Lehmann, E. and Romano, J. (2005). *Testing Statistical Hypotheses.* Springer Verlag.

Lehmann, E. L. (1952). Testing multiparameter hypotheses. *The Annals of Mathematical Statistics*, **23** 541–552.

Luenberger, D. (1969). *Optimization by Vector Space Methods.* John Wiley & Sons.

Manski, C. F. (2003). *Partial Identification of Probability Distributions.* Springer-Verlag, New York.

Moreau, J.-J. (1962). Décomposition orthogonale dun espace hilbertien selon deux cônes mutuellement polaires. *CR Acad. Sci. Paris*, **255** 238–240.

Murphy, S. and van der Vaart, A. W. (1997). Semiparametric likelihood ratio inference. *The Annals of Statistics*, **25** 1471–1509.

Newey, W. K. (1990). Semiparametric efficiency bounds. *Journal of Applied Econometrics*, **5** 99–135.

Newey, W. K. (1994). The asymptotic variance of semiparametric estimators. *Econometrica*, **62** 1349–1382.

Newey, W. K. and Powell, J. (2003). Instrumental variables estimation of nonparametric models. *Econometrica*, **71** 1565–1578.

Newey, W. K. and Powell, J. L. (1999). Two-step estimation, optimal moment conditions, and sample selection models. Working Paper, MIT.

Newey, W. K., Powell, J. L. and Vella, F. (1999). Nonparametric estimation of triangular simultaneous equations models. *Econometrica*, **67** 565–603.

Olley, G. S. and Pakes, A. (1996). The dynamics of productivity in the telecommunications equipment industry. *Econometrica*, **64** 1263–1297.

Parente, P. M. and Smith, R. J. (2011). Gel methods for nonsmooth moment indicators. *Econometric Theory*, **27** 74–113.

Ploberger, W. and Phillips, P. C. (2012). Optimal estimation under nonstandard conditions. *Journal of Econometrics*, **169** 258–265.

Robinson, P. M. (1988). Root-n-consistent semiparametric regression. *Econometrica*, **56** 931–954.

Santos, A. (2012). Inference in nonparametric instrumental variables with partial identification. *Econometrica*, **80** 213–275.

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

Shen, X. (1997). On methods of sieves and penalization. *The Annals of Statistics*, **25** 2555–2591.

Shen, X. and Shi, J. (2005). Sieve likelihood ratio inference on general parameter space. *Science in China Ser. A Mathematics*, **48** 67–78.

Stone, C. J. (1985). Additive regression and other nonparametric models. *The annals of Statistics*, **13** 689–705.

van der Vaart, A. (1989). On the asymptotic information bound. *The Annals of Statistics*, **17** 1487–1500.

van der Vaart, A. (1991a). An asymptotic representation theorem. *International Statistical Review*, **59** 97–121.

van der Vaart, A. (1991b). On Differentiable Functionals. *The Annals of Statistics*, **19** 178–204.

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

van der Vaart, A. and Wellner, J. (1996). *Weak Convergence and Empirical Processes*. Springer Verlag.

Zeidler, E. (1984). *Nonlinear Functional Analysis and its Applications III*. Springer-Verlag, New York.

Appendix A - Limiting Experiment

In this Appendix we embed specification tests and regular estimators in a common statistical experiment that highlights their connection to each other and to the local overidentification of P. The main result in this Appendix, Theorem A.1 below, plays an important role in the proofs of our main results in Section 3, and is therefore presented here for completeness. The proof of Theorem A.1 can be found in the Online Appendix.

Heuristically, in an asymptotic framework that is local to P, our parameter uncertainty is over what "direction" P is being approached from. We may intuitively interpret such a direction as the score g of $P_{1/\sqrt{n},g}$ and represent our parameter uncertainty as possessing only a "noisy" measure of g. Let $d_T \equiv \dim\{\bar{T}(P)\}$ and $d_{T^\perp} \equiv \dim\{\bar{T}(P)^\perp\}$ denote the (possibly infinite) dimensions of the tangent space and its orthogonal complement. Both $\bar{T}(P)$ and $\bar{T}(P)^\perp$ are Hilbert spaces with norm $\|\cdot\|_{P,2}$ and hence there exist orthonormal bases $\{\psi_k^T\}_{k=1}^{d_T}$ and $\{\psi_k^{T^\perp}\}_{k=1}^{d_{T^\perp}}$ for $\bar{T}(P)$ and $\bar{T}(P)^\perp$ respectively. We then consider a random variable $(\mathbb{Y}^T, \mathbb{Y}^{T^\perp}) \in \mathbf{R}^{d_T} \times \mathbf{R}^{d_{T^\perp}}$ whose law is such that the vectors $\mathbb{Y}^T \equiv (\mathbb{Y}_1^T, \ldots, \mathbb{Y}_{d_T}^T)'$ and $\mathbb{Y}^{T^\perp} \equiv (\mathbb{Y}_1^{T^\perp}, \ldots, \mathbb{Y}_{d_{T^\perp}}^{T^\perp})'$ have mutually independent coordinates and satisfy for some (unknown) $g_0 \in L_0^2(P)$ the relation

$$\mathbb{Y}_k^T \sim N\Big(\int g_0 \psi_k^T dP,\ 1\Big) \quad \text{for} \quad 1 \le k \le d_T$$
$$\mathbb{Y}_k^{T^\perp} \sim N\Big(\int g_0 \psi_k^{T^\perp} dP,\ 1\Big) \quad \text{for} \quad 1 \le k \le d_{T^\perp}. \tag{A.1}$$

Here, if $d_{T^\perp} = 0$, then we interpret $\mathbb{Y}^{T^\perp}$ as being equal to zero with probability one. Finally, we let Q_g denote the distribution of $(\mathbb{Y}^T, \mathbb{Y}^{T^\perp}) \in \mathbf{R}^{d_T} \times \mathbf{R}^{d_{T^\perp}}$ when (A.1) holds with $g_0 = g \in L_0^2(P)$.

Thus, by definition, we know that the (unknown) distribution Q_{g_0} of $(\mathbb{Y}^T, \mathbb{Y}^{T^\perp})$ belongs to the nonparametric family $\{Q_g : g \in L_0^2(P)\}$.

The following theorem formalizes the connection between specification tests, regular estimators, and the tangent space through the introduced limiting experiment.[8] Recall that $\stackrel{L}{\to}$ means convergence in law under $P^n \equiv \bigotimes_{i=1}^n P$.

Theorem A.1. *Under Assumption 2.1, the following two propositions hold:*

(i) Let ϕ_n be any local asymptotic level α specification test for $\mathbf{P}$ with a local asymptotic power function π. Then: there is a level α test $\phi : (\mathbb{Y}^T, \mathbb{Y}^{T^\perp}) \to [0,1]$ of

$$H_0 : \Pi_{T^\perp}(g_0) = 0 \qquad H_1 : \Pi_{T^\perp}(g_0) \neq 0 \tag{A.2}$$

based on one observation $(\mathbb{Y}^T, \mathbb{Y}^{T^\perp})$ such that $\pi(g_0) = \int \phi dQ_{g_0}$ for all $g_0 \in L_0^2(P)$.

(ii) (Convolution Theorem) Let $\hat{\theta}_n : \{X_i\}_{i=1}^n \to \mathbf{B}$ be any asymptotically linear regular estimator of any parameter $\theta(P) \in \mathbf{B}$. Then: for any $b^* \in \mathbf{B}^*$ there exist linear maps $F^T : \mathbf{R}^{d_T} \to \mathbf{R}$ and $F^{T^\perp} : \mathbf{R}^{d_{T^\perp}} \to \mathbf{R}$ such that under the law P^n it follows

$$\sqrt{n}\{b^*(\hat{\theta}_n) - b^*(\theta(P))\} \stackrel{L}{\to} F^T(\mathbb{Y}^T) + F^{T^\perp}(\mathbb{Y}^{T^\perp}), \tag{A.3}$$

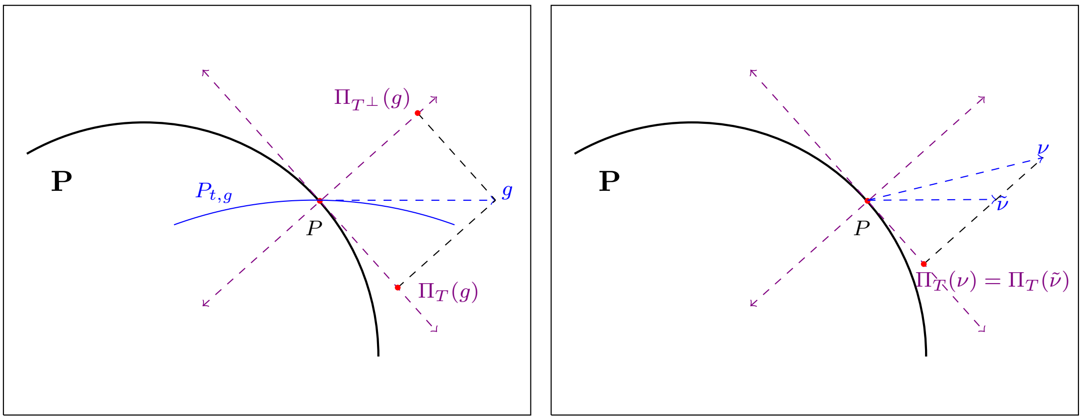
where $(\mathbb{Y}^T, \mathbb{Y}^{T^\perp}) \sim Q_{g_0}$ with $g_0 = 0$, and the map $F^T : \mathbf{R}^{d_T} \to \mathbf{R}$ depends on $\theta : \mathbf{P} \to \mathbf{B}$ and $b^* \in \mathbf{B}^*$ but not on the estimator $\hat{\theta}_n$.

Theorem A.1(i) relates the local properties of any specification test for $\mathbf{P}$ to a testing problem concerning the unknown distribution Q_{g_0} of $(\mathbb{Y}^T, \mathbb{Y}^{T^\perp})$ through the asymptotic representation theorem (van der Vaart, 1991a). Intuitively, any path $t \mapsto P_{t,g}$ that approaches P from outside the model $\mathbf{P}$ should be such that its score does not belong to the tangent set or, equivalently, $\Pi_{T^\perp}(g) \neq 0$; see Figure 1(a). In contrast, the score g of any submodel $t \mapsto P_{t,g} \in \mathbf{P}$ must belong to the tangent set, implying $\Pi_{T^\perp}(g) = 0$. Thus, any specification test for $\mathbf{P}$ should behave locally as a test of the null hypothesis in (A.2). Theorem A.1(i) formalizes these heuristics by showing that if π is the local asymptotic power function of a specification test for $\mathbf{P}$ (as in (9)), then π must also be the power function of a test of (A.2) based on a single observation $(\mathbb{Y}^T, \mathbb{Y}^{T^\perp})$ whose (unknown) law Q_{g_0} is known to belong to the nonparametric family $\{Q_g : g \in L_0^2(P)\}$.

Theorem A.1(ii) is essentially the convolution theorem of Hájek (1970), stated here in a manner that facilitates a connection to Theorem A.1(i). To gain intuition on this result, we focus on the scalar case ($\mathbf{B} = \mathbf{R}$) and suppose there are two asymptotically linear regular estimators $\hat{\theta}_n$ and $\tilde{\theta}_n$ of a common parameter with influence functions ν and $\tilde{\nu}$ respectively. Regularity constrains the projection of ν and $\tilde{\nu}$ onto the tangent space to be equal, which originates a term in the asymptotic distribution that is independent of the choice of estimator ($F^T(\mathbb{Y}^T)$); see Figure 1(b). The estimators, however, may differ on a component that is extraneous to the model ($\Pi_{T^\perp}(\nu) \neq \Pi_{T^\perp}(\tilde{\nu})$), contributing a "noise" term to the asymptotic distribution that depends on the choice of estimator ($F^{T^\perp}(\mathbb{Y}^{T^\perp})$). An efficient estimator is the one for which the "noise" component is zero.

[8] See Choi et al. (1996) and Hirano and Porter (2009) for other applications of the limit experiment.

Figure 1: The Tangent Space, Specification Tests, and Regular Estimators



(a) Intuition for Theorem A.1(i) (b) Intuition for Theorem A.1(ii)

Crucial for our purposes, is the observation that $\mathbb{Y}^{T^\perp}$ plays fundamental yet distinct roles in the asymptotic behavior of both specification tests and regular estimators. From a specification testing perspective, $\mathbb{Y}^{T^\perp}$ is a partially sufficient statistic for $\Pi_{T^\perp}(g)$ and is needed to construct any nontrivial test of (A.2). In contrast, from a regular estimation perspective, $\mathbb{Y}^{T^\perp}$ is an ancillary statistic that can only contribute "noise" to estimators.[9] Thus, the limit experiment requires P to be locally overidentified ($\bar{T}(P)^\perp \neq \{0\}$) in order to allow for both nontrivial tests and asymptotically distinct estimators.

Appendix B - Proofs for Sections 2 and 3

In this Appendix we present proofs of the theoretical results in Sections 2 and 3. All of the additional technical lemmas used in this Appendix can be found in the Online Appendix.

Proof of Lemma 2.1: Since $T(P)$ is linear by Assumption 2.1(ii), $\bar{T}(P)$ is a vector subspace of $L_0^2(P)$, and therefore $L_0^2(P) = \bar{T}(P) \oplus \bar{T}(P)^\perp$; see, e.g., Theorem 3.4.1 in Luenberger (1969). The claims of the Lemma then immediately follow from $\bar{T}(P) = L_0^2(P)$ if and only if $\bar{T}(P)^\perp = \{0\}$. ■

Proof of Theorem 3.1: To establish part (i) of the Theorem we let ν and $\tilde{\nu}$ denote the influence functions of $\hat{\theta}_n$ and $\tilde{\theta}_n$ respectively. Then note for any $b^* \in \mathbf{B}^*$ and $\lambda \in \mathbf{R}$

$$\sqrt{n}\{b^*(\lambda\hat{\theta}_n + (1-\lambda)\tilde{\theta}_n) - b^*(\theta(P))\}$$
$$= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\{\lambda b^*(\nu(X_i)) + (1-\lambda)b^*(\tilde{\nu}(X_i))\} + o_p(1) \xrightarrow{L} N(0,\sigma_\lambda^2) \quad \text{(B.1)}$$

for $\sigma_\lambda^2 = \|b^*(\lambda\nu + (1-\lambda)\tilde{\nu})\|_{P,2}^2$ by asymptotic linearity and the central limit theorem. Further note that if P is locally just identified, then Theorem A.1(ii) implies σ_λ^2 does not depend on λ. However, since $\|b^*(\lambda\nu + (1-\lambda)\tilde{\nu})\|_{P,2}^2$ being constant in λ implies that $\|b^*(\nu - \tilde{\nu})\|_{P,2} = 0$, and

[9] In regular estimation, only paths within the model are considered; see Definition 3.1. The resulting limiting experiment is then indexed by $\{Q_g : g \in \bar{T}(P)\}$, in which $\mathbb{Y}^{T^\perp}$ is ancillary.

$b^* \in \mathbf{B}^*$ was arbitrary, we can conclude that

$$b^*(\sqrt{n}\{\hat{\theta}_n - \tilde{\theta}_n\}) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} b^*(\nu(X_i) - \tilde{\nu}(X_i)) + o_p(1) = o_p(1) \tag{B.2}$$

for any $b^* \in \mathbf{B}^*$. Since $\sqrt{n}\{\hat{\theta}_n - \tilde{\theta}_n\}$ is asymptotically tight and measurable by Lemmas 1.4.3 and 1.4.4 in van der Vaart and Wellner (1996), result (B.2) and Lemma E.1 (in the Online Appendix) imply $\sqrt{n}\{\hat{\theta}_n - \tilde{\theta}_n\} = o_p(1)$ in $\mathbf{B}$, which establishes part (i) of the Theorem.

To establish part (ii) of the Theorem, we note that by Theorem A.1(i), there exists a level α test ϕ of (A.2) such that for any $g \in L_0^2(P)$ and path $t \mapsto P_{t,g}$

$$\lim_{n\to\infty} \int \phi_n dP^n_{1/\sqrt{n},g} = \int \phi dQ_g. \tag{B.3}$$

However, if P is locally just identified by $\mathbf{P}$, then $\bar{T}(P) = L_0^2(P)$, or equivalently $\bar{T}(P)^{\perp} = \{0\}$. Therefore, the null hypothesis in (A.2) holds for all $g \in L_0^2(P)$, which implies $\int \phi dQ_g \leq \alpha$ for all $g \in L_0^2(P)$, and part (ii) of the Theorem holds by (B.3). ■

Proof of Theorem 3.2: First, by Theorem 3.1, it follows that (ii) implies (i) and that (iii) implies (i). Therefore, it suffices to show that (i) (i.e., P being locally overidentified by $\mathbf{P}$) implies both (ii) and (iii) hold. To this end, we observe that if P is locally overidentified by $\mathbf{P}$, then Lemma 2.1 implies there exists a $0 \neq \tilde{f} \in \bar{T}(P)^{\perp}$, which without loss of generality we assume satisfies $\|\tilde{f}\|_{P,2} = 1$. We next aim to employ such a $\tilde{f}$ to verify that (ii) and (iii) indeed hold.

To establish that (i) implies (ii), we note that $\mathbb{G}_n(\tilde{f}) \equiv \sum_{i=1}^n \tilde{f}(X_i)/\sqrt{n}$ trivially satisfies Assumption 3.1(i) and Assumption 3.1(ii) with $\mathbb{G}_0 \sim N(0,1)$ since $\|\tilde{f}\|_{P,2} = 1$ and $\tilde{f} \in L_0^2(P)$. Thus, part (i) implying part (ii) is a special case of Lemma 3.1(i).

To establish that (i) implies (iii), we let $\mathbb{Z} \sim N(0,1)$ and note that Theorem 3.10.12 in van der Vaart and Wellner (1996) implies that for any path $t \mapsto P_{t,g} \in \mathcal{M}$

$$\mathbb{G}_n(\tilde{f}) \stackrel{L_{n,g}}{\to} \mathbb{Z} + \int \tilde{f} g dP \tag{B.4}$$

since $\|\tilde{f}\|_{P,2} = 1$. For $z_{1-\alpha/2}$ the $(1-\alpha/2)$ quantile of a standard normal distribution, we define the test $\phi_n \equiv 1\{|\mathbb{G}_n(\tilde{f})| > z_{1-\alpha/2}\}$. Then (B.4) and the Portmanteau Theorem imply that

$$\pi(g) \equiv \lim_{n\to\infty} \int \phi_n dP_{1/\sqrt{n},g} = P\left(\left|\mathbb{Z} + \int \tilde{f} g dP\right| > z_{1-\alpha/2}\right) \tag{B.5}$$

for any path $t \mapsto P_{t,g} \in \mathcal{M}$. Hence, (B.5) implies ϕ_n indeed has a local asymptotic power function. Moreover, since $\tilde{f} \in \bar{T}(P)^{\perp}$, result (B.5) implies $\pi(g) = \alpha$ whenever $g \in \bar{T}(P)$, which establishes ϕ_n is a local asymptotic level α specification test. In addition, for any $g \in \bar{T}(P)^{\perp}$ we have either $\int \tilde{f} g dP = 0$ (and hence $\pi(g) = \alpha$ by (B.5)), or $\int \tilde{f} g dP \neq 0$ (and hence $\pi(\tilde{f}) > \alpha$ by (B.5)). Thus this test is locally unbiased. Finally, there exists a path $t \mapsto P_{t,\tilde{f}} \in \mathcal{M}$ with score $\tilde{f} \in \bar{T}(P)^{\perp}$, in which case (B.5) implies $\pi(\tilde{f}) > \alpha$ and hence (i) implies (iii). ■

Proof of Corollary 3.1: First note that since every $f \in \mathcal{D}$ is bounded, $\theta_f(P) \equiv \int f dP$ is pathwise differentiable at P relative to $T(P)$ with derivative $\dot{\theta}_f(g) \equiv \int \Pi_T(f) g dP$; see Lemma F.1 (in the

Online Appendix). Therefore, by Theorem 5.2.1 in Bickel et al. (1993) its efficiency bound is given by $\Omega_f^* = \|\Pi_T(f)\|_{P,2}^2$. For any $f \in L^2(P)$ let $\Pi_{L_0^2(P)}(f)$ denote its projection onto $L_0^2(P)$ and note that $\Pi_{L_0^2(P)}(f) = \{f - \int f dP\}$, and hence $\text{Var}\{f(X)\} = \|\Pi_{L_0^2(P)}(f)\|_{P,2}^2$. By orthogonality of $\bar{T}(P)$ and $\bar{T}(P)^\perp$, then

$$\begin{aligned}\text{Var}\{f(X)\} = \|\Pi_{L_0^2(P)}(f)\|_{P,2}^2 &= \|\Pi_T(\Pi_{L_0^2(P)}(f)) + \Pi_{T^\perp}(\Pi_{L_0^2(P)}(f))\|_{P,2}^2 \\ &= \|\Pi_T(\Pi_{L_0^2(P)}(f))\|_{P,2}^2 + \|\Pi_{T^\perp}(\Pi_{L_0^2(P)}(f))\|_{P,2}^2 = \Omega_f^* + \|\Pi_{T^\perp}(f)\|_{P,2}^2, \end{aligned} \quad \text{(B.6)}$$

where in the final equality we used $\Pi_T(\Pi_{L_0^2(P)}(f)) = \Pi_T(f)$ and $\Pi_{T^\perp}(\Pi_{L_0^2(P)}(f)) = \Pi_{T^\perp}(f)$ for any $f \in L^2(P)$ due to $\bar{T}(P)$ and $\bar{T}(P)^\perp$ being subspaces of $L_0^2(P)$. Thus, by (B.6) $\text{Var}\{f(X)\} = \Omega_f^*$ for all $f \in \mathcal{D}$ if and only if $\Pi_{T^\perp}(f) = 0$ for all $f \in \mathcal{D}$, which by denseness of $\mathcal{D}$ is equivalent to $\bar{T}(P)^\perp = \{0\}$. ■

Proof of Lemma 3.1: For part (i) of the Lemma, note that since $\mathbb{G}_0$ is nondegenerate, Assumption 3.1(i) implies $s_{\tau^*} \neq 0$ for some $\tau^* \in \mathbf{T}$, and for a $0 \neq \tilde{b} \in \mathbf{B}$ we set

$$\tilde{\theta}_n \equiv \hat{\theta}_n + \tilde{b} \times n^{-1/2}\hat{\mathbb{G}}_n(\tau^*). \quad \text{(B.7)}$$

Notice $\hat{\theta}_n$ is asymptotically linear by hypothesis and denote its influence function by ν. Assumption 3.1(i), definition (B.7), and the continuous mapping theorem then yield

$$\sqrt{n}\{\tilde{\theta}_n - \theta(P)\} = \frac{1}{\sqrt{n}}\sum_{i=1}^n \{\nu(X_i) + \tilde{b} \times s_{\tau^*}(X_i)\} + o_p(1). \quad \text{(B.8)}$$

Setting $\tilde{\nu}(X_i) \equiv \nu(X_i) + \tilde{b} \times s_{\tau^*}(X_i)$, we obtain for any $b^* \in \mathbf{B}^*$ that $b^*(\tilde{\nu}) = \{b^*(\nu) + b^*(\tilde{b}) \times s_{\tau^*}\} \in L_0^2(P)$ since $b^*(\nu) \in L_0^2(P)$ due to $\hat{\theta}_n$ being asymptotically linear and $s_{\tau^*} \in \bar{T}(P)^\perp \subseteq L_0^2(P)$ by Assumption 3.1(i). Hence, (B.8) implies $\tilde{\theta}_n$ is indeed asymptotically linear and its influence function equals $\tilde{\nu}$. Moreover, by Lemma D.4 (in the Online Appendix), $(\sqrt{n}\{\hat{\theta}_n - \theta(P)\}, \frac{1}{\sqrt{n}}\sum_{i=1}^n s_{\tau^*}(X_i))$ converge jointly in distribution in $\mathbf{B} \times \mathbf{R}$ under P^n, and hence the continuous mapping theorem implies

$$\sqrt{n}\{\tilde{\theta}_n - \theta(P)\} = \sqrt{n}\{\hat{\theta}_n - \theta(P)\} + \tilde{b} \times \{\frac{1}{\sqrt{n}}\sum_{i=1}^n s_{\tau^*}(X_i)\} \xrightarrow{L} \mathbb{Z} \quad \text{(B.9)}$$

on $\mathbf{B}$ under P^n for some tight Borel random variable $\mathbb{Z}$. In addition, we have that

$$\sqrt{n}\{\hat{\theta}_n - \tilde{\theta}_n\} = -\tilde{b} \times \{\frac{1}{\sqrt{n}}\sum_{i=1}^n s_{\tau^*}(X_i)\} \xrightarrow{L} \Delta \quad \text{(B.10)}$$

by the central limit and continuous mapping theorems. Further note that since $\tilde{b} \neq 0$, we trivially have $\Delta \neq 0$ in $\mathbf{B}$ because $b^*(\Delta) \sim N(0, \|b^*(\tilde{b})s_{\tau^*}\|_{P,2}^2)$ and $\|b^*(\tilde{b})s_{\tau^*}\|_{P,2} > 0$ for some $b^* \in \mathbf{B}^*$ since $\tilde{b} \neq 0$. Thus, to conclude the proof of part (i) it only remains to show that $\tilde{\theta}_n$ is regular. To this end let $t \mapsto P_{t,g} \in \mathbf{P}$, and note Lemma 25.14 in van der Vaart (1998) yields

$$\sum_{i=1}^n \log\Big(\frac{dP_{1/\sqrt{n},g}}{dP}(X_i)\Big) = \frac{1}{\sqrt{n}}\sum_{i=1}^n g(X_i) - \frac{1}{2}\int g^2 dP + o_p(1) \quad \text{(B.11)}$$

under P^n, and thus Example 3.10.6 in van der Vaart and Wellner (1996) implies P^n and $P^n_{1/\sqrt{n},g}$ are mutually contiguous. Since $\tilde{\theta}_n$ is asymptotically linear, $\big(\sqrt{n}\{\tilde{\theta}_n - \theta(P)\}, \frac{1}{\sqrt{n}}\sum_{i=1}^n g(X_i)\big)$ converge jointly in $\mathbf{B} \times \mathbf{R}$ by Lemma D.4. Hence, by (B.11) and Lemma A.8.6 in Bickel et al. (1993) we obtain that

$$\sqrt{n}\{\tilde{\theta}_n - \theta(P)\} \stackrel{L_{n,g}}{\to} \mathbb{Z}_g \tag{B.12}$$

for some tight Borel $\mathbb{Z}_g$ on $\mathbf{B}$. Furthermore, since $T(P)$ is linear by Assumption 2.1(ii), and $\hat{\theta}_n$ is regular by hypothesis, Lemma D.4 and Theorem 5.2.3 in Bickel et al. (1993) imply there is a bounded linear map $\dot{\theta} : \bar{T}(P) \to \mathbf{B}$ such that for any $t \mapsto P_{t,g} \in \mathbf{P}$

$$\lim_{t \downarrow 0} \|t^{-1}\{\theta(P_{t,g}) - \theta(P)\} - \dot{\theta}(g)\|_{\mathbf{B}} = 0. \tag{B.13}$$

Therefore, combining (B.12) and (B.13) and the continuous mapping theorem yields

$$\sqrt{n}\{\tilde{\theta}_n - \theta(P_{1/\sqrt{n},g})\} \stackrel{L_{n,g}}{\to} \mathbb{Z}_g + \dot{\theta}(g). \tag{B.14}$$

Next we note that for any $b^* \in \mathbf{B}^*$, (B.9), (B.11), and the central limit theorem imply

$$\left(\begin{array}{c} \sqrt{n}\{b^*(\tilde{\theta}_n) - b^*(\theta(P))\} \\ \sum_{i=1}^n \log(\frac{dP_{1/\sqrt{n},g}}{dP}(X_i)) \end{array}\right) \stackrel{L}{\to} N\Big(\Big[\begin{array}{c} 0 \\ -\frac{1}{2}\int g^2 dP \end{array}\Big], \Sigma\Big) \tag{B.15}$$

under P^n, where since $\int g s_{\tau^*} dP = 0$ due to $g \in T(P)$ and $s_{\tau^*} \in \bar{T}(P)^{\perp}$, we have

$$\Sigma = \Big[\begin{array}{cc} \int (b^*(\nu) + b^*(\tilde{b}) s_{\tau^*})^2 dP & \int b^*(\nu) g dP \\ \int b^*(\nu) g dP & \int g^2 dP \end{array}\Big]. \tag{B.16}$$

In addition, since $b^*(\hat{\theta}_n)$ is an asymptotically linear regular estimator of $b^*(\theta(P))$, Proposition 3.3.1 in Bickel et al. (1993) and $g \in \bar{T}(P)$ imply $\int b^*(\nu) g dP = b^*(\dot{\theta}(g))$. Hence, results (B.15) and (B.16), and Lemma A.9.3 in Bickel et al. (1993) establish

$$\sqrt{n}\{b^*(\tilde{\theta}_n) - b^*(\theta(P_{1/\sqrt{n},g}))\} \stackrel{L_{n,g}}{\to} N\Big(0, \int (b^*(\nu) + b^*(\tilde{b}) s_{\tau^*})^2 dP\Big). \tag{B.17}$$

Define $\zeta_{b^*}(X_i) \equiv \{b^*(\nu(X_i)) + b^*(\tilde{b}) s_{\tau^*}(X_i)\}$, and for any finite collection $\{b_k^*\}_{k=1}^K \subset \mathbf{B}^*$ let $(\mathbb{W}_{b_1^*}, \ldots, \mathbb{W}_{b_K^*})$ denote a multivariate normal vector with $E[\mathbb{W}_{b_k^*}] = 0$ for all $1 \leq k \leq K$ and $E[\mathbb{W}_{b_k^*}\mathbb{W}_{b_j^*}] = E[\zeta_{b_k^*}(X_i)\zeta_{b_j^*}(X_i)]$ for any $1 \leq j \leq k \leq K$. Letting $C_b(\mathbf{R}^K)$ denote the set of continuous and bounded functions on $\mathbf{R}^K$, we then obtain from (B.14), (B.17), the Cramer-Wold device, and the continuous mapping theorem that

$$E\big[f\big(b_1^*(\mathbb{Z}_g + \dot{\theta}(g)), \ldots, b_K^*(\mathbb{Z}_g + \dot{\theta}(g))\big)\big] = E\big[f\big(b_1^*(\mathbb{W}_{b_1^*}), \ldots, b_K^*(\mathbb{W}_{b_K^*})\big)\big], \tag{B.18}$$

for any $f \in C_b(\mathbf{R}^K)$. Since $\mathcal{G} \equiv \{f \circ (b_1^*, \ldots, b_K^*) : f \in C_b(\mathbf{R}^K), \{b_k^*\}_{k=1}^K \subset \mathbf{B}^*, 1 \leq K < \infty\}$ is a vector lattice that separates points in $\mathbf{B}$, it follows from Lemma 1.3.12 in van der Vaart and Wellner (1996) that there is a unique tight Borel measure $\mathbb{W}$ on $\mathbf{B}$ satisfying (B.18). In particular, since the right hand side of (B.18) does not depend on g, we conclude the law of $\mathbb{Z}_g + \dot{\theta}(g)$ is constant in g, establishing the regularity of $\tilde{\theta}_n$.

For part (ii) of the Lemma, we let ν and $\tilde{\nu}$ denote the influence functions of $\hat{\theta}_n$ and $\tilde{\theta}_n$ respectively, and note that since $\|b^*\|_{\mathbf{B}^*} \leq 1$ for all $b^* \in \mathbf{T}$ it follows that

$$\sup_{b^* \in \mathbf{T}} \big|\hat{\mathbb{G}}_n(b^*) - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} b^*(\nu(X_i) - \tilde{\nu}(X_i))\big|$$
$$\leq \sup_{b^* \in \mathbf{B}^*} \|b^*\|_{\mathbf{B}^*} \times \big\|\sqrt{n}\{\hat{\theta}_n - \tilde{\theta}_n\} - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \{\nu(X_i) - \tilde{\nu}(X_i)\}\big\|_{\mathbf{B}} = o_p(1). \quad \text{(B.19)}$$

Moreover, note that since $b^*(\hat{\theta}_n)$ and $b^*(\tilde{\theta}_n)$ are both asymptotically linear regular estimators of the parameter $b^*(\theta(P)) \in \mathbf{R}$, Proposition 3.3.1 in Bickel et al. (1993) implies

$$\Pi_T(b^*(\nu)) = \Pi_T(b^*(\tilde{\nu})). \quad \text{(B.20)}$$

In particular, since $b^*(\nu) \in L_0^2(P)$, we may decompose $b^*(\nu) = \Pi_T(b^*(\nu)) + \Pi_{T^\perp}(b^*(\nu))$, and therefore applying an identical argument to $b^*(\tilde{\nu})$ we can conclude that

$$b^*(\nu - \tilde{\nu}) = \Pi_{T^\perp}(b^*(\nu)) - \Pi_{T^\perp}(b^*(\tilde{\nu})) \quad \text{(B.21)}$$

by result (B.20). It follows $b^*(\nu - \tilde{\nu}) \in \bar{T}(P)^\perp$ for any $b^* \in \mathbf{T}$, which together with (B.19) verifies Assumption 3.1(i) holds. Next, define $F : \mathbf{B} \to \ell^\infty(\mathbf{T})$ to be given by $F(b)(b^*) = b(b^*)$ for any $b \in \mathbf{B}$, and note F is linear and in addition

$$\|F(b)\|_\infty = \sup_{\|b^*\|_{\mathbf{B}^*} \leq 1} |b(b^*)| = \|b\|_{\mathbf{B}}, \quad \text{(B.22)}$$

due to the definition of $\mathbf{T}$ and Lemma 6.10 in Aliprantis and Border (2006). In particular, (B.22) implies F is continuous, and by the continuous mapping theorem we obtain

$$\hat{\mathbb{G}}_n = F(\sqrt{n}\{\hat{\theta}_n - \tilde{\theta}_n\}) \xrightarrow{L} F(\Delta) \ \text{ in } \ \ell^\infty(\mathbf{T}). \quad \text{(B.23)}$$

Let $\mathbb{G}_0 \equiv F(\Delta)$ and note Gaussianity of $\mathbb{G}_0$ follows by (B.19). Moreover, we note there must exist a $b^* \in \mathbf{B}^*$ such that $\|b^*(\nu - \tilde{\nu})\|_{P,2} > 0$, for otherwise Lemma E.1 (in the Online Appendix) would imply $\Delta = 0$ contradicting Assumption 3.2. Hence, $\mathbb{G}_0$ is in addition non-degenerate, which verifies Assumption 3.1(ii). ■

Proof of Theorem 3.3: For part (i) of the Theorem, we note that Lemma E.2 (in the Online Appendix), Assumption 3.3(i) and the continuous mapping theorem imply that for any path $t \mapsto P_{t,g} \in \mathcal{M}$,

$$\Psi(\hat{\mathbb{G}}_n) \xrightarrow{L_{n,g}} \Psi(\mathbb{G}_0 + \Delta_g), \quad \text{(B.24)}$$

where $\Delta_g : \mathbf{T} \to \mathbf{R}$ satisfies $\Delta_g(\tau) = \int s_\tau g dP$ for any $\tau \in \mathbf{T}$. Further note that by direct calculation $\Delta_{-g} = -\Delta_g$, and hence Lemma E.3 (in the Online Appendix) implies $-\Delta_g$ belongs to the support of $\mathbb{G}_0$ for any $g \in L_0^2(P)$. In particular, since $\Psi(0) = 0$ and $\Psi(b) \geq 0$ for all $b \in \ell^\infty(\mathbf{T})$, it follows that for any $c > 0$ there exists an open neighborhood N_c of $-\Delta_g \in \ell^\infty(\mathbf{T})$ such that $0 \leq \Psi(b + \Delta_g) \leq c$ for all $b \in N_c$. Thus, we can conclude for any $c > 0$ that

$$P(\Psi(\mathbb{G}_0 + \Delta_g) \leq c) \geq P(\mathbb{G}_0 \in N_c) > 0, \quad \text{(B.25)}$$

where the final inequality follows from $-\Delta_g$ belonging to the support of $\mathbb{G}_0$. Next, note that Theorem 7.1.7 in Bogachev (2007) implies $\mathbb{G}_0$ is a regular measure, and hence since it is tight by Assumption 3.1(ii) it follows that it is also a Radon measure. Together with the convexity of the map $\Psi(\cdot + \Delta_g) : \ell^\infty(\mathbf{T}) \to \mathbf{R}$, $\mathbb{G}_0$ being Radon allows us to apply Theorem 11.1 in Davydov et al. (1998) to conclude that the point

$$c_0 \equiv \inf\{c : P(\Psi(\mathbb{G}_0 + \Delta_g) \leq c) > 0\} \tag{B.26}$$

is the only possible discontinuity point of the c.d.f. of $\Psi(\mathbb{G}_0 + \Delta_g)$. However, since $\Psi(b) \geq 0$ for all $b \in \ell^\infty(\mathbf{T})$, result (B.25) holding for any $c > 0$ implies that $c_0 = 0$. In particular, $c_{1-\alpha} > 0$ by hypothesis implies that $c_{1-\alpha}$ is a continuity point of the c.d.f. of $\Psi(\mathbb{G}_0 + \Delta_g)$ for any $g \in L_0^2(P)$. Therefore, result (B.24) allows us to conclude: for any path $t \mapsto P_{t,g} \in \mathcal{M}$,

$$\pi(g) \equiv \lim_{n\to\infty} P_{1/\sqrt{n},g}(\Psi(\hat{\mathbb{G}}_n) > c_{1-\alpha}) = P(\Psi(\mathbb{G}_0 + \Delta_g) > c_{1-\alpha}), \tag{B.27}$$

which establishes that the test ϕ_n indeed has an asymptotic local power function. Moreover, if $t \mapsto P_{t,g} \in \mathbf{P}$, then Lemma E.2 (in the Online Appendix) implies $\Delta_g = 0$ and hence result (B.27) yields

$$\lim_{n\to\infty} P_{1/\sqrt{n},g}(\Psi(\hat{\mathbb{G}}_n) > c_{1-\alpha}) = P(\Psi(\mathbb{G}_0) > c_{1-\alpha}) = \alpha, \tag{B.28}$$

where we exploited that $c_{1-\alpha}$ is the $1-\alpha$ quantile of $\Psi(\mathbb{G}_0)$ and that the c.d.f. of $\Psi(\mathbb{G}_0)$ is continuous at $c_{1-\alpha}$. Thus, we conclude from (B.28) that ϕ_n is also an asymptotic level α specification test. On the other hand, we note that $\Delta_g \neq 0$ whenever $\Pi_S(g) \neq 0$ since $\Delta_g(\tau) = \int s_\tau g dP$ and $\bar{S}(P) = \overline{\text{lin}}\{s_\tau : \tau \in \mathbf{T}\}$. In addition, Theorem 3.6.1 in Bogachev (1998) implies the support of $\mathbb{G}_0$ is a separable vector subspace of $\ell^\infty(\mathbf{T})$, and hence $\Delta_g \neq 0$ belonging to the support of $\mathbb{G}_0$ and Lemma E.5 (in the Online Appendix) establish

$$P(\Psi(\mathbb{G}_0 + \Delta_g) < c_{1-\alpha}) < P(\Psi(\mathbb{G}_0) < c_{1-\alpha}) = 1 - \alpha. \tag{B.29}$$

We can now exploit that $c_{1-\alpha} > 0$ is a continuity point of the c.d.f. of $\Psi(\mathbb{G}_0 + \Delta_g)$ together with results (B.27) and (B.29) to conclude that for any path $t \mapsto P_{t,g} \in \mathcal{M}$ with $\Pi_S(g) \neq 0$,

$$\lim_{n\to\infty} P_{1/\sqrt{n},g}(\Psi(\hat{\mathbb{G}}_n) > c_{1-\alpha}) = 1 - P(\Psi(\mathbb{G}_0 + \Delta_g) \leq c_{1-\alpha}) > \alpha, \tag{B.30}$$

which satisfies (15). Finally, for any path $t \mapsto P_{t,g} \in \mathcal{M}$ with $\Pi_S(g) = 0$, we have $\Delta_g = 0$ and hence $\pi(g) = \alpha$ by equations (B.27) and (B.28). Thus the test ϕ_n is also locally unbiased, and we establish part (i) of the Theorem.

For part (ii) of the Theorem, we proceed as in Lemma 3.1(ii) and set $\mathbf{T} = \{b^* \in \mathbf{B}^* : \|b^*\|_{\mathbf{B}^*} \leq 1\}$ and $\hat{\mathbb{G}}_n(b^*) = \sqrt{n} b^*(\hat{\theta}_n - \tilde{\theta}_n)$ for any $b^* \in \mathbf{B}^*$. Since $s_{b^*} = b^*(\nu - \tilde{\nu})$ by Lemma 3.1(ii), we obtain by definition that $\bar{S}(P) = \overline{\text{lin}}\{b^*(\nu - \tilde{\nu}) : b^* \in \mathbf{T}\} = \overline{\text{lin}}\{b^*(\nu - \tilde{\nu}) : b^* \in \mathbf{B}^*\}$. Moreover, if $\Psi = \|\cdot\|_\infty$, then

$$\Psi(\hat{\mathbb{G}}_n) = \sup_{\|b^*\|_{\mathbf{B}^*} \leq 1} |b^*(\sqrt{n}\{\hat{\theta}_n - \tilde{\theta}_n\})| = \sqrt{n}\|\hat{\theta}_n - \tilde{\theta}_n\|_{\mathbf{B}}, \tag{B.31}$$

where the final equality follows by Lemma 6.10 in Aliprantis and Border (2006). Since $\Psi = \|\cdot\|_\infty$ satisfies Assumption 3.3, the second claim of the Theorem follows. ■

Proof of Lemma 3.2: Part (i) of the Lemma is immediate since $\bar{T}(P)^{\perp} \cap \bar{M}(P) \subseteq \bar{T}(P)^{\perp}$.

For part (ii) of the Lemma, we will exploit Theorem A.1(i) and its notation. Note that Theorem A.1(i) implies there exists a level α test $\phi : (\mathbb{Y}^T, \mathbb{Y}^{T^{\perp}}) \to [0,1]$ of the hypothesis in (A.2), and such that for any path $t \mapsto P_{t,g} \in \mathcal{M}$

$$\lim_{n\to\infty} \int \phi_n dP^n_{1/\sqrt{n},g} = \int \phi dQ_g, \tag{B.32}$$

where Q_g denotes the (unknown) distribution of $(\mathbb{Y}^T, \mathbb{Y}^{T^{\perp}})$ as defined in (A.1). Recall $\{s_\tau\}_{\tau=1}^d$ with $d < \infty$ is an orthonormal basis for $\bar{T}(P)^{\perp} \cap \bar{M}(P)$, we let d_r denote the (possibly infinite) dimension of $\bar{M}(P)^{\perp}$ and $\{r_k\}_{k=1}^{d_r}$ be an orthonormal basis for $\bar{M}(P)^{\perp}$. By (21), $\{s_\tau\}_{\tau=1}^d \cup \{r_k\}_{k=1}^{d_r}$ is then an orthonormal basis for $\bar{T}(P)^{\perp}$. Thus, in Theorem A.1(i) we may set $\{\psi_k^{T^{\perp}}\}_{k=1}^{d_{T^{\perp}}} = \{s_\tau\}_{\tau=1}^d \cup \{r_k\}_{k=1}^{d_r}$, which implies we may write $\mathbb{Y}^{T^{\perp}} = (\mathbb{M}, \mathbb{R}) \in \mathbf{R}^d \times \mathbf{R}^{d_r}$, where the vectors $\mathbb{M} \equiv (\mathbb{M}_1, \ldots, \mathbb{M}_d)'$ and $\mathbb{R} = (\mathbb{R}_1, \ldots, \mathbb{R}_{d_r})'$ have mutually independent coordinates, and whenever $(\mathbb{Y}^T, \mathbb{Y}^{T^{\perp}})$ are distributed according to Q_g, the induced distribution on $(\mathbb{M}, \mathbb{R})$ is

$$\begin{aligned} \mathbb{M}_\tau &\sim N\Big(\int g s_\tau dP,\ 1\Big) \quad \text{for} \quad 1 \le \tau \le d \\ \mathbb{R}_k &\sim N\Big(\int g r_k dP,\ 1\Big) \quad \text{for} \quad 1 \le k \le d_r. \end{aligned} \tag{B.33}$$

Let Φ denote the standard normal measure on $\mathbf{R}$. Note that $Q_0 = \bigotimes_{k=1}^{d_T} \Phi \times \bigotimes_{k=1}^{d} \Phi \times \bigotimes_{k=1}^{d_r} \Phi$, we can define a test $\bar{\phi} : \mathbb{M} \to [0,1]$ to be given by

$$\bar{\phi}(\mathbb{M}) \equiv E_{Q_0}[\phi(\mathbb{Y}^T, \mathbb{M}, \mathbb{R})|\mathbb{M}], \tag{B.34}$$

where the expectation is taken over $(\mathbb{Y}^T, \mathbb{Y}^{T^{\perp}}) \sim Q_0$. Since $\{g \in \bar{T}(P)^{\perp} \cap \bar{M}(P) : \|g\|_{P,2} \ge \kappa\} \subseteq \mathcal{G}(\kappa)$, we can conclude from result (B.32) that

$$\begin{aligned} \inf_{g \in \mathcal{G}(\kappa)} \lim_{n\to\infty} \int \phi_n dP^n_{1/\sqrt{n},g} &\le \inf_{g \in \bar{T}(P)^{\perp} \cap \bar{M}(P) : \|g\|_{P,2} \ge \kappa} \int \phi dQ_g \\ &= \inf_{g \in \bar{T}(P)^{\perp} \cap \bar{M}(P) : \|g\|_{P,2} \ge \kappa} \int \bar{\phi} d\{\bigotimes_{\tau=1}^{d} \Phi(\cdot - \int g s_\tau dP)\}, \end{aligned} \tag{B.35}$$

where in the equality we exploited (B.34), the independence of $(\mathbb{Y}^T, \mathbb{R})$ and $\mathbb{M}$, and that for any $g \in \bar{T}(P)^{\perp} \cap \bar{M}(P)$ it follows that $(\mathbb{Y}^T, \mathbb{R}) \sim \bigotimes_{k=1}^{d_T} \Phi \times \bigotimes_{k=1}^{d_r} \Phi$ under Q_g as a result of g being orthogonal to $\{\psi_k^T\}_{k=1}^{d_T} \cup \{r_k\}_{k=1}^{d_r}$. Finally, note that

$$\inf_{g \in \bar{T}(P)^{\perp} \cap \bar{M}(P) : \|g\|_{P,2} \ge \kappa} \int \bar{\phi} d\{\bigotimes_{\tau=1}^{d} \Phi(\cdot - \int g s_\tau dP)\} = \inf_{h \in \mathbf{R}^d : \|h\| \ge \kappa} \int \bar{\phi} d\{\bigotimes_{\tau=1}^{d} \Phi(\cdot - h_k)\} \tag{B.36}$$

by Parseval's equality and where $h = (h_1, \ldots, h_d)$. Let $\chi_d^2(\kappa)$ denote a chi-squared random variable with d degrees of freedom and noncentrality parameter κ. It then follows from $\int \bar{\phi} d\{\bigotimes_{\tau=1}^{d} \Phi\} \le \alpha$ due to (B.34) and ϕ being a level α test of (A.2), results (B.35) and (B.36) and Problem 8.29 in

Lehmann and Romano (2005) that

$$\inf_{g\in\mathcal{G}(\kappa)} \lim_{n\to\infty} \int \phi_n dP^n_{1/\sqrt{n},g} \le P\big(\chi^2_d(\kappa) > q_{d,1-\alpha}\big) \tag{B.37}$$

where $q_{d,1-\alpha}$ denotes the $(1-\alpha)$ quantile of a chi-squared random variable with d degrees of freedom. However, note that since $\{s_\tau\}_{\tau=1}^d$ is orthonormal by hypothesis, Assumption 3.1(i) implies $\|\hat{\mathbb{G}}_n\|^2 \xrightarrow{L} \chi^2_d(0)$ under P^n and therefore $c_{1-\alpha} = q_{d,1-\alpha}$. Furthermore, Lemma E.2 (in the Online Appendix) implies that for $\mathbb{G}_0 \sim N(0, I_d)$ with I_d the $d\times d$ identity matrix and $\Delta_g \in \mathbf{R}^d$ given by $\Delta_g = (\int gs_1 dP, \ldots, \int gs_d dP)'$ we must have for any path $t \mapsto P_{t,g} \in \mathcal{M}$

$$\hat{\mathbb{G}}_n \xrightarrow{L_{n,g}} \mathbb{G}_0 + \Delta_g. \tag{B.38}$$

In particular, since $\|\Delta_g\| = \|\Pi_{T^\perp}(g)\|_{P,2}$ for any $g \in \bar{M}(P)$, we obtain from (B.38) that

$$\inf_{g\in\mathcal{G}(\kappa)} \lim_{n\to\infty} P_{1/\sqrt{n},g}(\|\hat{\mathbb{G}}_n\|^2 > c_{1-\alpha})$$
$$= \inf_{g\in\mathcal{G}(\kappa)} P(\|\mathbb{G}_0 + \Delta_g\|^2 > q_{d,1-\alpha}) = P(\chi^2_d(\kappa) > q_{d,1-\alpha}). \tag{B.39}$$

Therefore, part (ii) of the Lemma follows from (B.37) and (B.39).

For part (iii) of the Lemma, it suffices to verify that $b^*(\nu - \tilde{\nu}) \in \bar{T}(P)^\perp \cap \bar{M}(P)$ for all $b^* \in \mathbf{B}^*$. To this end, we note that $b^*(\hat{\theta}_n)$ and $b^*(\tilde{\theta}_n)$ are both asymptotically linear regular (with respect to $\mathbf{P}$) estimators of $b^*(\theta(P))$ with influence functions $b^*(\nu)$ and $b^*(\tilde{\nu})$ respectively. We also have that

$$b^*(\tilde{\nu}) - b^*(\nu) = \Pi_{T^\perp}(b^*(\tilde{\nu})) \tag{B.40}$$

since by Proposition 3.3.1 in Bickel et al. (1993), $b^*(\nu) \in \bar{T}(P)$ due to $b^*(\hat{\theta}_n)$ being efficient (with respect to $\mathbf{P}$), and $\Pi_T(b^*(\tilde{\nu})) = b^*(\nu)$ due to $b^*(\tilde{\theta}_n)$ being regular (with respect to $\mathbf{M}$ and $\mathbf{P}$). However, $b^*(\tilde{\nu})$ being efficient with respect to $\mathbf{M}$ and Proposition 3.3.1 in Bickel et al. (1993) imply $b^*(\tilde{\nu}) \in \bar{M}(P)$. Since $\bar{M}(P)$ is a vector subspace and $b^*(\nu) \in \bar{T}(P) \subseteq \bar{M}(P)$, result (B.40) additionally implies $\Pi_{T^\perp}(b^*(\tilde{\nu})) \in \bar{M}(P)$, and thus $b^*(\nu - \tilde{\nu}) \in \bar{T}(P)^\perp \cap \bar{M}(P)$ as claimed. ■

Appendix C - Proofs for $T(P)$ a Convex Cone

Proof of Lemma 5.1: Since $\bar{T}(P)$ is a convex cone by Assumption 5.1, Proposition 46.5.4 in Zeidler (1984) implies $L^2_0(P) = \bar{T}(P) \oplus \bar{T}(P)^-$. The Lemma then follows since $\bar{T}(P)^- = \{0\}$ if and only if $\bar{T}(P) = L^2_0(P)$. ■

Lemma C.1. *Let Assumption 5.1 hold and P be locally just identified by $\mathbf{P}$. Then: for all bounded function $f : \mathbf{X} \to \mathbf{R}$, the sample mean, $n^{-1}\sum_{i=1}^n f(X_i)$, is an asymptotically locally admissible estimator of $\int f dP$ under any Ψ-loss.*

Proof of Lemma C.1: We aim to show that if P is locally just identified by $\mathbf{P}$, then $n^{-1}\sum_{i=1}^n f(X_i)$ is an asymptotically locally admissible estimator of $\int f dP$. To this end, we note that for any bounded $f : \mathbf{X} \to \mathbf{R}$, Theorem 3.10.12 in van der Vaart and Wellner (1996) implies that for any path $t \mapsto P_{t,g} \in \mathcal{M}$,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \{f(X_i) - \int f dP_{1/\sqrt{n},g}\} \stackrel{L_{n,g}}{\to} \mathbb{G}_0 \tag{C.1}$$

where $\mathbb{G}_0 \sim N(0, \text{Var}\{f(X_i)\})$. Therefore, since Ψ is bounded and continuous we obtain from (C.1) that for any path $t \mapsto P_{t,g} \in \mathbf{P}$

$$\limsup_{n\to\infty} E_{P_{1/\sqrt{n},g}}\Big[\Psi\Big(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \{f(X_i) - \int f dP_{1/\sqrt{n},g}\}\Big)\Big] = E[\Psi(\mathbb{G}_0)]. \tag{C.2}$$

By way of contradiction, next suppose that $\frac{1}{n}\sum_{i=1}^{n} f(X_i)$ is not an asymptotically locally admissible estimator of $\int f dP$ under Ψ-loss. It then follows that there must exist another estimator $\hat{\theta}_n : \{X_i\}_{i=1}^{n} \to \mathbf{R}$ of $\int f dP$ satisfying for any path $t \mapsto P_{t,g} \in \mathbf{P}$ and some tight law $\mathbb{Z}_g$

$$\sqrt{n}\{\hat{\theta}_n - \int f dP_{1/\sqrt{n},g}\} \stackrel{L_{n,g}}{\to} \mathbb{Z}_g, \tag{C.3}$$

and moreover, by result (C.2), for any $t \mapsto P_{t,g} \in \mathbf{P}$, $\hat{\theta}_n$ must additionally be such that

$$\limsup_{n\to\infty} E_{P_{1/\sqrt{n},g}}\Big[\Psi\Big(\sqrt{n}\{\hat{\theta}_n - \int f dP_{1/\sqrt{n},g}\}\Big)\Big] \leq E[\Psi(\mathbb{G}_0)] \tag{C.4}$$

with strict inequality holding for some $t \mapsto P_{t,g} \in \mathbf{P}$. In particular, since Ψ is bounded and continuous, results (C.3) and (C.4) imply that

$$E[\Psi(\mathbb{G}_0)] \geq \sup_{g\in T(P)} E[\Psi(\mathbb{Z}_g)] = \sup_{g\in \bar{T}(P)} E[\Psi(\mathbb{Z}_g)] \tag{C.5}$$

where the equality follows from Lemma E.6 (in the Online Appendix), which establishes both that $\mathbb{Z}_g$ is well defined for $g \in \bar{T}(P)$ and that the supremums over $T(P)$ and $\bar{T}(P)$ must be equal. Since P is just identified by $\mathbf{P}$, however, we have $\bar{T}(P) = L_0^2(P)$, which implies $f - \int f dP \in \bar{T}(P)$. Therefore, result (C.5), Theorem 2.6 in van der Vaart (1989), and Proposition 8.6 in van der Vaart (1998) together establish that under $\bigotimes_{i=1}^{n} P$ we must have

$$\sqrt{n}\{\hat{\theta}_n - \int f dP\} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Big(f(X_i) - \int f dP\Big) + o_p(1). \tag{C.6}$$

Equivalently, $\sqrt{n}\{\hat{\theta}_n - \frac{1}{n}\sum_{i=1}^{n} f(X_i)\} = o_p(1)$ under $\bigotimes_{i=1}^{n} P$ and, by contiguity, also under $\bigotimes_{i=1}^{n} P_{1/\sqrt{n},g}$ for any path $t \mapsto P_{t,g}$. However, by results (C.1) and (C.3) we can then conclude that $\mathbb{Z}_g$ must equal $\mathbb{G}_0$ in distribution, thus establishing the desired contradiction since as a result (C.4) cannot hold strictly for any path $t \mapsto P_{t,g} \in \mathbf{P}$. ■

Proof of Theorem 5.1: We note that if P is locally just identified, then by Lemma C.1 it follows that (ii) implies (i). Similarly, we also note that by Theorem 3.1(ii) it follows that (iii) implies (i). Thus, to conclude the proof we need only show that (i) implies (ii) and (iii). To this end, note that if P is locally overidentified by $\mathbf{P}$, then Lemma 5.1 implies there exists a $0 \neq \tilde{f} \in \bar{T}(P)^{-}$, which without loss of generality we assume satisfies $\|\tilde{f}\|_{P,2} = 1$. The fact that (i) implies (ii) then follows by applying Theorem 5.4 with $s_{\mathcal{T}^\star} = \tilde{f}$ and f defined by $f(x) = \tilde{f}(x)1\{|\tilde{f}(x)| \leq M\}$, which satisfies $\int \tilde{f} f dP > 0$ for M large enough. Finally, to establish (i) implies (iii) we note that by Theorem 3.10.12 in van der Vaart and Wellner (1996)

$$\mathbb{G}_n(\tilde{f}) \stackrel{L_{n,g}}{\to} \mathbb{Z} + \int \tilde{f} g dP \quad \text{for any path } t \mapsto P_{t,g} \in \mathcal{M} \tag{C.7}$$

where $\mathbb{Z} \sim N(0,1)$. We define the test $\phi_n \equiv 1\{\mathbb{G}_n(\tilde{f}) > z_{1-\alpha}\}$ for $z_{1-\alpha}$ the $1-\alpha$ quantile of $\mathbb{Z}$. Then equation (C.7) implies that

$$\pi(g) \equiv \lim_{n\to\infty} \int \phi_n dP_{1/\sqrt{n},g} = P\Big(\mathbb{Z} + \int \tilde{f} g dP > z_{1-\alpha}\Big) \tag{C.8}$$

for any path $t \mapsto P_{t,g} \in \mathcal{M}$. Thus, whenever the path $t \mapsto P_{t,g} \in \mathbf{P}$, it follows from $g \in \bar{T}(P)$ and $\tilde{f} \in \bar{T}(P)^-$ that $\int g\tilde{f} dP \leq 0$ and hence by (C.8) that $\pi(g) \leq \alpha$ – i.e. ϕ_n has asymptotic local level α. On the other hand, there exists a path $t \mapsto P_{t,\tilde{f}} \in \mathcal{M}$ with score $\tilde{f} \in \bar{T}(P)^-$. This and (C.8) together imply $\pi(\tilde{f}) > \alpha$, hence we conclude (i) implies (iii). ■

Proof of Theorem 5.2: Fix a path $t \mapsto P_{t,g}$ such that its score $g \in L_0^2(P)$ satisfies $\lambda \Pi_{T^-}(g) \in \bar{T}(P)$ for any $\lambda \leq 0$, and note that by Proposition 46.5.4 in Zeidler (1984)

$$g = \Pi_T(g) + \Pi_{T^-}(g). \tag{C.9}$$

Moreover, we note that if $\Pi_{T^-}(g) = 0$ then $g \in \bar{T}(P)$ and thus $\pi(g) \leq \alpha$ since $\pi(g) \leq \alpha$ for all $g \in \bar{T}(P)$ by hypothesis. We therefore assume without loss of generality that $\Pi_{T^-}(g) \neq 0$ and observe that by the hypotheses of the Theorem there exist a $f^\star \in \{f_1, f_2\}$ such that $f^\star \in \bar{T}(P)^-$, $f^\star$ is linearly independent of $\Pi_{T^-}(g)$, and $f^\star$ satisfies $\lambda f^\star \in \bar{T}(P)$ for all $\lambda \leq 0$. Defining

$$H \equiv \big\{h \in L_0^2(P) : h = \Pi_T(g) + \gamma_1 \Pi_{T^-}(g) + \gamma_2 f^\star \text{ for some } (\gamma_1, \gamma_2) \in \mathbf{R}^2\big\}, \tag{C.10}$$

we may then construct for any $h \in H$ a path $t \mapsto \bar{P}_{t,h}$ whose score is h and such that $\bar{P}_{t,h} \ll P \ll \bar{P}_{t,h}$; see, e.g., Example 3.2.1 in Bickel et al. (1993). Recall $\mathcal{B}$ is the σ-algebra on $\mathbf{X}$, and consider the sequence of experiments $\mathcal{E}_n$ given by

$$\mathcal{E}_n \equiv \Big(\mathbf{X}^n, \mathcal{B}^n, \bigotimes_{i=1}^n \bar{P}_{1/\sqrt{n},h} : h \in H\Big). \tag{C.11}$$

Setting $h_0 \equiv \Pi_T(g)$, then observe that Lemma 25.14 in van der Vaart (1998) implies

$$\sum_{i=1}^n \log\Big(\frac{d\bar{P}_{1/\sqrt{n},h}}{d\bar{P}_{1/\sqrt{n},h_0}}(X_i)\Big) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \big(h(X_i) - h_0(X_i)\big) - \frac{1}{2}\int (h^2 - h_0^2) dP + o_p(1) \tag{C.12}$$

under $P^n \equiv \bigotimes_{i=1}^n P$, and where we exploited that $\bar{P}_{t,h} \ll P \ll \bar{P}_{t,h_0}$. Since similarly

$$\sum_{i=1}^n \log\Big(\frac{d\bar{P}_{1/\sqrt{n},h_0}}{dP}(X_i)\Big) = \frac{1}{\sqrt{n}} \sum_{i=1}^n h_0(X_i) - \frac{1}{2}\int h_0^2 dP + o_p(1) \tag{C.13}$$

under P^n by Lemma 25.14 in van der Vaart (1998), it follows by LeCam's 3rd Lemma (see, e.g., Lemma A.8.6 in Bickel et al. (1993)) that for an arbitrary finite subset $\{h_j\}_{j=1}^J \equiv I \subseteq H$ and L_{n,h_0} denoting the law under $\bigotimes_{i=1}^n \bar{P}_{1/\sqrt{n},h_0}$ we have

$$\Big(\sum_{i=1}^n \log\Big(\frac{d\bar{P}_{1/\sqrt{n},h_1}}{d\bar{P}_{1/\sqrt{n},h_0}}(X_i)\Big), \ldots, \sum_{i=1}^n \log\Big(\frac{d\bar{P}_{1/\sqrt{n},h_J}}{d\bar{P}_{1/\sqrt{n},h_0}}(X_i)\Big)\Big)' \stackrel{L_{n,h_0}}{\to} N(-\mu_I, \Sigma_I) \tag{C.14}$$

where $\Sigma_I \equiv \int (h_1 - h_0, \ldots, h_J - h_0)(h_1 - h_0, \ldots, h_J - h_0)' dP$ and the mean is given by $\mu_I \equiv \frac{1}{2}\big(\int (h_1 - h_0)^2 dP, \ldots, \int (h_J - h_0)^2 dP\big)'$. Next, define $v_h \in \mathbf{R}^2$ and $\Omega \in \mathbf{R}^{2\times 2}$ by

$$v_h \equiv \begin{pmatrix} \int \{\Pi_{T^-}(g)\} h dP \\ \int f^\star h dP \end{pmatrix} \qquad \Omega \equiv \begin{pmatrix} \int \{\Pi_{T^-}(g)\}^2 dP & \int \{\Pi_{T^-}(g)\} f^\star dP \\ \int \{\Pi_{T^-}(g)\} f^\star dP & \int \{f^\star\}^2 dP \end{pmatrix} \tag{C.15}$$

and note that the linear independence of $f^\star$ and $\Pi_{T^-}(g)$ in $L_0^2(P)$ imply Ω is invertible. For any $h \in H$, then let Q_h be the bivariate normal law on $\mathbf{R}^2$ satisfying

$$Q_h \stackrel{L}{=} N(\Omega^{-1}\{v_h - 2v_{h_0}\}, \Omega^{-1}). \tag{C.16}$$

Further observe that for any $h \in H$ and $U \in \mathbf{R}^2$ we can obtain by direct calculation

$$\log(\frac{dQ_h}{dQ_{h_0}}(U)) = U'(v_h - v_{h_0}) + \frac{1}{2}\{v_{h_0}'\Omega^{-1}v_{h_0} - (v_h - 2v_{h_0})'\Omega^{-1}(v_h - 2v_{h_0})\} \tag{C.17}$$

and therefore exploiting (C.17) and $(v_{h_i} - v_{h_0})'\Omega^{-1}(v_{h_k} - v_{h_0}) = \int (h_i - h_0)(h_k - h_0) dP$ for any $h_i, h_k \in H$ implies that for any finite subset $\{h_j\}_{j=1}^J \equiv I \subseteq H$ we have

$$\Big(\log(\frac{dQ_{h_1}}{dQ_{h_0}}), \ldots, \log(\frac{dQ_{h_J}}{dQ_{h_0}})\Big) \sim N(-\mu_I, \Sigma_I) \tag{C.18}$$

under Q_{h_0}. Since (C.14) and Corollary 12.3.1 in Lehmann and Romano (2005) imply $\{P_{1/\sqrt{n},h}\}$ and $\{P_{1/\sqrt{n},h_0}\}$ are mutually contiguous for any $h \in H$, results (C.14) and (C.18) together with Lemma 10.2.1 in LeCam (1986) establish $\mathcal{E}_n$ converges weakly to

$$\mathcal{E} \equiv \big(\mathbf{R}^2, \mathcal{A}^2, Q_h : h \in H\big), \tag{C.19}$$

where $\mathcal{A}$ denotes the Borel σ-algebra on $\mathbf{R}$.

By the asymptotic representation theorem, see for example Theorem 7.1 in van der Vaart (1991a), it then follows from ϕ_n having a local asymptotic power function π that there exists a test ϕ based on a single observation of $U \sim Q_h$ such that for all $h \in H$

$$\pi(h) \equiv \lim_{n\to\infty} \int \phi_n d\bar{P}^n_{1/\sqrt{n},h} = \int \phi dQ_h. \tag{C.20}$$

Further note that any $h \in H$ can be written as $h = \Pi_T(g) + \gamma_1(h)\Pi_{T^-}(g) + \gamma_2(h) f^\star$ for some $\gamma(h) = (\gamma_1(h), \gamma_2(h))' \in \mathbf{R}^2$ and that moreover $\gamma(h) = \Omega^{-1}\{v_h - v_{h_0}\}$. In addition we observe that $\lambda f^\star, \lambda \Pi_{T^-}(g) \in \bar{T}(P)^-$ whenever $\lambda \geq 0$ and $\lambda f^\star, \lambda\Pi_{T^-}(g) \in \bar{T}(P)$ whenever $\lambda \leq 0$ together with the linear independence of $\Pi_{T^-}(g)$ imply that $h \in \bar{T}(P)$ if and only if $\gamma_1(h) \leq 0$ and $\gamma_2(h) \leq 0$.

Thus, the hypothesis on π and result (C.20) yield

$$\begin{array}{ll} \int \phi dQ_h \leq \alpha & \text{if } \gamma_1(h) \leq 0 \text{ and } \gamma_2(h) \leq 0 \\ \int \phi dQ_h \geq \alpha & \text{if } \gamma_1(h) > 0 \text{ or } \gamma_2(h) > 0 \end{array} . \tag{C.21}$$

However, (C.21), $h \mapsto \gamma(h)$ being bijective between H and $\mathbf{R}^2$, and Lehmann (1952) p. 542 imply that $\int \phi dQ_h = \alpha$ for all $h \in H$. In particular, since $g \in H$, the claim of the Theorem finally follows from $\int \phi dQ_g = \alpha$, result (C.20) and Lemma D.1 (in the Online Appendix). ■

Proof of Theorem 5.3: For any path $t \mapsto P_{t,g} \in \mathcal{M}$ we first note that applying Lemma E.2 (in the Online Appendix) with Assumption 5.2 in place of Assumption 3.1 implies

$$\hat{\mathbb{G}}_n \overset{L_{n,g}}{\to} \mathbb{G}_0 + \Delta_g, \tag{C.22}$$

for $\Delta_g \in \ell^\infty(\mathbf{T})$ given by $\Delta_g(\tau) \equiv \int s_\tau g dP$. Let $x \vee y \equiv \max\{x, y\}$. Define $\Delta_g^\omega \in \ell^\infty(\mathbf{T})$ to be $\Delta_g^\omega(\tau) \equiv \omega(\tau) \times \Delta_g(\tau)$, we then obtain from (C.22) and the continuous mapping theorem

$$\|\hat{\mathbb{G}}_n^\omega \vee 0\|_\infty \overset{L_{n,g}}{\to} \|(\mathbb{G}_0^\omega + \Delta_g^\omega) \vee 0\|_\infty. \tag{C.23}$$

Moreover, note that $\mathbb{G}_0^\omega$ is a regular measure by Theorem 7.1.7 in Bogachev (2007), and hence since $\mathbb{G}_0^\omega$ is also tight due to $\omega \in \ell^\infty(\mathbf{T})$ and $\mathbb{G}_0$ being tight by Assumption 5.2(ii), we conclude $\mathbb{G}_0^\omega$ is Radon. Together with the convexity of the map $\| \cdot \vee 0\|_\infty : \ell^\infty(\mathbf{T}) \to \mathbf{R}$, $\mathbb{G}_0^\omega$ being Radon allows us to apply Theorem 11.1 in Davydov et al. (1998) to obtain

$$c_0 \equiv \inf\{c : P(\|(\mathbb{G}_0^\omega + \Delta_g^\omega) \vee 0\|_\infty \leq c) > 0\} \tag{C.24}$$

is the only possible discontinuity point of the c.d.f. of $\|(\mathbb{G}_0^\omega + \Delta_g^\omega) \vee 0\|_\infty$. However, note that since $c_{1-\alpha}^\omega > 0$ by hypothesis, we must have $\|\omega\|_\infty > 0$, and therefore for any $c > 0$

$$P(\|(\mathbb{G}_0^\omega + \Delta_g^\omega) \vee 0\|_\infty \leq c) \geq P(\|\mathbb{G}_0 + \Delta_g\|_\infty \leq \frac{c}{\|\omega\|_\infty}) > 0 \tag{C.25}$$

where we exploited that $\omega \geq 0$, and the final inequality follows from Proposition 12.1 in Davydov et al. (1998) and $-\Delta_g = \Delta_{-g}$ belonging to the support of $\mathbb{G}_0$ by Lemma E.3 (in the Online Appendix).[10] Since $c_{1-\alpha}^\omega > 0$ by hypothesis, it follows from (C.24) and (C.25) that $c_{1-\alpha}^\omega$ is a continuity point of the c.d.f. of $\|(\mathbb{G}_0^\omega + \Delta_g^\omega) \vee 0\|_\infty$. Therefore, we obtain from (C.22) that

$$\lim_{n\to\infty} P_{1/\sqrt{n},g}(\|\hat{\mathbb{G}}_n^\omega \vee 0\|_\infty > c_{1-\alpha}^\omega) = P(\|(\mathbb{G}_0^\omega + \Delta_g^\omega) \vee 0\|_\infty > c_{1-\alpha}^\omega), \tag{C.26}$$

which verifies that ϕ_n^ω indeed has an asymptotic local power function. Moreover, note that if $t \mapsto P_{t,g} \in \mathbf{P}$, then $g \in \bar{T}(P)$ by definition and hence $\int s_\tau g dP \leq 0$ for all $\tau \in \mathbf{T}$ since $s_\tau \in \bar{T}(P)^-$. Thus, $\omega \geq 0$ implies $\Delta_g^\omega \leq 0$, and therefore (C.26) yields

$$\lim_{n\to\infty} P_{1/\sqrt{n},g}(\|\hat{\mathbb{G}}_n^\omega \vee 0\|_\infty > c_{1-\alpha}^\omega) \leq P(\|\mathbb{G}_0^\omega \vee 0\|_\infty > c_{1-\alpha}^\omega) = \alpha, \tag{C.27}$$

[10] Lemma E.3 requires Assumption 3.1 in place of Assumption 5.2, but the proof of Lemma E.3 also holds under the latter assumption with no modifications.

where we exploited that $c^{\omega}_{1-\alpha}$ is the $1-\alpha$ quantile of $\|\mathbb{G}^{\omega}_0 \vee 0\|_\infty$ and the c.d.f. of $\|\mathbb{G}^{\omega}_0 \vee 0\|_\infty$ is continuous at $c^{\omega}_{1-\alpha}$. Since (C.27) holds for any path $t \mapsto P_{t,g} \in \mathbf{P}$, we conclude ϕ^{ω}_n is indeed an asymptotic level α specification test.

To establish (63), note $\bar{C}(P) \subseteq L^2_0(P)$ is a closed convex cone by definition, and let $\bar{C}(P)^- \equiv \{g \in L^2_0(P) : \int gf dP \le 0 \text{ for all } f \in \bar{C}(P)\}$. For any $g \in L^2_0(P)$, Proposition 46.5.4 in Zeidler (1984) implies $g = \Pi_C(g) + \Pi_{C^-}(g)$ and $\int \{\Pi_C(g)\}\{\Pi_{C^-}(g)\}dP = 0$. In particular, if a path $t \mapsto P_{t,g} \in \mathcal{M}$ is such that $\Pi_C(g) \neq 0$, then

$$\int g\{\Pi_C(g)\}dP = \int \{\Pi_C(g) + \Pi_{C^-}(g)\}\{\Pi_C(g)\}dP = \int \{\Pi_C(g)\}^2 dP > 0. \tag{C.28}$$

Since $\Pi_C(g) \in \bar{C}(P)$ and $\bar{C}(P)$ is the closed convex cone generated by $\{s_\tau\}_{\tau \in \mathbf{T}}$, there exists an integer $K < \infty$, positive scalars $\{\alpha_k\}_{k=1}^K$, and $\{\tau_k\}_{k=1}^K \subseteq \mathbf{T}$ such that

$$\Big\|\Pi_C(g) - \sum_{k=1}^K \alpha_k s_{\tau_k}\Big\|_{P,2} < \frac{1}{2}\frac{\|\Pi_C(g)\|^2_{P,2}}{\|g\|_{P,2}}. \tag{C.29}$$

Therefore, results (C.28) and (C.29) together with the Cauchy-Schwarz inequality yield

$$\int g\{\sum_{k=1}^K \alpha_k s_{\tau_k}\}dP \ge \int g\{\Pi_C(g)\}dP - \Big|\int g\{\Pi_C(g) - \sum_{k=1}^K \alpha_k s_{\tau_k}\}dP\Big| \ge \frac{1}{2}\|\Pi_C(g)\|^2_{P,2} > 0. \tag{C.30}$$

Since $\alpha_k \ge 0$ for all $1 \le k \le K$, result (C.30) implies that $\int g s_{\tau^\star} dP > 0$ for some $\tau^\star \in \mathbf{T}$. To conclude, we then let $\omega^\star(\tau) \equiv 1\{\tau = \tau^\star\}$ and note $\mathbb{G}^{\omega^\star}_0(\tau^\star) \sim N(0, \int s^2_{\tau^\star} dP)$. Furthermore, since $\int s^2_{\tau^\star} dP > 0$ because $\int g s_{\tau^\star} dP > 0$, and $\|\mathbb{G}^{\omega^\star}_0 \vee 0\|_\infty = \max\{\mathbb{G}_0(\tau^\star), 0\}$ almost surely, it follows that $c^{\omega^\star}_{1-\alpha} > 0$ provided $\alpha \in (0, \frac{1}{2})$. We may then exploit result (C.26) since $c^{\omega^\star}_{1-\alpha} > 0$, and employ $\int g s_{\tau^\star} dP > 0$ to obtain

$$\lim_{n\to\infty} P_{1/\sqrt{n},g}\big(\|\hat{\mathbb{G}}^{\omega^\star}_n \vee 0\|_\infty > c^{\omega^\star}_{1-\alpha}\big) = P\big(\mathbb{G}_0(\tau^\star) + \int g s_{\tau^\star} dP > c^{\omega^\star}_{1-\alpha}\big) > \alpha, \tag{C.31}$$

which establishes the second claim of the Theorem. ■

Proof of Theorem 5.4: Let $R(\tau^\star) \equiv \{\lambda s_{\tau^\star} : \lambda \ge 0\}$ which is a closed convex cone and set $R(\tau^\star)^-$ to be the polar cone of $R(\tau^\star)$, which satisfies

$$R(\tau^\star)^- = \Big\{g \in L^2_0(P) : \int g s_{\tau^\star} dP \le 0\Big\}. \tag{C.32}$$

In addition, for any $g \in L^2_0(P)$ we let $\Pi_R(g)$ and $\Pi_{R^-}(g)$ denote the metric projections of g onto $R(\tau^\star)$ and $R(\tau^\star)^-$ respectively, and we note by direct calculation that

$$\Pi_R\Big(f - \int f dP\Big) = \beta(f, \tau^\star) \times s_{\tau^\star} \tag{C.33}$$

for any $f \in L^2(P)$. Moreover, by Proposition 46.5.4 in Zeidler (1984) it also follows that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n \{f(X_i) - \int f dP\} = \frac{1}{\sqrt{n}}\sum_{i=1}^n \Big\{\{\Pi_{R^-}(f - \int f dP)\}(X_i) + \beta(f, \tau^\star) s_{\tau^\star}(X_i)\Big\}, \tag{C.34}$$

where $\int \Pi_{R^-}(f - \int f dP)\beta(f,\tau^\star)s_{\tau^\star} dP = 0$. Let $\Delta_g \equiv \int \Pi_R(f - \int f dP) g dP$. Then we obtain from results (C.33) and (C.34), Assumption 5.2(i) and Theorem 3.10.12 in van der Vaart and Wellner (1996) that for any path $t \mapsto P_{t,g} \in \mathcal{M}$ we have

$$\Big(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\{f(X_i) - \int f dP\},\ \beta(f,\tau^\star)\hat{\mathbb{G}}_n(\tau^\star)\Big) \stackrel{L_{n,g}}{\to} \Big(\mathbb{G}_R + \mathbb{G}_{R^-} + \int f g dP,\ \mathbb{G}_R + \Delta_g\Big) \qquad (C.35)$$

where $(\mathbb{G}_R, \mathbb{G}_{R^-})$ are independent normals with $\text{Var}\{\mathbb{G}_R + \mathbb{G}_{R^-}\} = \text{Var}\{f(X_i)\}$ and $\text{Var}\{\mathbb{G}_R\} = \|\beta(f,\tau^\star)s_{\tau^\star}\|_{P,2}^2$. Moreover, for any bounded $f : \mathbf{X} \to \mathbf{R}$, Lemma F.1 (in the Online Appendix) implies

$$\Big|\sqrt{n}\int f(dP_{1/\sqrt{n},g} - dP) - \int f g dP\Big| = o(1). \qquad (C.36)$$

for any path $t \mapsto P_{t,g} \in \mathcal{M}$. Therefore, results (C.35) and (C.36), the definition of $\hat{\mu}_n(f,\tau^\star)$, and the continuous mapping theorem allow us to conclude

$$\sqrt{n}\Big\{\hat{\mu}_n(f,\tau^\star) - \int f dP_{1/\sqrt{n},g}\Big\} \stackrel{L_{n,g}}{\to} \mathbb{G}_{R^-} + \min\{\mathbb{G}_R, -\Delta_g\} \equiv \mathbb{Z}_g, \qquad (C.37)$$

which implies $\hat{\mu}_n(f,\tau^\star)$ indeed satisfies (60). We next aim to show that (65) holds for any path $t \mapsto P_{t,g} \in \mathbf{P}$ provided $\int f s_{\tau^\star} dP > 0$, which implies $\beta(f,\tau^\star) > 0$. We thus assume $\beta(f,\tau^\star) \neq 0$, and note this implies $\text{Var}\{\mathbb{G}_R\} > 0$. Hence, since $\mathbb{G}_0 = \mathbb{G}_R + \mathbb{G}_{R^-}$ by results (C.1) and (C.35), we can exploit the definition of $\mathbb{Z}_g$ in (C.37) to obtain for any $t > 0$ that

$$\begin{aligned} P(|\mathbb{Z}_g| \le t) &= P(|\mathbb{G}_0| \le t, \mathbb{G}_R \le -\Delta_g) + P(|\mathbb{G}_{R^-} - \Delta_g| \le t,\ \mathbb{G}_R > -\Delta_g) \\ &= P(|\mathbb{G}_0| \le t) + P(|\mathbb{G}_{R^-} - \Delta_g| \le t,\ \mathbb{G}_R > -\Delta_g) - P(|\mathbb{G}_0| \le t,\ \mathbb{G}_R > -\Delta_g). \end{aligned} \qquad (C.38)$$

Let $\sigma_{R^-}^2 \equiv \text{Var}\{\mathbb{G}_{R^-}\}$. Note that if $\sigma_{R^-}^2 = 0$, then (C.38) implies $P(|\mathbb{Z}_g| \le t) < P(|\mathbb{G}_0| \le t)$ whenever $t > |\Delta_g|$ due to $\text{Var}\{\mathbb{G}_R\} > 0$. If $\sigma_{R^-}^2 > 0$, then for Φ the c.d.f. of a standard normal random variable we can conclude from $\mathbb{G}_0 = \mathbb{G}_R + \mathbb{G}_{R^-}$ and the independence of $\mathbb{G}_R$ and $\mathbb{G}_{R^-}$ that

$$\begin{aligned} P\big(|\mathbb{G}_{R^-} - \Delta_g| \le t \mid \mathbb{G}_R + \Delta_g > 0\big) &= \Phi(\frac{t + \Delta_g}{\sigma_{R^-}}) - \Phi(\frac{-t + \Delta_g}{\sigma_{R^-}}) \\ P(|\mathbb{G}_0| \le t \mid \mathbb{G}_R + \Delta_g > 0) &= E\Big[\Phi(\frac{t - \mathbb{G}_R}{\sigma_{R^-}}) - \Phi(\frac{-t - \mathbb{G}_R}{\sigma_{R^-}}) \mid \mathbb{G}_R > -\Delta_g\Big]. \end{aligned} \qquad (C.39)$$

We note that the function $F_t(a) \equiv \Phi((t-a)/\sigma_{R^-}) - \Phi((-t-a)/\sigma_{R^-})$ is decreasing in $a \in [0,\infty)$ whenever $t \ge 0$. Since $s_{\tau^\star} \in \bar{T}(P)^-$, we have $\Delta_g \equiv \int\{\Pi_R(f - \int f dP)\} g dP \le 0$ whenever $g \in \bar{T}(P)$. It follows from (C.39) that for any $g \in \bar{T}(P)$ we have

$$P(|\mathbb{Z}_g| \le t) > P(|\mathbb{G}_0| \le t) \quad \text{for all} \ \ t > 0\ . \qquad (C.40)$$

Thus, since $\Psi(b) = \Psi(|b|)$, $\Psi(b) \ge \Psi(b')$ whenever $|b| \ge |b'|$, and Ψ is nonconstant, result (C.40) implies

$$E[\Psi(\mathbb{Z}_g)] < E[\Psi(\mathbb{G}_0)]. \qquad (C.41)$$

Since (C.41) holds for any $g \in \bar{T}(P)$, we then obtain from (C.37) and Definition 5.1(iii)

$$\limsup_{n\to\infty} E_{P_{1/\sqrt{n},g}}\left[\Psi\left(\sqrt{n}\{\hat{\mu}_n(f,\tau^\star) - \int f dP_{1/\sqrt{n},g}\}\right)\right] = E[\Psi(\mathbb{Z}_g)] < E[\Psi(\mathbb{G}_0)] \tag{C.42}$$

for any path $t \mapsto P_{t,g} \in \mathbf{P}$, which together with (C.2) establishes (65). ■

Online Appendix to

Overidentification in Regular Models

Xiaohong Chen Andres Santos

April 2018

This Online Supplementary Appendix contains additional results and proofs to support the main text. Appendix D contains the proof of the limiting experiment results in Appendix A and additional lemmas. Appendix E presents the technical lemmas and their proofs that are used in the proofs of Appendix B and Appendix C. Appendix F contains the proofs of the results in Section 4. Appendix G provides sufficient conditions for verifying Assumption 3.1 in the general nonparametric conditional moment restriction models studied in Section 4. Appendix H provides additional discussion of the Examples in Section 4 as well as a final example.

Appendix D - Proofs for Appendix A and Additional Lemmas

In this Appendix we provide the proofs of Theorem A.1 and additional technical lemmas.

Proof of Theorem A.1: To establish part (i), we first note that for any $g \in L^2_0(P)$ it is possible to construct a path $t \mapsto P_{t,g}$ whose score is g; see Example 3.2.1 in Bickel et al. (1993) for a concrete construction. Further, any two paths $t \mapsto \tilde{P}_{t,g}$ and $t \mapsto P_{t,g}$ with the same score $g \in L^2_0(P)$ satisfy

$$\lim_{n\to\infty} \Big| \int \phi_n d\tilde{P}^n_{1/\sqrt{n},g} - \int \phi_n dP^n_{1/\sqrt{n},g} \Big| \leq \lim_{n\to\infty} \int |dP^n_{1/\sqrt{n},g} - d\tilde{P}^n_{1/\sqrt{n},g}| = 0 \tag{D.1}$$

for any $0 \leq \phi_n \leq 1$ by Lemma D.1 (below). Thus, for each $g \in L^2_0(P)$ we may select an arbitrary path $t \mapsto P_{t,g}$ whose score is indeed g, and for $\mathcal{B}$ the σ-algebra on $\mathbf{X}$ we consider the sequence of experiments

$$\mathcal{E}_n \equiv \big(\mathbf{X}^n, \mathcal{B}^n, P^n_{1/\sqrt{n},g} : g \in L^2_0(P)\big). \tag{D.2}$$

Next, since $\{\psi^T_k\}_{k=1}^{d_T} \cup \{\psi^{T^\perp}_k\}_{k=1}^{d_{T^\perp}}$ forms an orthonormal basis for $L^2_0(P)$, we obtain from Lemma D.3 (below) that $\mathcal{E}_n$ converges weakly to the experiment $\mathcal{E}$ given by

$$\mathcal{E} \equiv \big(\mathbf{R}^{d_T} \times \mathbf{R}^{d_{T^\perp}}, \mathcal{A}^{d_T} \times \mathcal{A}^{d_{T^\perp}}, Q_g : g \in L^2_0(P)\big), \tag{D.3}$$

where $\mathcal{A}$ denotes the Borel σ-algebra on $\mathbf{R}$ and we exploited that for $d_P \equiv \dim\{L^2_0(P)\}$ we have $\mathbf{R}^{d_T} \times \mathbf{R}^{d_{T^\perp}} = \mathbf{R}^{d_P}$ and $\mathcal{A}^{d_T} \times \mathcal{A}^{d_{T^\perp}} = \mathcal{A}^{d_P}$. The existence of a test function $\phi : (\mathbb{Y}^T, \mathbb{Y}^{T^\perp}) \to [0,1]$ satisfying $\pi(g_0) = \int \phi dQ_{g_0}$ for all $g_0 \in L^2_0(P)$ then follows from Theorem 7.1 in van der Vaart (1991). To establish part (i) of the Theorem, it thus only remains to show that ϕ must control size in (A.2). To this end, note that $\Pi_{T^\perp}(g_0) = 0$ if and only if $g_0 \in \bar{T}(P)$. Fixing $\delta > 0$ then observe that for any $g_0 \in \bar{T}(P)$ there exists a $\tilde{g} \in T(P)$ such that $\|g_0 - \tilde{g}\|_{P,2} < \delta$. Moreover, since

$\tilde{g} \in T(P)$, there exists a path $t \mapsto \tilde{P}_{t,\tilde{g}} \in \mathbf{P}$ with score $\tilde{g}$ and hence we can conclude that

$$\begin{aligned} \int \phi dQ_{g_0} &= \lim_{n\to\infty} \int \phi_n dP^n_{1/\sqrt{n},g_0} \\ &\leq \lim_{n\to\infty} \int \phi_n d\tilde{P}^n_{1/\sqrt{n},\tilde{g}} + \limsup_{n\to\infty} \int |dP^n_{1/\sqrt{n},g_0} - d\tilde{P}^n_{1/\sqrt{n},\tilde{g}}| \\ &\leq \alpha + 2\{1 - \exp\{-\frac{\delta^2}{4}\}\}^{1/2}, \end{aligned} \quad \text{(D.4)}$$

where the first inequality employed $0 \leq \phi_n \leq 1$, and the second inequality exploited Lemma D.1 and that ϕ_n is a local asymptotic level α test. Since $\delta > 0$ is arbitrary, we conclude from (11) and (D.4) that $\pi(g_0) = \int \phi dQ_{g_0} \leq \alpha$ whenever $g_0 \in \bar{T}(P)$, and hence part (i) of the Theorem follows.

For part (ii) of the Theorem, we first note that since $T(P)$ is linear by Assumption 2.1(ii), and $\hat{\theta}_n$ is regular by hypothesis, Lemma D.4 (below) and Theorem 5.2.3 in Bickel et al. (1993) imply θ is pathwise differentiable at P – i.e. there exists a bounded linear operator $\dot{\theta} : \bar{T}(P) \to \mathbf{B}$ such that for any submodel $t \mapsto P_{t,g} \in \mathbf{P}$ it follows

$$\lim_{t\downarrow 0} \|t^{-1}\{\theta(P_{t,g}) - \theta(P)\} - \dot{\theta}(g)\|_{\mathbf{B}} = 0. \quad \text{(D.5)}$$

Then note that for any $b^* \in \mathbf{B}^*$, $b^* \circ \dot{\theta} : \bar{T}(P) \to \mathbf{R}$ is a continuous linear functional. Hence, since $\bar{T}(P)$ is a Hilbert space under $\|\cdot\|_{P,2}$, the Riesz representation theorem implies there exists a $\dot{\theta}_{b^*} \in \bar{T}(P)$ such that for all $g \in \bar{T}(P)$ we have

$$b^*(\dot{\theta}(g)) = \int \dot{\theta}_{b^*} g dP. \quad \text{(D.6)}$$

Moreover, since $\hat{\theta}_n$ is an asymptotically linear regular estimator of $\theta(P)$, it follows that $b^*(\hat{\theta}_n)$ is an asymptotically linear regular estimator of $b^*(\theta(P))$ with influence function $b^* \circ \nu$. Proposition 3.3.1 in Bickel et al. (1993) then implies that for all $g \in \bar{T}(P)$

$$\int (\dot{\theta}_{b^*} - b^* \circ \nu) g dP = 0. \quad \text{(D.7)}$$

In particular, (D.7) implies that $\dot{\theta}_{b^*} = \Pi_T(b^* \circ \nu)$, and therefore by asymptotic linearity

$$\sqrt{n}\{b^*(\hat{\theta}_n) - b^*(\theta(P))\} \xrightarrow{L} N\big(0,\ \|\dot{\theta}_{b^*}\|^2_{P,2} + \|\Pi_{T^\perp}(b^* \circ \nu)\|^2_{P,2}\big) \quad \text{(D.8)}$$

where we have exploited the central limit theorem and $b^* \circ \nu = \Pi_T(b^* \circ \nu) + \Pi_{T^\perp}(b^* \circ \nu)$. To conclude, we next define the maps $F^T(\mathbb{Y}^T)$ and $F^{T^\perp}(\mathbb{Y}^{T^\perp})$ to be given by

$$\begin{aligned} F^T(\mathbb{Y}^T) &= \sum_{k=1}^{d_T} \mathbb{Y}^T_k \int \{\dot{\theta}_{b^*}\} \psi^T_k dP \\ F^{T^\perp}(\mathbb{Y}^{T^\perp}) &= \sum_{k=1}^{d_{T^\perp}} \mathbb{Y}^{T^\perp}_k \int \{\Pi_{T^\perp}(b^* \circ \nu)\} \psi^{T^\perp}_k dP. \end{aligned} \quad \text{(D.9)}$$

We aim to show that if $(\mathbb{Y}^T, \mathbb{Y}^{T^\perp}) \sim Q_{g_0}$ with $g_0 = 0$, then $F^T(\mathbb{Y}^T) \sim N(0, \|\dot{\theta}_{b^*}\|^2_{P,2})$ which is immediate if $d_T < \infty$, and thus we assume $d_T = \infty$. Defining the partial sums

$$\mathbb{V}_K \equiv \sum_{k=1}^{K} \mathbb{Y}_k^T \int \{\dot{\theta}_{b^*}\}\psi_k^T dP \tag{D.10}$$

we then observe $\mathbb{V}_K \sim N(0, \sigma_K^2)$ where $\sigma_K^2 \equiv \sum_{k=1}^K \int\{\int \dot{\theta}_{b^*}\psi_k^T\}^2 dP$ and $\sigma_K^2 \uparrow \|\dot{\theta}_{b^*}\|_{P,2}^2$ by Parseval's identity. By the martingale convergence theorem, see, e.g., Theorem 12.1.1 in Williams (1991), it follows $\mathbb{V}_K$ converges almost surely and thus that $F^T(\mathbb{Y}^T)$ is well defined. Moreover, for any continuous bounded function $f : \mathbf{R} \to \mathbf{R}$ it follows

$$\begin{aligned} E[f(F^T(\mathbb{Y}^T))] = \lim_{K\to\infty} E[f(\mathbb{V}_K)] &= \lim_{K\to\infty} \frac{1}{\sqrt{2\pi}\sigma_K} \int f(z)\exp\{-\frac{z^2}{2\sigma_K^2}\}dz \\ &= \frac{1}{\sqrt{2\pi}\|\dot{\theta}_{b^*}\|_{P,2}} \int f(z)\exp\{-\frac{z^2}{2\|\dot{\theta}_{b^*}\|_{P,2}^2}\}dz \end{aligned} \tag{D.11}$$

due to $\sigma_K^2 \uparrow \|\dot{\theta}_{b^*}\|_{P,2}^2$. We conclude from (D.11) that $F^T(\mathbb{Y}^T) \sim N(0, \|\dot{\theta}_{b^*}\|_{P,2}^2)$ when $(\mathbb{Y}^T, \mathbb{Y}^{T^\perp}) \sim Q_{g_0}$ with $g_0 = 0$. Identical arguments imply $F^{T^\perp}(\mathbb{Y}^{T^\perp}) \sim N(0, \|\Pi_{T^\perp}(b^* \circ \nu)\|_{P,2}^2)$. Thus part (ii) of the Theorem follows from (D.8) and independence of $\mathbb{Y}^T$ and $\mathbb{Y}^{T^\perp}$. ■

Lemma D.1. *If $t \mapsto P_{t,g_1}$ and $t \mapsto P_{t,g_2}$ are arbitrary paths, then it follows that:*

$$\limsup_{n\to\infty} \int |dP^n_{1/\sqrt{n},g_1} - dP^n_{1/\sqrt{n},g_2}| \leq 2\{1 - \exp\{-\frac{1}{4}\|g_1 - g_2\|_{P,2}^2\}\}^{1/2}. \tag{D.12}$$

Proof of Lemma D.1: Since $t \mapsto P_{t,g_1}$ and $t \mapsto P_{t,g_2}$ satisfy (1), we must have

$$\lim_{n\to\infty} n \int [dP^{1/2}_{1/\sqrt{n},g_1} - dP^{1/2}_{1/\sqrt{n},g_2}]^2 = \frac{1}{4}\int [g_1 dP^{1/2} - g_2 dP^{1/2}]^2 = \frac{1}{4}\|g_1 - g_2\|_{P,2}^2. \tag{D.13}$$

Moreover, by Theorem 13.1.2 in Lehmann and Romano (2005) we can also conclude

$$\begin{aligned} \frac{1}{2}\int |dP^n_{1/\sqrt{n},g_1} - dP^n_{1/\sqrt{n},g_2}| &\leq \{1 - [\int \{dP^n_{1/\sqrt{n},g_1}\}^{1/2}\{dP^n_{1/\sqrt{n},g_2}\}^{1/2}]^2\}^{1/2} \\ = \{1 - [\int dP^{1/2}_{1/\sqrt{n},g_1} dP^{1/2}_{1/\sqrt{n},g_2}]^{2n}\}^{1/2} &= \{1 - [1 - \frac{1}{2}\int [dP^{1/2}_{1/\sqrt{n},g_1} - dP^{1/2}_{1/\sqrt{n},g_2}]^2]^{2n}\}^{1/2} \end{aligned} \tag{D.14}$$

where in the first equality we exploited $P^n_{1/\sqrt{n},g_1}$ and $P^n_{1/\sqrt{n},g_2}$ are product measures, while the second equality follows from direct calculation. Thus, by (D.13) and (D.14)

$$\begin{aligned} \limsup_{n\to\infty} \frac{1}{2}\int |dP^n_{1/\sqrt{n},g_1} - dP^n_{1/\sqrt{n},g_2}| \\ &\leq \limsup_{n\to\infty}\{1 - [1 - \frac{1}{2n}\int n[dP^{1/2}_{1/\sqrt{n},g_1} - dP^{1/2}_{1/\sqrt{n},g_2}]^2]^{2n}\}^{1/2} \\ &= \{1 - \exp\{-\frac{1}{4}\|g_1 - g_2\|_{P,2}^2\}\}^{1/2} \end{aligned} \tag{D.15}$$

which establishes the claim of the Lemma. ■

Lemma D.2. *Let $\{P_n\}$, $\{Q_n\}$, $\{V_n\}$ be probability measures defined on a common space. If*

$\{dQ_n/dP_n\}$ is asymptotically tight under P_n and $\int |dP_n - dV_n| = o(1)$, then

$$|\frac{dQ_n}{dP_n} - \frac{dQ_n}{dV_n}| \overset{P_n}{\to} 0. \tag{D.16}$$

Proof of Lemma D.2: Throughout let $\mu_n = P_n + Q_n + V_n$, note μ_n dominates P_n, Q_n, and V_n, and set $p_n \equiv dP_n/d\mu_n$, $q_n \equiv dQ_n/d\mu_n$, and $v_n \equiv dV_n/d\mu_n$. We then obtain

$$\int |\frac{dP_n}{dV_n} - 1| dV_n = \int |\frac{p_n}{v_n} - 1| v_n d\mu_n = \int_{v_n>0} |\frac{p_n}{v_n} - \frac{v_n}{v_n}| v_n d\mu_n$$
$$\leq \int |p_n - v_n| d\mu_n = \int |dP_n - dV_n| = o(1), \tag{D.17}$$

where the second to last equality follows by definition, and the final equality by assumption. Hence, by (D.17) and Markov's inequality we obtain $dP_n/dV_n \overset{V_n}{\to} 1$. Moreover, since $\int |dV_n - dP_n| = o(1)$ implies $\{P_n\}$ and $\{V_n\}$ are mutually contiguous, we conclude

$$\frac{dP_n}{dV_n} \overset{P_n}{\to} 1. \tag{D.18}$$

Next observe that for any continuous and bounded function $f : \mathbf{R} \to \mathbf{R}$ we have that

$$\int f(\frac{dQ_n}{dP_n} - \frac{dQ_n}{dV_n}) dP_n = \int f(\frac{q_n}{p_n} - \frac{q_n}{v_n}) p_n d\mu_n$$
$$= \int_{p_n>0} f(\frac{q_n}{p_n}(1 - \frac{p_n}{v_n})) p_n d\mu_n = \int f(\frac{dQ_n}{dP_n}(1 - \frac{dP_n}{dV_n})) dP_n \to f(0), \tag{D.19}$$

where the final result follows from (D.18), dQ_n/dP_n being asymptotically tight under P_n and continuity and boundedness of f. Since (D.19) holds for any continuous and bounded f, we conclude $dQ_n/dP_n - dQ_n/dV_n$ converges in law (under P_n) to zero, and hence also in P_n probability, which establishes (D.16). ■

Lemma D.3. *Let $H \subseteq L^2_0(P)$, assume for each $g \in H$ there is a path $t \mapsto P_{t,g}$ with score g, recall $\mathcal{B}$ is the σ-algebra on $\mathbf{X}$, let $\mathcal{A}$ be the Borel σ-algebra on $\mathbf{R}$, and set*

$$\mathcal{E}_n \equiv \big(\mathbf{X}^n, \mathcal{B}^n, P^n_{1/\sqrt{n},g} : g \in H\big). \tag{D.20}$$

If $0 \in H$, $\{\psi_k\}_{k=1}^{d_P}$ is an orthonormal basis for $L^2_0(P)$, and Φ denotes the standard normal measure on $\mathbf{R}$, then $\mathcal{E}_n$ converges weakly to the dominated experiment $\mathcal{E}$

$$\mathcal{E} \equiv \big(\mathbf{R}^{d_P}, \mathcal{A}^{d_P}, Q_g : g \in H\big), \tag{D.21}$$

where for each $g \in H$, $Q_g(\cdot) = Q_0(\cdot - T(g))$ for $T(g) \equiv \{\int g\psi_k dP\}_{k=1}^{d_P}$ and $Q_0 = \bigotimes_{k=1}^{d_P} \Phi$.

Proof of Lemma D.3: The conclusion of the Lemma is well known (see e.g. Subsection 8.2 in van der Vaart (1991)), but we were unable to find a concrete reference and hence we include its proof for completeness. Since the Lemma is straightforward when the dimension of $L^2_0(P)$ is finite

($d_P < \infty$) we focus on the case $d_P = \infty$. To analyze $\mathcal{E}$, let

$$\ell^2 \equiv \{\{c_k\}_{k=1}^{\infty} \in \mathbf{R}^{\infty} : \sum_{k=1}^{\infty} c_k^2 < \infty\}, \tag{D.22}$$

and note that by Example 2.3.5 in Bogachev (1998), ℓ^2 is the Cameron-Martin space of Q_0.[1] Hence, since for any $g \in L_0^2(P)$ we have $\{\int g\psi_k dP\}_{k=1}^{\infty} \in \ell^2$ due to $\{\psi_k\}_{k=1}^{\infty}$ being an orthonormal basis for $L_0^2(P)$, Theorem 2.4.5 in Bogachev (1998) implies

$$Q_g \equiv Q_0(\cdot - T(g)) \ll Q_0 \tag{D.23}$$

for all $g \in L_0^2(P)$, and thus $\mathcal{E}$ is dominated by Q_0. Denoting an element of $\mathbf{R}^{\infty}$ by $\omega = \{\omega_k\}_{k=1}^{\infty}$, we then obtain from $\{\int g\psi_k dP\}_{k=1}^{\infty} \in \ell^2$ and the Martingale convergence theorem, see for example Theorem 12.1.1 in Williams (1991), that

$$Q_0(\omega : \lim_{n\to\infty} \sum_{k=1}^{n} \omega_k \int \psi_k g dP \text{ exists}) = 1 \tag{D.24}$$

$$\lim_{n\to\infty} \int (\sum_{k=n+1}^{\infty} \omega_k \int g\psi_k dP)^2 dQ_0(\omega) = 0. \tag{D.25}$$

Therefore, Example 2.3.5 and Corollary 2.4.3 in Bogachev (1998) yield for any $g \in L_0^2(P)$

$$\begin{aligned} \log(\frac{dQ_g}{dQ_0}(\omega)) &= \sum_{k=1}^{\infty} \omega_k \int g\psi_k dP - \frac{1}{2}\int (\sum_{k=1}^{\infty} \omega_k \int g\psi_k dP)^2 dQ_0(\omega) \\ &= \sum_{k=1}^{\infty} \omega_k \int g\psi_k dP - \frac{1}{2}\int g^2 dP, \end{aligned} \tag{D.26}$$

where the right hand side of the first equality is well defined Q_0 almost surely by (D.24), while the second equality follows from (D.25) and $\sum_{k=1}^{\infty}(\int g\psi_k dP)^2 = \int g^2 dP$ due to $\{\psi_k\}_{k=1}^{\infty}$ being an orthonormal basis for $L_0^2(P)$.

Next, select an arbitrary finite subset $\{g_j\}_{j=1}^{J} \equiv I \subseteq H$ and vector $(\lambda_1, \ldots, \lambda_J)' \equiv \lambda \in \mathbf{R}^J$. From result (D.26) we then obtain Q_0 almost surely that

$$\sum_{j=1}^{J} \lambda_j \log(\frac{dQ_{g_j}}{dQ_0}(\omega)) = \sum_{k=1}^{\infty} \omega_k \int (\sum_{j=1}^{J} \lambda_j g_j)\psi_k dP - \sum_{j=1}^{J} \frac{\lambda_j}{2} \int g_j^2 dP. \tag{D.27}$$

In particular, we can conclude from Example 2.10.2 and Proposition 2.10.3 in Bogachev (1998) together with (D.25) and $\sum_{j=1}^{J} \lambda_j g_j \in L_0^2(P)$ that, under Q_0, we have

$$\sum_{j=1}^{J} \lambda_j \log(\frac{dQ_j}{dQ_0}) \sim N\Big(-\sum_{j=1}^{J} \frac{\lambda_j}{2} \int g_j^2 dP, \int (\sum_{j=1}^{J} \lambda_j g_j)^2 dP\Big). \tag{D.28}$$

[1] See page 44 in Bogachev (1998) for a definition of a Cameron Martin space.

Thus, for $\mu_I \equiv \frac{1}{2}(\int g_1^2 dP, \ldots, \int g_J^2 dP)'$ and $\Sigma_I \equiv \int (g_1, \ldots, g_J)'(g_1, \ldots, g_J) dP$, we have

$$\Big(\log(\frac{dQ_{g_1}}{dQ_0}), \ldots, \log(\frac{dQ_{g_J}}{dQ_0})\Big)' \sim N(-\mu_I, \Sigma_I), \tag{D.29}$$

under Q_0 due to (D.28) holding for arbitrary $\lambda \in \mathbf{R}^J$.

To obtain an analogous result for the sequence of experiments $\mathcal{E}_n$, let $\{X_i\}_{i=1}^n \sim P^n$. From Lemma 25.14 in van der Vaart (1998) we obtain under P^n

$$\sum_{i=1}^n \log\Big(\frac{dP_{1/\sqrt{n}, g_j}}{dP}(X_i)\Big) = \frac{1}{\sqrt{n}} \sum_{i=1}^n g_j(X_i) - \frac{1}{2}\int g_j^2 dP + o_p(1) \tag{D.30}$$

for any $1 \leq j \leq J$. Thus, defining $P^n_{1/\sqrt{n}, g_j} \equiv \bigotimes_{i=1}^n P_{1/\sqrt{n}, g_j}$, we can conclude that

$$\Big(\log(\frac{dP^n_{1/\sqrt{n}, g_1}}{dP^n}), \ldots, \log(\frac{dP^n_{1/\sqrt{n}, g_J}}{dP^n})\Big)' \overset{L}{\to} N(-\mu_I, \Sigma_I), \tag{D.31}$$

under P^n by (D.30), the central limit theorem, and the definitions of μ_I and Σ_I. Furthermore, also note Lemma D.1 implies $\int |dP^n - dP^n_{1/\sqrt{n},0}| = o(1)$ and hence

$$\Big(\frac{dP^n_{1/\sqrt{n}, g_1}}{dP^n}, \ldots, \frac{dP^n_{1/\sqrt{n}, g_J}}{dP^n}\Big)' = \Big(\frac{dP^n_{1/\sqrt{n}, g_1}}{dP^n_{1/\sqrt{n},0}}, \ldots, \frac{dP^n_{1/\sqrt{n}, g_J}}{dP^n_{1/\sqrt{n},0}}\Big)' + o_p(1) \tag{D.32}$$

under P^n by Lemma D.2 and result (D.31). Thus, by (D.31) and (D.32) we obtain

$$\Big(\log(\frac{dP^n_{1/\sqrt{n}, g_1}}{dP^n_{1/\sqrt{n},0}}), \ldots, \log(\frac{dP^n_{1/\sqrt{n}, g_J}}{dP^n_{1/\sqrt{n},0}})\Big)' \overset{L}{\to} N(-\mu_I, \Sigma_I), \tag{D.33}$$

under P^n, and since $\int |dP^n - dP^n_{1/\sqrt{n},0}| = o(1)$ also under $P^n_{1/\sqrt{n},0}$. Hence, the Lemma follows from (i) (D.29), (ii) (D.33), and (iii) $\{P^n_{1/\sqrt{n},g}\}$ and $\{P^n_{1/\sqrt{n},0}\}$ being mutually contiguous for any $g \in H$ by (D.30) and Corollary 12.3.1 in Lehmann and Romano (2005), which together verify the conditions of Lemma 10.2.1 in LeCam (1986). ■

Lemma D.4. *Let Assumption 2.1(i) hold,* $\mathbf{B}$ *be a Banach space, and* $\hat{\theta}_n$ *be an asymptotically linear estimator for* $\theta(P) \in \mathbf{B}$ *such that* $\sqrt{n}\{\hat{\theta}_n - \theta(P)\} \overset{L}{\to} \mathbb{D}$ *under* P^n *on* $\mathbf{B}$ *for some tight Borel* $\mathbb{D}$*. Then: for any function* $h \in L_0^2(P)$*,* $\big(\sqrt{n}\{\hat{\theta}_n - \theta(P)\}, \frac{1}{\sqrt{n}}\sum_{i=1}^n h(X_i)\big)$ *converges in distribution under* P^n *on* $\mathbf{B} \times \mathbf{R}$*.*

Proof of Lemma D.4: For notational simplicity, let $\eta(P) \equiv (\theta(P), 0) \in \mathbf{B} \times \mathbf{R}$ and define $\hat{\eta}_n \equiv (\hat{\theta}_n, \frac{1}{n}\sum_{i=1}^n h(X_i)) \in \mathbf{B} \times \mathbf{R}$. Further let $(\mathbf{B} \times \mathbf{R})^*$ denote the dual space of $\mathbf{B} \times \mathbf{R}$ and note that for any $d^* \in (\mathbf{B} \times \mathbf{R})^*$ there are $b^*_{d^*} \in \mathbf{B}^*$ and $r^*_{d^*} \in \mathbf{R}$ such that $d^*((b, r)) = b^*_{d^*}(b) + r^*_{d^*}(r)$ for all $(b, r) \in \mathbf{B} \times \mathbf{R}$. For ν the influence function of $\hat{\theta}_n$ then define $\zeta_{d^*}(X_i) \equiv \{b^*_{d^*}(\nu(X_i)) + r^*_{d^*}(h(X_i))\}$ to obtain that under $\bigotimes_{i=1}^n P$ we have

$$d^*(\sqrt{n}\{\hat{\eta}_n - \eta(P)\}) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \zeta_{d^*}(X_i) + o_p(1) \tag{D.34}$$

by asymptotic linearity of $\hat{\theta}_n$. Thus, for any finite set $\{d_k^*\}_{k=1}^K \subset (\mathbf{B} \times \mathbf{R})^*$, we have

$$(d_1^*(\sqrt{n}\{\hat{\eta}_n - \eta(P)\}), \ldots, d_K^*(\sqrt{n}\{\hat{\eta}_n - \eta(P)\})) \xrightarrow{L} (\mathbb{W}_{d_1^*}, \ldots, \mathbb{W}_{d_K^*}) \tag{D.35}$$

for $(\mathbb{W}_{d_1^*}, \ldots, \mathbb{W}_{d_K^*})$ a multivariate normal random variable satisfying $E[\mathbb{W}_{d_k^*}] = 0$ for all $1 \leq k \leq K$ and $E[\mathbb{W}_{d_j^*}\mathbb{W}_{d_k^*}] = E[\zeta_{d_j^*}(X_i)\zeta_{d_k^*}(X_i)]$ for all $1 \leq j \leq k \leq K$.

Next note that since $\sqrt{n}\{\hat{\theta}_n - \theta(P)\}$ is asymptotically measurable and asymptotically tight by Lemma 1.3.8 in van der Vaart and Wellner (1996), it follows that $\sqrt{n}\{\hat{\eta}_n - \eta(P)\}$ is asymptotically measurable and asymptotically tight on $\mathbf{B} \times \mathbf{R}$ by Lemmas 1.4.3 and 1.4.4 in van der Vaart and Wellner (1996). Hence, we conclude by Theorem 1.3.9 in van der Vaart and Wellner (1996) that any sequence $\{n_k\}$ has a subsequence $\{n_{k_j}\}$ with

$$\sqrt{n}_{k_j}\{\hat{\eta}_{n_{k_j}} - \eta(P)\} \xrightarrow{L} \mathbb{W} \tag{D.36}$$

under $\bigotimes_{i=1}^{n_{k_j}} P$ for $\mathbb{W}$ some tight Borel Law on $\mathbf{B} \times \mathbf{R}$. However, letting $C_b(\mathbf{R}^K)$ denote the set of continuous and bounded functions on $\mathbf{R}^K$, we obtain from (D.35), (D.36), and the continuous mapping theorem that for any $\{d_k^*\}_{k=1}^K \subset (\mathbf{B} \times \mathbf{R})^*$ and $f \in C_b(\mathbf{R}^K)$

$$E[f((d_1^*(\mathbb{W}), \ldots, d_K^*(\mathbb{W})))] = E[f((\mathbb{W}_{d_1^*}, \ldots, \mathbb{W}_{d_K^*}))]. \tag{D.37}$$

Since $\mathcal{G} \equiv \{f \circ (d_1^*, \ldots, d_K^*) : f \in C_b(\mathbf{R}^K), \{d_k^*\}_{k=1}^K \subset (\mathbf{B} \times \mathbf{R})^*, 1 \leq K < \infty\}$ is a vector lattice that separates points in $\mathbf{B} \times \mathbf{R}$, Lemma 1.3.12 in van der Vaart and Wellner (1996) implies there is a unique tight Borel measure $\mathbb{W}$ on $\mathbf{B} \times \mathbf{R}$ satisfying (D.37). Thus, since the original sequence $\{n_k\}$ was arbitrary, we conclude all limit points of the law of $\sqrt{n}\{\hat{\eta}_n - \eta(P)\}$ coincide, and the Lemma follows. ■

Appendix E - Technical Lemmas Used in Appendix B and Appendix C

In this Appendix we present technical lemmas that are used in Appendix B and Appendix C.

Lemma E.1. *If $\mathbb{Z}_n \in \mathbf{B}$ is asymptotically tight and asymptotically measurable and satisfies $b^*(\mathbb{Z}_n) \xrightarrow{p} 0$ for any $b^* \in \mathbf{B}^*$, then it follows $\mathbb{Z}_n = o_p(1)$ in $\mathbf{B}$.*

Proof of Lemma E.1: For an arbitrary subsequence $\{n_j\}_{j=1}^{\infty}$, Theorem 1.3.9(ii) in van der Vaart and Wellner (1996) implies there exists a further subsequence $\{n_{j_k}\}_{k=1}^{\infty}$ along which $\mathbb{Z}_{n_{j_k}}$ converges in distribution to a tight limit $\mathbb{Z}$. Moreover, note that by the continuous mapping theorem $b^*(\mathbb{Z}) = 0$ for all $b^* \in \mathbf{B}^*$. Therefore, letting $C_b(\mathbf{R}^K)$ denote the set of bounded and continuous functions on $\mathbf{R}^K$ and defining $\mathcal{G} \equiv \{f \circ (b_1^*, \ldots, b_K^*) : f \in C_b(\mathbf{R}^K), \{b_k^*\}_{k=1}^K \subset \mathbf{B}^*, 1 \leq K < \infty\}$ we then obtain for any $g \in \mathcal{G}$

$$E[g(\mathbb{Z})] = g(0). \tag{E.1}$$

In particular, since $\mathcal{G}$ is a vector lattice that contains the constant functions and separates points in $\mathbf{B}$, Lemma 1.3.12 in van der Vaart and Wellner (1996) implies $\mathbb{Z} = 0$ almost surely. We conclude that $\mathbb{Z}_{n_{j_k}}$ converges in probability to zero along $\{n_{j_k}\}_{k=1}^{\infty}$. Thus, since the original subsequence $\{n_j\}_{j=1}^{\infty}$ was arbitrary, it follows that $\mathbb{Z}_n = o_p(1)$. ■

Lemma E.2. *Let Assumptions 2.1(i) and 3.1 hold, and for any $g \in L_0^2(P)$ define $\Delta_g : \mathbf{T} \to \mathbf{R}$ to be given by $\Delta_g(\tau) \equiv \int s_\tau g dP$. It then follows that for any path $t \mapsto P_{t,g} \in \mathcal{M}$,*

$$\hat{\mathbb{G}}_n \stackrel{L_{n,g}}{\to} \mathbb{G}_0 + \Delta_g \quad in \quad \ell^\infty(\mathbf{T}). \tag{E.2}$$

Further under Assumption 2.1(ii), $\Delta_g = 0$ whenever $\Pi_S(g) = 0$ where $S(P) = \{s_\tau \in \bar{T}(P)^\perp : \tau \in \mathbf{T}\}$.

Proof of Lemma E.2: We first note Lemma 25.14 in van der Vaart (1998) implies

$$\sum_{i=1}^n \log\Big(\frac{dP_{1/\sqrt{n},g}}{dP}(X_i)\Big) = \frac{1}{\sqrt{n}}\sum_{i=1}^n g(X_i) - \frac{1}{2}\int g^2 dP + o_p(1) \tag{E.3}$$

under P^n for any path $t \mapsto P_{t,g} \in \mathcal{M}$. Thus, by Example 3.10.6 in van der Vaart and Wellner (1996), P^n and $P^n_{1/\sqrt{n},g}$ are mutually contiguous, and hence applying Lemma D.4 and Lemma A.8.6 in Bickel et al. (1993) yields that for any path $t \mapsto P_{t,g} \in \mathcal{M}$

$$\hat{\mathbb{G}}_n \stackrel{L_{n,g}}{\to} \mathbb{G}_0 + \Delta_g \quad \text{in} \quad \ell^\infty(\mathbf{T}), \tag{E.4}$$

which establishes (E.2). Moreover, if $t \mapsto P_{t,g} \in \mathbf{P}$, then by definition $g \in T(P)$ and hence $\int g s_\tau dP = 0$ for all $\tau \in \mathbf{T}$ due to $s_\tau \in \bar{T}(P)^\perp$ by Assumption 3.1(i). More generally, $\Delta_g = 0$ for any path $t \mapsto P_{t,g} \in \mathcal{M}$ with $\Pi_S(g) = 0$. ■

Lemma E.3. *Let Assumptions 2.1(i) and 3.1 hold, and for any $g \in L_0^2(P)$ define $\Delta_g : \mathbf{T} \to \mathbf{R}$ by $\Delta_g(\tau) \equiv \int s_\tau g dP$. It then follows that Δ_g is in the support of $\mathbb{G}_0$.*

Proof of Lemma E.3: Define $\mathcal{S} \equiv S(P) = \{s_\tau : \tau \in \mathbf{T}\}$ and let the map $B : \mathbf{T} \to \mathcal{S}$ be given by $B(\tau) = s_\tau$ for any $\tau \in \mathbf{T}$. In addition, for any $s \in \mathcal{S}$ we define a selection $E : \mathcal{S} \to \mathbf{T}$ that assigns to each $s \in \mathcal{S}$ a unique element $E(s) \in B^{-1}(s)$. Our first goal is to show

$$P\Big(\sup_{s\in\mathcal{S}} \sup_{\tau \in B^{-1}(s)} |\mathbb{G}_0(\tau) - \mathbb{G}_0(E(s))| = 0\Big) = 1, \tag{E.5}$$

and to this end we fix $\epsilon, \eta > 0$, and note that since $\mathbb{G}_0$ is tight by Assumption 3.1(ii) we obtain by Lemma 1.3.8 in van der Vaart and Wellner (1996) that $\hat{\mathbb{G}}_n$ is asymptotically tight. Thus, since $\mathbb{G}_0$ is Gaussian, Theorem 1.5.7 in van der Vaart and Wellner (1996) implies that for any $\epsilon, \eta > 0$ there exists a $\delta(\epsilon, \eta) > 0$ such that

$$\limsup_{n\to\infty} P\Big(\sup_{\tau_1,\tau_2 : \|s_{\tau_1} - s_{\tau_2}\|_{P,2} > \delta(\epsilon,\eta)} |\hat{\mathbb{G}}_n(\tau_1) - \hat{\mathbb{G}}_n(\tau_2)| > \epsilon\Big) < \eta. \tag{E.6}$$

Moreover, since $\|s_{E(s)} - s_\tau\|_{P,2} = 0$ for any $\tau \in B^{-1}(s)$, by the Portmanteau Theorem, see e.g.

Theorem 1.3.4(ii) in van der Vaart and Wellner (1996), we obtain

$$P\left(\sup_{s\in\mathcal{S}}\sup_{\tau\in B^{-1}(s)}|\mathbb{G}_0(\tau)-\mathbb{G}_0(E(s))|>\epsilon\right)$$
$$\leq P\left(\sup_{\tau_1,\tau_2:\|s_{\tau_1}-s_{\tau_2}\|_{P,2}>\delta(\epsilon,\eta)}|\mathbb{G}_0(\tau_1)-\mathbb{G}_0(\tau_2)|>\epsilon\right)<\eta. \quad \text{(E.7)}$$

In particular, since $\epsilon,\eta>0$ were arbitrary, we conclude that (E.5) holds by result (E.7) and the monotone convergence theorem.

Next, define a map $\Upsilon:\ell^\infty(\mathbf{T})\to\ell^\infty(\mathcal{S})$ to be given by $\Upsilon(f)(s)=f(E(s))$ for any $f\in\ell^\infty(\mathbf{T})$, and note that Υ is linear and continuous. Thus, setting $\mathbb{S}_0=\Upsilon(\mathbb{G}_0)$, we note $\mathbb{S}_0$ is a tight Gaussian process on $\ell^\infty(\mathcal{S})$, which by Assumption 3.1(i) satisfies

$$E[\mathbb{S}_0(s_1)\mathbb{S}_0(s_2)]=\int s_1s_2dP \quad \text{(E.8)}$$

for any $s_1,s_2\in\mathcal{S}$. Similarly, let $S_g\in\ell^\infty(\mathcal{S})$ be given by $S_g\equiv\Upsilon(\Delta_g)$ and note that

$$S_g(s)=\Delta_g(E(s))=\int gs_{E(s)}dP=\int gsdP \quad \text{(E.9)}$$

for any $s\in\mathcal{S}$. Further note that by Lemma 1.5.9 in van der Vaart and Wellner (1996) and Gaussianity of $\mathbb{G}_0$, $\mathbf{T}$ is totally bounded under the semimetric $d(\tau_1,\tau_2)\equiv\|s_{\tau_1}-s_{\tau_2}\|_{P,2}$ and the sample paths of $\mathbb{G}_0$ are almost surely uniformly continuous with respect to $d(\cdot,\cdot)$. It follows that $\mathbb{S}_0$ is almost surely uniformly continuous on $\mathcal{S}$ with respect to $\|\cdot\|_{P,2}$, which implies its sample paths almost surely admit a unique extension to $\bar{\mathcal{S}}$ for $\bar{\mathcal{S}}$ the closure of $\mathcal{S}$ under $\|\cdot\|_{P,2}$, and thus we may view $\mathbb{S}_0$ as an element of the space

$$C(\bar{\mathcal{S}})\equiv\{S:\bar{\mathcal{S}}\to\mathbf{R}\text{ that are continuous under }\|\cdot\|_{P,2}\}. \quad \text{(E.10)}$$

Moreover, since $\mathbf{T}$ being totally bounded under $d(\cdot,\cdot)$ implies $\mathcal{S}$ is totally bounded under $\|\cdot\|_{P,2}$, it follows that $\bar{\mathcal{S}}$ is compact under $\|\cdot\|_{P,2}$. Since $\mathbb{S}_0$ is Radon by Theorem A.3.11 in Bogachev (1998), we can conclude from (E.8) and (E.9), and Lemma E.4 (below) that S_g belongs to the support of $\mathbb{S}_0$. In particular, we conclude for any $\epsilon>0$ that

$$0<P(\|S_g-\mathbb{S}_0\|_\infty>\epsilon)=P(\|\Upsilon(\Delta_g)-\Upsilon(\mathbb{G}_0)\|_\infty>\epsilon)=P(\|\Delta_g-\mathbb{G}_0\|_\infty>\epsilon) \quad \text{(E.11)}$$

where the first equality follows by definition of $\Upsilon:\ell^\infty(\mathbf{T})\to\ell^\infty(\mathcal{S})$, and the second equality is implied by (E.7) and $\Delta_g(\tau)=\Upsilon(\Delta_g)(s)$ for any $\tau\in B^{-1}(s)$. Thus, since $\epsilon>0$ was arbitrary, we conclude from (E.11) that Δ_g is in the support of $\mathbb{G}_0$. ■

Lemma E.4. *Let Assumption 2.1(i) hold, $\mathcal{S}\subset L_0^2(P)$ be compact under $\|\cdot\|_{P,2}$, for any $g\in L_0^2(P)$ let $S_g:\mathcal{S}\to\mathbf{R}$ be given by $S_g(s)=\int sgdP$, and define*

$$C(\mathcal{S})\equiv\{S:\mathcal{S}\to\mathbf{R}\textit{ is continuous under }\|\cdot\|_{P,2}\}, \quad \text{(E.12)}$$

which is endowed with the norm $\|S\|_\infty=\sup_{s\in\mathcal{S}}|S(s)|$. If $\mathbb{S}_0$ is a centered Radon Gaussian measure on $C(\mathcal{S})$ satisfying $E[\mathbb{S}_0(s_1)\mathbb{S}_0(s_2)]=\int s_1s_2dP$ for any $s_1,s_2\in\mathcal{S}$, then it follows that S_g belongs *to the support of* $\mathbb{S}_0$ *for any* $g \in L_0^2(P)$.

Proof of Lemma E.4: Fix $g \in L_0^2(P)$ and note the Cauchy-Schwarz inequality yields

$$|S_g(s_1) - S_g(s_2)| \leq \int |g||s_1 - s_2|dP \leq \|g\|_{P,2}\|s_1 - s_2\|_{P,2} \tag{E.13}$$

for any $s_1, s_2 \in \mathcal{S}$, and therefore $S_g \in C(\mathcal{S})$. Let $\bar{V}$ denote the closure of the linear span of $\mathcal{S}$ in $L_0^2(P)$ and set $\Pi_{\bar{V}}(g)$ to equal the metric projection of g onto $\bar{V}$. For any $s \in \mathcal{S}$, then define $S_{\Pi_{\bar{V}}(g)} : \mathcal{S} \to \mathbf{R}$ by $S_{\Pi_{\bar{V}}(g)}(s) = \int \{\Pi_{\bar{V}}(g)\} s dP$ and note that

$$S_g(s) = \int gsdP = \int \{\Pi_{\bar{V}}(g)\}sdP = S_{\Pi_{\bar{V}}(g)}(s). \tag{E.14}$$

Moreover, since $\Pi_{\bar{V}}(g) \in \bar{V}$, it follows there is a sequence sequence $\{g_k\}_{k=1}^{\infty}$ such that

$$\lim_{k\to\infty} \|g_k - \Pi_{\bar{V}}(g)\|_{P,2} = 0, \tag{E.15}$$

where each g_k satisfies for some $\{\alpha_{j,k}, s_{j,k}\}_{j=1}^{k}$ with $(\alpha_{j,k}, s_{j,k}) \in \mathbf{R} \times \mathcal{S}$ the relation

$$g_k = \sum_{j=1}^{k} \alpha_{j,k} s_{j,k}. \tag{E.16}$$

Defining $S_{g_k} : \mathcal{S} \to \mathbf{R}$ by $S_{g_k}(s) = \int g_k s dP$, we then conclude from results (E.14) and (E.15) together with the Cauchy-Schwarz inequality that

$$\begin{aligned} \lim_{k\to\infty} \|S_{g_k} - S_g\|_\infty = \lim_{k\to\infty} \|S_{g_k} - S_{\Pi_{\bar{V}}(g)}\|_\infty &\leq \lim_{k\to\infty} \sup_{s\in\mathcal{S}} \int |s||g_k - \Pi_{\bar{V}}(g)|dP \\ &\leq \sup_{s\in\mathcal{S}} \|s\|_{P,2} \times \lim_{k\to\infty} \|g_k - \Pi_{\bar{V}}(g)\|_{P,2} = 0, \end{aligned} \tag{E.17}$$

where in the final equality we exploited that $\sup_{s\in\mathcal{S}} \|s\|_{P,2} < \infty$ since $\mathcal{S}$ is compact under $\|\cdot\|_{P,2}$ by hypothesis. In particular, since the topological support of $\mathbb{S}_0$ is a closed subset of $C(\mathcal{S})$, result (E.17) implies that to establish the Lemma it suffices to show S_{g_k} belongs to the support of $\mathbb{S}_0$ for all k. To this end, we let $ca(\mathcal{S})$ denote the set of finite signed Borel (w.r.t. $\|\cdot\|_{P,2}$) measures on $\mathcal{S}$, and note that by Theorem 14.15 in Aliprantis and Border (2006), it follows $ca(\mathcal{S})$ is the dual space of $C(\mathcal{S})$. Next, for any k we define a measure $\nu_k \in ca(\mathcal{S})$ by setting for each Borel set $A \subseteq C(\mathcal{S})$

$$\nu_k(A) = \sum_{j=1}^{k} \alpha_{j,k} 1\{s_{j,k} \in A\}. \tag{E.18}$$

Following the notation in Bogachev (1998), for any k we additionally introduce the linear map $R(\nu_k) : ca(\mathcal{S}) \to \mathbf{R}$ which for any $\mu \in ca(\mathcal{S})$ is given by

$$R(\nu_k)(\mu) = E[\{\int \mathbb{S}_0(s)\nu_k(ds)\}\{\int \mathbb{S}_0(s)\mu(ds)\}]. \tag{E.19}$$

By results (E.18) and (E.19), Fubini's theorem, see e.g. Corollary 3.4.2 in Bogachev (2007), and

$E[\mathbb{S}_0(s_1)\mathbb{S}_0(s_2)] = \int s_1 s_2 dP$ for any $s_1, s_2 \in \mathcal{S}$ we then obtain

$$R(\nu_k)(\mu) = E[\{\sum_{j=1}^{k} \alpha_{j,k}\mathbb{S}_0(s_{j,k})\}\{\int \mathbb{S}_0(s)\mu(ds)\}]$$

$$= \int E[\{\sum_{j=1}^{k} \alpha_{j,k}\mathbb{S}_0(s_{j,k})\}\mathbb{S}_0(s)]\mu(ds) = \int S_{g_k}(s)\mu(ds), \quad \text{(E.20)}$$

where the last equality follows from (E.16). Result (E.20) implies we may identify the linear map $R(\nu_k) : ca(\mathcal{S}) \to \mathbf{R}$ with S_{g_k}, and therefore Theorem 3.2.3 in Bogachev (1998) implies S_{g_k} is in the Cameron-Martin space of $\mathbb{S}_0$. However, by Theorem 3.6.1 in Bogachev (1998) the Cameron-Martin space of $\mathbb{S}_0$ is a subset of its support, and hence we conclude S_{g_k} is in the support of $\mathbb{S}_0$. The Lemma then follows from (E.17). ∎

Lemma E.5. *Let $\mathbb{G}_0$ be a centered Gaussian measure on a separable Banach space $\mathbf{B}$ and $0 \neq \Delta \in \mathbf{B}$ belong to the support of $\mathbb{G}_0$. Further suppose $\Psi : \mathbf{B} \to \mathbf{R}_+$ is continuous, convex and nonconstant, and satisfies $\Psi(0) = 0$, $\Psi(b) = \Psi(-b)$ for all $b \in \mathbf{B}$, and $\{b \in \mathbf{B} : \Psi(b) \leq t\}$ is bounded for any $0 < t < \infty$. For any $t > 0$ it then follows that*

$$P(\Psi(\mathbb{G}_0 + \Delta) < t) < P(\Psi(\mathbb{G}_0) < t).$$

Proof of Lemma E.5: Let $\|\cdot\|_{\mathbf{B}}$ denote the norm of $\mathbf{B}$, fix $t > 0$, and define

$$C \equiv \{b \in \mathbf{B} : \Psi(b) < t\}. \quad \text{(E.21)}$$

For $\mathbf{B}^*$ the dual space of $\mathbf{B}$ let $\|\cdot\|_{\mathbf{B}^*}$ denote its norm, and $\nu_C : \mathbf{B}^* \to \mathbf{R}$ be given by

$$\nu_C(b^*) = \sup_{b \in C} b^*(b), \quad \text{(E.22)}$$

which constitutes the support functional of C. Then note for any $b^* \in \mathbf{B}^*$ we have

$$\nu_C(-b^*) = \sup_{b \in C} -b^*(b) = \sup_{b \in C} b^*(-b) = \sup_{b \in -C} b^*(b) = \nu_C(b^*), \quad \text{(E.23)}$$

due to $C = -C$ since $\Psi(b) = \Psi(-b)$ for all $b \in \mathbf{B}$. Moreover, note that $0 \in C$ since $\Psi(0) = 0 < t$, and hence there exists a $M_0 > 0$ such that $\{b \in \mathbf{B} : \|b\|_{\mathbf{B}} \leq M_0\} \subseteq C$ by continuity of Ψ. Thus, by definition of $\|\cdot\|_{\mathbf{B}^*}$ we obtain for any $b^* \in \mathbf{B}^*$ that

$$\nu_C(b^*) = \sup_{b \in C} b^*(b) \geq \sup_{\|b\|_{\mathbf{B}} \leq M_0} b^*(b) = M_0 \times \sup_{\|b\|_{\mathbf{B}} \leq 1} |b^*(b)| = M_0\|b^*\|_{\mathbf{B}^*}. \quad \text{(E.24)}$$

Analogously, note that by assumption $M_1 \equiv \sup_{b \in C} \|b\|_{\mathbf{B}} < \infty$, and thus for any $b^* \in \mathbf{B}^*$

$$\nu_C(b^*) = \sup_{b \in C} b^*(b) \leq \|b^*\|_{\mathbf{B}^*} \times \sup_{b \in C} \|b\|_{\mathbf{B}} = M_1\|b^*\|_{\mathbf{B}^*}. \quad \text{(E.25)}$$

We next aim to define a norm on $\mathbf{B}$ under which C is the open unit sphere. To this end, recall that the original norm $\|\cdot\|_{\mathbf{B}}$ on $\mathbf{B}$ may be written as

$$\|b\|_{\mathbf{B}} = \sup_{\|b^*\|_{\mathbf{B}^*}=1} b^*(b), \tag{E.26}$$

see for instance Lemma 6.10 in Aliprantis and Border (2006). Similarly, instead define

$$\|b\|_{\mathbf{B},C} \equiv \sup_{\|b^*\|_{\mathbf{B}^*}=1} \frac{b^*(b)}{\nu_C(b^*)}, \tag{E.27}$$

and note that: (i) $\|b_1 + b_2\|_{\mathbf{B},C} \leq \|b_1\|_{\mathbf{B},C} + \|b_2\|_{\mathbf{B},C}$ for any $b_1, b_2 \in \mathbf{B}$ by direct calculation, (ii) $\|\alpha b\|_{\mathbf{B},C} = |\alpha|\|b\|_{\mathbf{B},C}$ for any $\alpha \in \mathbf{R}$ and $b \in \mathbf{B}$ by (E.23) and (E.27), and (iii) results (E.24), (E.25), (E.26), and (E.27) imply that for any $b \in \mathbf{B}$

$$M_0\|b\|_{\mathbf{B},C} \leq \|b\|_{\mathbf{B}} \leq M_1\|b\|_{\mathbf{B},C} \tag{E.28}$$

which establishes $\|b\|_{\mathbf{B},C} = 0$ if an only if $b = 0$, and hence we conclude $\|\cdot\|_{\mathbf{B},C}$ is indeed a norm on $\mathbf{B}$. In fact, (E.28) implies that the norms $\|\cdot\|_{\mathbf{B}}$ and $\|\cdot\|_{\mathbf{B},C}$ are equivalent, and hence $\mathbf{B}$ remains a separable Banach space and its Borel σ-algebra unchanged when endowed with $\|\cdot\|_{\mathbf{B},C}$ in place of $\|\cdot\|_{\mathbf{B}}$.

Next, note that the continuity of Ψ implies C is open, and thus for any $b_0 \in C$ there is an $\epsilon > 0$ such that and $\{b \in \mathbf{B} : \|b - b_0\|_{\mathbf{B}} \leq \epsilon\} \subset C$. We then obtain

$$\nu_C(b^*) \geq \sup_{\|b-b_0\|_{\mathbf{B}}\leq\epsilon} b^*(b) = \sup_{\|b\|_{\mathbf{B}}\leq 1} \{b^*(b_0) + \epsilon b^*(b)\} = b^*(b_0) + \epsilon\|b^*\|_{\mathbf{B}^*}, \tag{E.29}$$

where the final equality follows as in (E.24). Thus, from (E.25) and (E.29) we conclude $1 - \epsilon/M_1 \geq b^*(b_0)/\nu_C(b^*)$ for all b^* with $\|b^*\|_{\mathbf{B}^*} = 1$, and hence we conclude

$$C \subseteq \{b \in \mathbf{B} : \|b\|_{\mathbf{B},C} < 1\}. \tag{E.30}$$

Suppose on the other hand that $\|b_0\|_{\mathbf{B},C} < 1$, and note (E.27) implies for some $\delta > 0$

$$b^*(b_0) < \nu_C(b^*)(1 - \delta) \tag{E.31}$$

for all $b^* \in \mathbf{B}^*$ with $\|b^*\|_{\mathbf{B}^*} = 1$. Setting $\eta \equiv \delta M_0$ and arguing as in (E.29) then yields

$$\begin{aligned}\sup_{\|b^*\|_{\mathbf{B}^*}=1} \sup_{\|b-b_0\|_{\mathbf{B}}\leq\eta} \{b^*(b) - \nu_C(b^*)\} &= \sup_{\|b^*\|_{\mathbf{B}^*}=1} \{b^*(b_0) + \eta\|b^*\|_{\mathbf{B}^*} - \nu_C(b^*)\} \\ &< \sup_{\|b^*\|_{\mathbf{B}^*}=1} \{\eta - \nu_C(b^*)\delta\} = \sup_{\|b^*\|_{\mathbf{B}^*}=1} \delta(M_0 - \nu_C(b^*)) \leq 0,\end{aligned} \tag{E.32}$$

where the first inequality follows from (E.31), the second equality by definition of η, and the final inequality follows from (E.24). Since C is convex by hypothesis, (E.32) and Theorem 5.12.5 in Luenberger (1969) imply $\{b \in \mathbf{B} : \|b - b_0\|_{\mathbf{B}} \leq \eta\} \subseteq \bar{C}$. We conclude b_0 is in the interior of $\bar{C}$, and since C is convex and open, Lemma 5.28 in Aliprantis and Border (2006) yields that $b_0 \in C$. Thus, we can conclude that

$$\{b \in \mathbf{B} : \|b\|_{\mathbf{B},C} < 1\} \subseteq C, \tag{E.33}$$

which together with (E.30) yields $C = \{b \in \mathbf{B} : \|b\|_{\mathbf{B},C} < 1\}$. Therefore, $\mathbf{B}$ being separable under $\|\cdot\|_{\mathbf{B},C}$, $0 \neq \Delta$ being in the support of $\mathbb{G}_0$ by hypothesis, and Corollary 2 in Lewandowski et al. (1995) finally enable us to derive

$$P(\Psi(\mathbb{G}_0 + \Delta) < t) = P(\mathbb{G}_0 + \Delta \in C) < P(\mathbb{G}_0 \in C) = P(\Psi(\mathbb{G}_0) < t), \tag{E.34}$$

which establishes the claim of the Lemma. ■

Lemma E.6. *Suppose Assumption 5.1 holds, $f : \mathbf{X} \to \mathbf{R}$ is bounded, and let $\hat{\theta}_n : \{X_i\}_{i=1}^n \to \mathbf{R}$ be an estimator of $\int f dP$. If $\hat{\theta}_n$ is such that for any path $t \mapsto P_{t,g} \in \mathbf{P}$*

$$\sqrt{n}\{\hat{\theta}_n - \int f dP_{1/\sqrt{n},g}\} \overset{L_{n,g}}{\to} \mathbb{Z}_g \tag{E.35}$$

for some tight law $\mathbb{Z}_g$, then (E.35) holds for any path $t \mapsto P_{t,g} \in \mathcal{M}$ with $g \in \bar{T}(P)$. Moreover, if a sequence $\{g_j\}_{j=1}^\infty \subseteq T(P)$ satisfies $\|g_j - g_0\|_{P,2} = o(1)$ for some $g_0 \in \bar{T}(P)$, then it follows that $\mathbb{Z}_{g_j} \to \mathbb{Z}_{g_0}$ in the weak topology.

Proof of Lemma E.6: Fix a score $g_0 \in \bar{T}(P)$, select a sequence $\{g_j\}_{j=1}^\infty \subseteq T(P)$ with $\|g_j - g_0\|_{P,2} = o(1)$, and set Δ_j to equal

$$\Delta_j \equiv 2\{1 - \exp\{-\frac{1}{4}\|g_0 - g_j\|_{P,2}\}\}^{1/2} \tag{E.36}$$

and note $\Delta_j = o(1)$. We further observe that since $f : \mathbf{X} \to \mathbf{R}$ is bounded, we obtain

$$|\sqrt{n}\{\int f dP_{1/\sqrt{n},g} - \int f dP\} - \int f g dP| \leq \|f\|_\infty \int |\frac{g}{2} dP^{1/2}(dP^{1/2}_{1/\sqrt{n},g} - dP^{1/2})| \\ + \|f\|_\infty \int |\sqrt{n}\{dP^{1/2}_{1/\sqrt{n},g} - dP^{1/2}\} - \frac{g}{2} dP^{1/2}|(dP^{1/2}_{1/\sqrt{n},g} + dP^{1/2}) \tag{E.37}$$

for any path $t \mapsto P_{t,g}$. In particular, result (E.37) and the Cauchy-Schwarz inequality implies $t \mapsto \int f dP_{t,g}$ has pathwise derivative $\int f g dP$ at $t = 0$. Hence, since $g_j \in T(P)$ implies there exists a path $t \mapsto P_{t,g_j} \in \mathbf{P}$ such that (E.35) holds, we obtain

$$\sqrt{n}\{\hat{\theta}_n - \int f dP_{1/\sqrt{n},g_0}\} \overset{L_{n,g_j}}{\to} \mathbb{Z}_{g_j} + \int f(g_j - g_0) dP \tag{E.38}$$

for any j by the pathwise differentiability of $t \mapsto \int f dP_{t,g}$ and the continuous mapping theorem. For any continuous and bounded map $F : \mathbf{R} \to \mathbf{R}$, we then note

$$\begin{aligned}\limsup_{n\to\infty} &\int F(\sqrt{n}\{\hat{\theta}_n - \int f dP_{1/\sqrt{n},g_0}\}) dP_{1/\sqrt{n},g_0} \\ &\leq \limsup_{n\to\infty} \int F(\sqrt{n}\{\hat{\theta}_n - \int f dP_{1/\sqrt{n},g_0}\}) dP_{1/\sqrt{n},g_j} + \|F\|_\infty \Delta_j \\ &= \liminf_{n\to\infty} \int F(\sqrt{n}\{\hat{\theta}_n - \int f dP_{1/\sqrt{n},g_0}\}) dP_{1/\sqrt{n},g_j} + \|F\|_\infty \Delta_j \\ &\leq \liminf_{n\to\infty} \int F(\sqrt{n}\{\hat{\theta}_n - \int f dP_{1/\sqrt{n},g_0}\}) dP_{1/\sqrt{n},g_0} + 2\|F\|_\infty \Delta_j\end{aligned} \tag{E.39}$$

where the inequalities follow from (E.36) and Lemma D.1, and the equality from the limit existing by (E.38) and $F : \mathbf{R} \to \mathbf{R}$ being continuous and bounded. Since $\Delta_j = o(1)$, (E.39) implies the following limit exists for any continuous and bounded $F : \mathbf{R} \to \mathbf{R}$

$$L(F) \equiv \lim_{n\to\infty} \int F(\sqrt{n}\{\hat{\theta}_n - \int f dP_{1/\sqrt{n},g_0}\})dP_{1/\sqrt{n},g_0}. \tag{E.40}$$

In addition, for any $\epsilon > 0$ there exists a $j(\epsilon)$ such that $\Delta_{j(\epsilon)} < \epsilon/2$ and, since $\mathbb{Z}_{g_j(\epsilon)}$ is tight, a compact set K_ϵ such that $P(\mathbb{Z}_{g_{j(\epsilon)}} + \int f(g_{j(\epsilon)} - g_0)dP \in K_\epsilon) \geq 1 - \epsilon/2$. For any $\delta > 0$ let $K_\epsilon^\delta \equiv \{a \in \mathbf{R} : \inf_{b\in K_\epsilon} \|a - b\| < \delta\}$, and note Portmanteu's Theorem, see Theorem 1.3.4(ii) in van der Vaart and Wellner (1996), (E.38), and Lemma D.1 yield

$$\begin{aligned}
&\liminf_{n\to\infty} P_{1/\sqrt{n},g_0}(\sqrt{n}\{\hat{\theta}_n - \int f dP_{1/\sqrt{n},g_0}\} \in K_\epsilon^\delta) \\
&\qquad \geq \liminf_{n\to\infty} P_{1/\sqrt{n},g_{j(\epsilon)}}(\sqrt{n}\{\hat{\theta}_n - \int f dP_{1/\sqrt{n},g_0}\} \in K_\epsilon^\delta) - \Delta_{j(\epsilon)} \\
&\qquad \geq P(\mathbb{Z}_{g_{j(\epsilon)}} + \int (g_{j(\epsilon)} - g_0)dP \in K_\epsilon) - \Delta_{j(\epsilon)} \\
&\qquad \geq 1 - \epsilon.
\end{aligned} \tag{E.41}$$

Since ϵ was arbitrary, result (E.41) implies that the law of $\sqrt{n}\{\hat{\theta}_n - \int f dP_{1/\sqrt{n},g_0}\}$ under $P^n_{1/\sqrt{n},g_0}$ is asymptotically tight. Prohorov's theorem, see e.g. Theorem 1.3.9 in van der Vaart and Wellner (1996), then yields that every subsequence of $\sqrt{n}\{\hat{\theta}_n - \int f dP_{1/\sqrt{n},g_0}\}$ has a further subsequence that converges in distribution under $P^n_{1/\sqrt{n},g_0}$. However, in combination with result (E.40), these observations imply that $\sqrt{n}\{\hat{\theta}_n - \int f dP_{1/\sqrt{n},g_0}\}$ must itself converge in distribution under $P^n_{1/\sqrt{n},g_0}$, and we denote the limit law by $\mathbb{Z}_{g_0}$:

$$\sqrt{n}\{\hat{\theta}_n - \int f dP_{1/\sqrt{n},g_0}\} \overset{L_{n,g_0}}{\to} \mathbb{Z}_{g_0}. \tag{E.42}$$

Moreover, for any continuous and bounded $F : \mathbf{R} \to \mathbf{R}$, (E.38), (E.39), (E.42) imply

$$E[F(\mathbb{Z}_{g_0})] - \|F\|_\infty \Delta_j \leq E[F(\mathbb{Z}_{g_j} + \int f(g_j - g_0)dP)] \leq E[F(\mathbb{Z}_{g_0})] + \|F\|_\infty \Delta_j. \tag{E.43}$$

Since $\Delta_j = o(1)$, it therefore follows from (E.43) that $\mathbb{Z}_{g_j} + \int f(g_j - g_0)dP \to \mathbb{Z}_{g_0}$ in the weak topology. However, by the Cauchy-Schwarz inequality and $\|g_j - g_0\|_{P,2} = o(1)$ we can further conclude that $\int f(g_j - g_0)dP = o(1)$, and thus by the continuous mapping theorem we obtain that $\mathbb{Z}_{g_j} \to \mathbb{Z}_{g_0}$ in the weak topology. ■

Appendix F - Proofs of Main Results in Section 4

In this Appendix we first provide the proofs for Theorem 4.1, Lemma 4.1, Lemma 4.2, and Corollary 4.1. We then establish a theorem (Theorem F.1) that includes Theorem 4.2 as a special case for models defined by sequential moment restrictions.

Lemma F.1. *Let $\mathcal{F} \subset L_0^2(P)$ be such that $|f| \leq F$ for all $f \in \mathcal{F}$ and some $F \in L^2(P)$. Then, for* *any path $t \mapsto P_{t,g} \in \mathcal{M}$ satisfying $\int F^2 dP_{t,g} = O(1)$ it follows that*

$$\lim_{t\downarrow 0} \sup_{f\in\mathcal{F}} |\int \frac{f}{t}\{dP_{t,g} - dP\} - \int fgdP| = 0.$$

Proof of Lemma F.1: Since $t \mapsto P_{t,g}$ is a path, the Cauchy-Schwarz inequality implies

$$\lim_{t\downarrow 0} \sup_{f\in\mathcal{F}} \Big|\int f\big[\frac{1}{t}\{dP_{t,g}^{1/2} - dP^{1/2}\} - \frac{g}{2}dP^{1/2}\big]\big(dP_{t,g}^{1/2} + dP^{1/2}\big)\Big| = 0, \tag{F.1}$$

where we exploited that $F \in L^2(P)$ and $\int F^2 dP_{t,g} = O(1)$ by hypothesis. Next, for any $M < \infty$ we obtain by the Cauchy-Schwarz and triangle inequalities that

$$\begin{aligned}\sup_{f\in\mathcal{F}} |\int fgdP^{1/2}(dP_{t,g}^{1/2} - dP^{1/2})| \leq M\{\int g^2 dP\}^{1/2}\{\int (dP_{t,g}^{1/2} - dP)^{1/2})^2\}^{1/2} \\ + \{\int_{|F|>M} g^2 dP\}^{1/2}\{\int F^2(dP_{t,g}^{1/2} - dP^{1/2})^2\}^{1/2}.\end{aligned} \tag{F.2}$$

Therefore, result (F.2), $t \mapsto P_{t,g}$ being a path, $\int F^2 dP_{t,g} = O(1)$ by hypothesis, $g \in L_0^2(P)$, and $P(|F(X)| > M)$ converging to zero as M diverges to infinity, yields

$$\begin{aligned}\lim_{t\downarrow 0} \sup_{f\in\mathcal{F}} |\int fgdP^{1/2}(dP_{t,g}^{1/2} - dP^{1/2})| \\ \leq \lim_{M\uparrow\infty} \{\int_{|F|>M} g^2 dP\}^{1/2} \times \lim_{t\downarrow 0}\{\int F^2(dP_{t,g}^{1/2} - dP^{1/2})^2\}^{1/2} = 0.\end{aligned} \tag{F.3}$$

Hence, results (F.1) and (F.3) and the triangle inequality together establish

$$\begin{aligned}\lim_{t\downarrow 0} \sup_{f\in\mathcal{F}} |\frac{1}{t}\int f(dP_{t,g} - dP) - \int fgdP| \leq \lim_{t\downarrow 0} \sup_{f\in\mathcal{F}} |\frac{1}{2}\int fgdP^{1/2}(dP_{t,g}^{1/2} - dP^{1/2})| \\ + \lim_{t\downarrow 0} \sup_{f\in\mathcal{F}} |\int f\{\frac{1}{t}(dP_{t,g}^{1/2} - dP^{1/2}) - \frac{1}{2}gdP^{1/2}\}(dP_{t,g}^{1/2} + dP^{1/2})| = 0,\end{aligned} \tag{F.4}$$

and therefore the claim of the Lemma follows. ■

Proof of Theorem 4.1: Consider any path $t \mapsto P_{t,g}$ satisfying Condition A, and an arbitrary bounded function $\psi_j \in L^2(W_j)$ for any $1 \leq j \leq J$. Then from P satisfying (26) by Assumption 4.1(i) and $t \mapsto P_{t,g}$ satisfying Condition A(i) we obtain

$$\begin{aligned}0 &= \frac{1}{t}\{\int \rho_j(\cdot, h_t)\psi_j dP_{t,g} - \int \rho_j(\cdot, h_P)\psi_j dP\} \\ &= \frac{1}{t}\{\int \rho_j(\cdot, h_t)\psi_j(dP_{t,g} - dP) + \int (\rho_j(\cdot, h_t) - \rho_j(\cdot, h_P))\psi_j dP\}.\end{aligned} \tag{F.5}$$

Furthermore, since the path $t \mapsto P_{t,g}$ satisfies Condition A(iii) we obtain by Lemma F.1

$$\lim_{t\downarrow 0} \frac{1}{t}\int \rho_j(\cdot, h_t)\psi_j(dP_{t,g} - dP) = \lim_{t\downarrow 0}\int \rho_j(\cdot, h_t)\psi_j gdP = \int \rho_j(\cdot, h_P)\psi_j gdP \tag{F.6}$$

where the final equality follows by Assumption 4.1(iii), the Cauchy-Schwarz inequality, ψ_j being bounded, and $\|h_t - h_P\|_{\mathbf{H}} = o(1)$ by Condition A(ii). On the other hand, since $m_j(W_j, \cdot) : \mathbf{H} \to L^2(W_j)$ is Fréchet differentiable and $\|t^{-1}(h_t - h) - \Delta\|_{\mathbf{H}} = o(1)$ as $t \downarrow 0$ by Condition A(ii) for some $\Delta \in \mathbf{H}$, we can in addition conclude that

$$\lim_{t \downarrow 0} \frac{1}{t} E[\psi_j(W_j)\{m_j(W_j, h_t) - m_j(W_j, h_P)\}] = E[\psi_j(W_j)\nabla m_j(W_j, h_P)[\Delta]]. \tag{F.7}$$

Therefore, combining results (F.5), (F.6), and (F.7) we can obtain that any path $t \mapsto P_{t,g}$ satisfying Condition A must have a score $g \in L_0^2(P)$ satisfying the restriction

$$E[\{\sum_{j=1}^{J} \psi_j(W_j)\rho_j(Z, h_P)\}g(X)] = -E[\{\sum_{j=1}^{J} \psi_j(W_j)\nabla m_j(W_j, h_P)[\Delta]\}], \tag{F.8}$$

for any collection $(\psi_1, \ldots, \psi_J) \in \bigotimes_{j=1}^{J} L^2(W_j)$ of bounded functions. However, since the set of bounded functions of W_j is dense in $L^2(W_j)$ for any $1 \le j \le J$, the Cauchy Schwarz inequality and $E[\rho_j^2(Z, h_P)|W_j]$ being bounded almost surely by Assumption 4.2(i) imply that (F.8) actually holds for all $(\psi_1, \ldots, \psi_J) \in \bigotimes_{j=1}^{J} L^2(W_j)$. In particular, we note that if we select $(\psi_1, \ldots, \psi_J) \in \bar{\mathcal{R}}^{\perp}$, then the right hand side of (F.8) is equal to zero, and therefore, result (F.8) implies the set inclusion

$$\{f \in L_0^2(P) : f = \sum_{j=1}^{J} \rho_j(Z, h_P)\psi_j(W_j) \text{ for some } (\psi_1, \ldots, \psi_J) \in \bar{\mathcal{R}}^{\perp}\} \subseteq \bar{T}(P)^{\perp}. \tag{F.9}$$

In order to establish the Theorem we therefore only need to show the reverse inclusion in (F.9). As a preliminary result towards this goal, we first aim to establish that

$$\{f \in L_0^2(P) : f = \sum_{j=1}^{J} \rho_j(Z, h_P)\psi_j(W_j) \text{ for some } (\psi_1, \ldots, \psi_J) \in \bar{\mathcal{R}}^{\perp}\} = \bar{\mathcal{V}} \cap \bar{T}(P)^{\perp}. \tag{F.10}$$

To this end, note that by result (F.9) and the definition of $\bar{\mathcal{V}}$ we obtain the set inclusion

$$\{f \in L_0^2(P) : f = \sum_{j=1}^{J} \rho_j(Z, h_P)\psi_j(W_j) \text{ for some } (\psi_1, \ldots, \psi_J) \in \bar{\mathcal{R}}^{\perp}\} \subseteq \bar{\mathcal{V}} \cap \bar{T}(P)^{\perp}. \tag{F.11}$$

Next, we note that since $\bar{\mathcal{R}}$ is a closed linear subspace of $\bigotimes_{j=1}^{J} L^2(W_j)$, Theorem 3.4.1 in Luenberger (1969) implies we may decompose $\bigotimes_{j=1}^{J} L^2(W_j) = \bar{\mathcal{R}} \oplus \bar{\mathcal{R}}^{\perp}$. For any $(f_1, \ldots, f_J) \in \bigotimes_{j=1}^{J} L^2(W_j)$ we in turn denote its projection onto $\bar{\mathcal{R}}$ and $\bar{\mathcal{R}}^{\perp}$ as $(\Pi_{\mathcal{R}} f_1, \ldots, \Pi_{\mathcal{R}} f_J)$ and $(\Pi_{\mathcal{R}^{\perp}} f_1, \ldots, \Pi_{\mathcal{R}^{\perp}} f_J)$ respectively. Selecting an arbitrary $f \in \mathcal{V} \cap \bar{T}(P)^{\perp}$, which by definition of $\mathcal{V}$ must be of the form $f = \sum_{j=1}^{J} \rho_j(\cdot, h_P)\psi_j^f$ for some $(\psi_1^f, \ldots, \psi_J^f) \in \bigotimes_{j=1}^{J} L^2(W_j)$, we then observe that result (F.8) implies that for any path $t \mapsto P_{t,g}$ satisfying Condition A we must have the equality

$$E[\{\sum_{j=1}^{J} \psi_j^f(W_j)\rho_j(Z, h_P)\}g(X)] = -E[\sum_{j=1}^{J} \{\Pi_{\mathcal{R}} \psi_j^f(W_j)\}\nabla m_j(W_j, h_P)[\Delta]]. \tag{F.12}$$

However, by Assumption 4.1(iv), if $(\Pi_{\mathcal{R}} \psi_1^f, \ldots, \Pi_{\mathcal{R}} \psi_J^f) \neq 0$ (in $\bigotimes_{j=1}^{J} L^2(W_j)$), then there is a path

$t \mapsto P_{t,g}$ satisfying Condition A with $\|t^{-1}(h_t - h_P) - \Delta\|_{\mathbf{H}} = 0$ and

$$E[\sum_{j=1}^{J}\{\Pi_{\mathcal{R}}\psi_j^f(W_j)\}\nabla m_j(W,h_P)[\Delta]] \neq 0. \tag{F.13}$$

Therefore, if $f \in \mathcal{V}$ is such that $(\Pi_{\mathcal{R}}\psi_1^f, \ldots, \Pi_{\mathcal{R}}\psi_J^f) \neq 0$, then results (F.12) and (F.13) establish that there exists a path $t \mapsto P_{t,g}$ satisfying Condition A and for which

$$E[\{\sum_{j=1}^{J}\psi_j^f(W_j)\rho_j(Z,h_P)\}g(X)] = -E[\sum_{j=1}^{J}\{\Pi_{\mathcal{R}}\psi_j^f(W_j)\}\nabla m_j(W,h_P)[\Delta]] \neq 0, \tag{F.14}$$

thus violating that $f \in \mathcal{V} \cap \bar{T}(P)^{\perp}$. In particular it follows that any $f \in \mathcal{V} \cap \bar{T}(P)^{\perp}$ satisfies $(\Pi_{\mathcal{R}}\psi_1^f, \ldots, \Pi_{\mathcal{R}}\psi_J^f) = 0$, and hence from $\bigotimes_{j=1}^{J} L^2(W_j) = \bar{\mathcal{R}} \oplus \bar{\mathcal{R}}^{\perp}$ we obtain

$$\mathcal{V} \cap \bar{T}(P)^{\perp} \subseteq \{f \in L_0^2(P) : f = \sum_{j=1}^{J}\rho_j(Z,h_P)\psi_j(W_j) \text{ for some } (\psi_1, \ldots, \psi_J) \in \bar{\mathcal{R}}^{\perp}\}. \tag{F.15}$$

Next, let $\bar{f} \in \bar{\mathcal{V}}$ be arbitrary, and note that $\|\bar{f} - \sum_{j=1}^{J}\rho_j(\cdot, h_P)\psi_j\|_{P,2}$ diverges to infinity as $\sum_{j=1}^{J}\|\psi_j\|_{P,2}$ diverges to infinity due to Assumption 4.2(ii). Thus, we obtain

$$\begin{aligned} 0 = & \inf_{(\psi_1,\ldots,\psi_J)\in\bigotimes_{j=1}^{J}L^2(W_j)} \|\bar{f} - \sum_{j=1}^{J}\rho_j(\cdot,h_P)\psi_j\|_{P,2} \\ & = \min_{(\psi_1,\ldots,\psi_J)\in\bigotimes_{j=1}^{J}L^2(W_j)} \|\bar{f} - \sum_{j=1}^{J}\rho_j(\cdot,h_P)\psi_j\|_{P,2}, \end{aligned} \tag{F.16}$$

where the first equality holds because $\bar{f} \in \bar{\mathcal{V}}$, and attainment in the second equality is implied by Proposition 38.14 in Zeidler (1984). However, attainment in (F.16) implies that $\bar{f} \in \mathcal{V}$, and hence since $\bar{f} \in \bar{\mathcal{V}}$ was arbitrary we can conclude $\bar{\mathcal{V}} = \mathcal{V}$. The claim in (F.10) then holds by result (F.11) and result (F.15).

In order to establish the Theorem we next aim to show Assumption 4.1 implies

$$\bar{\mathcal{V}}^{\perp} \subseteq \bar{T}(P). \tag{F.17}$$

Selecting an arbitrary $g \in \bar{\mathcal{V}}^{\perp} \cap L^{\infty}(P)$, we define a path with density (w.r.t. P) equal

$$\frac{dP_{t,g}}{dP} = 1 + tg, \tag{F.18}$$

which we note implies $P_{t,g}$ is indeed a probability measure for t small enough since $g \in L^{\infty}(P)$. The score of such a path is equal to g by direct calculation. Moreover, for any $\psi_j \in L^2(W_j)$ and $1 \leq j \leq J$ we have that $\rho_j(\cdot, h_P)\psi_j \in \mathcal{V}$ implies

$$\int \rho_j(\cdot,h_P)\psi_j dP_{t,g} = E[\rho_j(Z,h_P)(1+tg(X))\psi_j(W_j)] = 0 \tag{F.19}$$

where we exploited that $g \in \bar{\mathcal{V}}^{\perp}$ and $E[\rho_j(Z,h_P)|W_j] = 0$. Since $\psi_j \in L^2(W_j)$ was arbitrary, (F.19) in fact implies the path $t \mapsto P_{t,g}$ satisfies Condition A(i) with $h_t = h_P$ for all t, and hence also Conditions A(ii)-(iii). We conclude that $t \mapsto P_{t,g}$ satisfies Condition A, and as a result that $g \in \bar{T}(P)$. Since $g \in L^\infty(P) \cap \bar{\mathcal{V}}^\perp$ was arbitrary, it follows that $L^\infty(P) \cap \bar{\mathcal{V}}^\perp \subseteq \bar{T}(P)$ and by Assumption 4.1(v) that (F.17) indeed holds. Thus, we further obtain from result (F.17) that $\bar{T}(P)^\perp \subseteq (\bar{\mathcal{V}}^\perp)^\perp$ and since $(\bar{\mathcal{V}}^\perp)^\perp = \bar{\mathcal{V}}$ by Theorem 3.4.1 in Luenberger (1969), we can conclude that $\bar{T}(P)^\perp \subseteq \bar{\mathcal{V}}$. The theorem therefore follows from result (F.10). ■

Proof of Lemma 4.1: Recall that the map $\nabla m(W, h_P) : \mathbf{H} \to \bigotimes_{j=1}^J L^2(W_j)$ equals

$$\nabla m(W, h_P)[h] \equiv (\nabla m_1(W_1, h_P)[h], \ldots, \nabla m_J(W_J, h_P)[h])' \tag{F.20}$$

which is linear and continuous by the stated assumption that $m_j(W_j, \cdot) : \mathbf{H} \to L^2(W_j)$ is Fréchet differentiable at h_P for $1 \le j \le J$. Since $\bigotimes_{j=1}^J L^2(W_j)$ is its own dual, the adjoint $\nabla m(W, h_P)^*$ of the map $\nabla m(W, h_P)$ has domain $\bigotimes_{j=1}^J L^2(W_j)$. Moreover, because $\nabla m_j(W_j, h_P)^*$ is the adjoint of $\nabla m_j(W_j, h_P)$, it follows that

$$\nabla m(W, h_P)^*[f] = \sum_{j=1}^J \nabla m_j(W_j, h_P)^*[f_j] \tag{F.21}$$

for any $f = (f_1, \ldots, f_J) \in \bigotimes_{j=1}^J L^2(W_j)$. Letting $\mathcal{N}(\nabla m(W, h_P)^*)$ denote the null space of $\nabla m(W, h_P)^* : \bigotimes_{j=1}^J L^2(W_j) \to \mathbf{H}^*$, and noting that $\mathcal{R}$ (as defined in (29)) equals the range of $\nabla m(W, h_P) : \mathbf{H} \to \bigotimes_{j=1}^J L^2(W_j)$, we obtain by Theorem 6.6.1 in Luenberger (1969) that for $[\mathcal{R}]^\perp$ the orthocomplement of $\mathcal{R}$ in $\bigotimes_{j=1}^J L^2(W_j)$ we have

$$[\mathcal{R}]^\perp = \mathcal{N}(\nabla m(W, h_P)^*). \tag{F.22}$$

Furthermore, since $[\mathcal{R}]^\perp = \bar{\mathcal{R}}^\perp$ by continuity, and $\bar{\mathcal{R}} = \bigotimes_{j=1}^J L^2(W_j)$ if and only if $\bar{\mathcal{R}}^\perp = \{0\}$, Equation (F.22) yields $\bar{\mathcal{R}} = \bigotimes_{j=1}^J L^2(W_j)$ if and only if $\mathcal{N}(\nabla m(W, h_P)^*) = \{0\}$, which together with (F.21) establishes the Lemma. ■

Proof of Lemma 4.2: Since each $\nabla m_j(W_j, h_P) : \mathbf{H} \to L^2(W_j)$ is linear and $\mathbf{H} = \bigotimes_{j=1}^J \mathbf{H}_j$, for each j there are linear maps $\nabla m_{j,k}(W_j, h_P) : \mathbf{H}_k \to L^2(W_j)$ such that

$$\nabla m_j(W_j, h_P)[h] = \sum_{k=1}^J \nabla m_{j,k}(W_j, h_P)[h_k] \tag{F.23}$$

for all $(h_1, \ldots, h_J) = h \in \mathbf{H}$. Moreover, since $\nabla m_{j,k}(W_j, h_P)[h_k] = 0$ for any $h_k \in \mathbf{H}_k$ whenever $k > j$ by hypothesis, the decomposition in (F.23) implies that

$$\nabla m_j(W_j, h_P)[h] = \sum_{k=1}^j \nabla m_{j,k}(W_j, h_P)[h_k]. \tag{F.24}$$

We first suppose that $\bar{\mathcal{R}}_j = L^2(W_j)$ for all j and aim to show $\bar{\mathcal{R}} = \bigotimes_{j=1}^J L^2(W_j)$. To this end, let $(f_1, \ldots, f_J) \in \bigotimes_{j=1}^J L^2(W_j)$ and $\epsilon > 0$ be arbitrary. Then observe that

$$\|\nabla m_{1,1}(W_1, h_P)[h_1^*] - f_1\|_{P,2} < \epsilon \tag{F.25}$$

for some $h_1^* \in \mathbf{H}_1$ since $\bar{\mathcal{R}}_1 = L^2(W_1)$. For $2 \le j \le J$ we may then exploit that $\bar{\mathcal{R}}_j = L^2(W_j)$ to inductively select $h_j^* \in \mathbf{H}_j$ to satisfy the inequality

$$\|\nabla m_{j,j}(W_j, h_P)[h_j^*] - (f_j - \sum_{k=1}^{j-1} \nabla m_{j,k}(W_j, h_P)[h_k^*])\|_{P,2} < \epsilon. \tag{F.26}$$

Therefore, setting $(h_1^*, \ldots, h_J^*) = h^* \in \mathbf{H}$ and employing (F.24) and (F.26) we obtain

$$\sum_{j=1}^{J} \|\nabla m_j(W_j, h_P)[h^*] - f_j\|_{P,2} < J\epsilon \tag{F.27}$$

which, since $(f_1, \ldots, f_J) \in \bigotimes_{j=1}^J L^2(W_j)$ and $\epsilon > 0$ were arbitrary, implies that $\bar{\mathcal{R}} = \bigotimes_{j=1}^J L^2(W_j)$.

We next suppose $\bar{\mathcal{R}} = \bigotimes_{j=1}^J L^2(W_j)$ and aim to show $\bar{\mathcal{R}}_j = L^2(W_j)$ for all j. First note that by (F.24), it is immediate that $\bar{\mathcal{R}} = \bigotimes_{j=1}^J L^2(W_j)$ implies $\bar{\mathcal{R}}_1 = L^2(W_1)$. Thus, we focus on showing $\bar{\mathcal{R}}_j = L^2(W_j)$ for all $j > 1$. To this end, we select an arbitrary $1 < k^* \le J$ and $g^* \in L^2(W_{k^*})$, and define $(f_1^*, \ldots, f_J^*)$ to satisfy $f_j^* = g^*$ if $j = k^*$ and $f_j^* = 0$ if $j \ne k^*$. Next, note that since $\bar{\mathcal{R}} = \bigotimes_{j=1}^J L^2(W_j)$ by hypothesis, there is a sequence $(h_{1n}, \ldots, h_{Jn}) = h_n \in \mathbf{H}$ such that

$$\lim_{n\to\infty} \sum_{j=1}^{J} \|\nabla m_j(W_j, h_P)[h_n] - f_j^*\|_{P,2} = 0. \tag{F.28}$$

In particular, since $k^* > 1$, result (F.24) and h_n satisfying (F.28) together yield that

$$\lim_{n\to\infty} \|\nabla m_{1,1}(W_1, h_P)[h_{1n}]\|_{P,2} = 0. \tag{F.29}$$

Moreover, employing (F.24) and requirement (34) we obtain for any $2 \le j \le J$ that

$$\begin{aligned} &\|\nabla m_{j,j}(W_j, h_P)[h_{jn}] - f_j^*\|_{P,2} \\ &\qquad \le \|\nabla m_j(W_j, h_P)[h_n] - f_j^*\|_{P,2} + C\sum_{k=1}^{j-1} \|\nabla m_{k,k}(W_k, h_P)[h_{kn}]\|_{P,2}. \end{aligned} \tag{F.30}$$

Evaluating (F.30) at any $j < k^*$ and proceeding inductively from (F.29) then implies

$$\lim_{n\to\infty} \|\nabla m_{j,j}(W_j, h_P)[h_{jn}]\|_{P,2} = 0 \tag{F.31}$$

since $f_j^* = 0$ for all $j < k^*$. Finally evaluating (F.30) at $j = k^*$ and employing (F.31) implies $f_{k^*}^* = g^* \in \bar{\mathcal{R}}_{k^*}$. Since $1 < k^* \le J$ and $g^* \in L^2(W_{k^*})$ were arbitrary, it follows that $\bar{\mathcal{R}}_j = L^2(W_j)$ for all j. Thus, we conclude $\bar{\mathcal{R}} = \bigotimes_{j=1}^J L^2(W_j)$ if and only if $\bar{\mathcal{R}}_j = L^2(W_j)$ for all $1 \le j \le J$. Finally, let $\nabla m_{j,j}(W_j, h_P)^* : L^2(W_j) \to \mathbf{H}_j^*$ denote the adjoint of $\nabla m_{j,j}(W_j, h_P) : \mathbf{H}_j \to L^2(W_j)$. Theorem 6.6.1 in Luenberger (1969) then implies that $\bar{\mathcal{R}}_j^\perp = \{h \in L^2(W_j) : \nabla m_{j,j}(W_j, h_P)[h] = 0\}$. Therefore, we further obtain that $\bar{\mathcal{R}}_j = L^2(W_j)$ for all $1 \le j \le J$ if and only if $\{h \in L^2(W_j) : \nabla m_{j,j}(W_j, h_P)[h] = 0\} = \{0\}$ for all $1 \le j \le J$. ■

Proof of Corollary 4.1: Notice that the conditions of Lemma 4.2 are trivially satisfied. Therefore, Lemma 4.2 implies that $\bar{\mathcal{R}} = \bigotimes_{j=1}^J L^2(W_j)$ if and only if $\bar{\mathcal{R}}_j = L^2(W_j)$ for all $1 \le j \le J$, where

$\bar{\mathcal{R}}_j$ denotes the closure of $\mathcal{R}_j$ in $L^2(W_j)$, and $\mathcal{R}_j$ is given by

$$\mathcal{R}_j = \{f \in L^2(W_j) : f = d_j(W_j)h_j(W_j) \text{ for some } h_j \in \mathbf{H}_j\}. \quad (\text{F.32})$$

Hence, the claim of the Corollary follows if for all $1 \leq j \leq J$, $\bar{\mathcal{R}}_j = L^2(W_j)$ if and only $\mathbf{H}_j$ is dense in $L^2(W_j)$ and $P(d_j(W_j) \neq 0) = 1$. The conditions of the Corollary are equivalent for all $1 \leq j \leq J$, and therefore without loss of generality we focus on the case $j = 1$. To this end, we first suppose $\bar{\mathcal{R}}_1 = L^2(W_1)$ and define $f_1 \in L^2(W_1)$ by

$$f_1(W_1) \equiv 1\{d_1(W_1) = 0\}. \quad (\text{F.33})$$

Next observe that since $\bar{\mathcal{R}}_1 = L^2(W_1)$ by hypothesis, it follows $f_1 \in \bar{\mathcal{R}}_1$ and therefore

$$0 = \inf_{h_1 \in \mathbf{H}_1} E[\{d_1(W_1)h_1(W_1) - f_1(W_1)\}^2] \geq E[\{f_1(W_1)\}^2] = P(d_1(W_1) = 0), \quad (\text{F.34})$$

where in the first equality we exploited (F.32), the inequality follows from definition (F.33) implying $d_1(W_1)f_1(W_1) = 0$ almost surely, and the final equality results from (F.33). Hence, we conclude that if $\bar{\mathcal{R}}_1 = L^2(W_1)$, then $P(d_1(W_1) \neq 0) = 1$. Moreover, for any $h_1 \in \mathbf{H}_1 \subseteq L^2(W_1)$ we have $d_1 h_1 \in L^2(W_1)$ since d_1 is bounded, and thus

$$\begin{aligned} 0 &= \inf_{h_1 \in \mathbf{H}_1} E[\{d_1(W_1)h_1(W_1) - d_1(W_1)f(W_1)\}^2] \\ &= \min_{h_1 \in \bar{\mathbf{H}}_1} E[\{d_1(W_1)\}^2\{h_1(W_1) - f(W_1)\}^2], \end{aligned} \quad (\text{F.35})$$

for any $f \in L^2(W_1)$, and where the first equality follows from $\bar{\mathcal{R}}_1 = L^2(W_1)$, while the final equality holds for $\bar{\mathbf{H}}_1$ the closure of $\mathbf{H}_1$ in $L^2(W_1)$, and attainment of the infimum is guaranteed by the criterion being convex and diverging to infinity as $\|h_1\|_{P,2} \uparrow \infty$ and Proposition 38.15 in Zeidler (1984). Thus we conclude from (F.34) and (F.35) that for any $f \in L^2(W_1)$ there exists a $h_1 \in \bar{\mathbf{H}}_1$ such that $P(f(W_1) = h_1(W_1)) = 1$. Since $\mathbf{H}_1 \subseteq L^2(W_1)$ by hypothesis, we conclude that in fact $\bar{\mathbf{H}}_1 = L^2(W_1)$.

We next suppose instead that $\bar{\mathbf{H}}_1 = L^2(W_1)$ and $P(d_1(W_1) \neq 0) = 1$ and aim to establish that $\bar{\mathcal{R}}_1 = L^2(W_1)$. First, since $\nabla m_1(W_1, h_P)[h] = d_1 h_1$ for any $(h_1, \ldots, h_J) = h \in \mathbf{H}$ and d_1 is bounded, we may view $\nabla m_1(W_1, h_P)$ as a map from $\bar{\mathbf{H}}_1$ into $L^2(W_1)$ by, with some abuse of notation, setting $\nabla m_1(W_1, h_P)[h_1] = d_1 h_1$ for any $h_1 \in \bar{\mathbf{H}}_1 \setminus \mathbf{H}_1$ as well. Furthermore, since $\bar{\mathbf{H}}_1 = L^2(W_1)$ by hypothesis, direct calculation reveals that $\nabla m_1(W_1, h_P) : L^2(W_1) \to L^2(W_1)$ is self adjoint. Thus, Theorem 6.6.3 in Luenberger (1969) implies $\bar{\mathcal{R}}_1 = L^2(W_1)$ if and only if $\nabla m_1(W_1, h_P) : L^2(W_1) \to L^2(W_1)$ is injective. However, injectivity of $\nabla m_1(W_1, h_P) : L^2(W_1) \to L^2(W_1)$ is equivalent to $P(d_1(W_1) \neq 0) = 1$, and therefore $\bar{\mathcal{R}}_1 = L^2(W_1)$. ■

Previously, Ai and Chen (2012) derived the semiparametric efficiency bound for a general class of "smooth" functionals of P defined by nonparametric sequential moment restriction model (31). The next Theorem, which is a restatement of Theorem 4.2, exploits their results and our Corollary 3.1 to obtain an alternative characterization of local just identification of P by model (31). In the following we let Ω_f^* denote the semiparametric efficient variance bound for estimating population mean $\theta_f(P) \equiv \int f dP$ for f in any dense subset $\mathcal{D}$ of $L^2(P)$.

Theorem F.1. *Let Assumption 4.3 hold. Then: There is a dense subset $\mathcal{D}$ of $L^2(P)$ such that $\Omega_f^* = Var\{f(X)\}$ for all $f \in \mathcal{D}$ if and only if $\bar{\mathcal{R}} = \bigotimes_{j=1}^J L^2(W_j)$. Hence: P is locally just identified by model (31) if and only if $\bar{\mathcal{R}} = \bigotimes_{j=1}^J L^2(W_j)$.*

Proof of Theorem F.1: We let $L^\infty(P) \equiv \{f : |f| \text{ is bounded } P-\text{a.s.}\}$, $L^\infty(W_j)$ and $L^\infty(Z)$ be the subsets of $L^\infty(P)$ depending only on W_j and Z respectively. We recall that $L^2(W_j)$ and $L^2(Z)$ are analogously defined. In addition, we note that Assumption 4.3(i) implies that if $j \leq j'$ then it follows that

$$W_j = F(W_{j'}) \tag{F.36}$$

for some measurable function $F : \mathbf{R}^{d_{w_{j'}}} \to \mathbf{R}^{d_{w_j}}$; see, e.g., Theorem 20.1 in Billingsley (2008). We define a subset $\mathcal{Q} \subseteq L^2(P)$ as

$$\mathcal{Q} \equiv \{f : f(X) = \{\sum_{j=1}^J \rho_j(Z, h_P) q_j(W_j) + C\} \text{ a.s. for some } q_j \in L^2(W_j), C \in \mathbf{R}\}, \tag{F.37}$$

and $\bar{\mathcal{Q}}$ as the closure of $\mathcal{Q}$ under $\|\cdot\|_{P,2}$. We set the desired subset $\mathcal{D}$ to equal $\mathcal{D} \equiv L^\infty(P) \setminus \bar{\mathcal{Q}}$ and note $\mathcal{D}$ is a subset of $L^2(P)$ since $\mathcal{D} \subseteq L^\infty(P) \subset L^2(P)$. To establish that $\mathcal{D}$ is dense in $L^2(P)$ we let j^* be the smallest j satisfying $1 \leq j \leq J$ and such that $L^2(W_{j^*})$ is infinite dimensional – note existence of j^* is guaranteed by Assumption 4.3(vi). We next aim to show that

$$L^2(W_{j^*}) \cap \bar{\mathcal{Q}} \neq L^2(W_{j^*}) \tag{F.38}$$

and to this end, we note that since $L^2(W_{j^*})$ is infinite dimensional and $L^2(W_j)$ is finite dimensional for all $j < j^*$ it follows that there exists a $g \in L^2(W_{j^*})$ with $\|g\|_{P,2} > 0$ and

$$E\big[g(W_{j^*})\{\sum_{j=1}^{j^*-1} \rho_j(Z, h_P) q_j(W_j) + C\}\big] = 0 \tag{F.39}$$

for all $C \in \mathbf{R}$ and $q_j \in L^2(W_j)$ – here if $j^* = 1$ then (F.39) should be understood as just requiring $E[g(W_{j^*})] = 0$. On the other hand, Assumption 4.3(i) and the law of iterated expectations together imply that for any $q_j \in L^2(W_j)$ we have

$$E\big[g(W_{j^*})\{\sum_{j=j^*}^{J} \rho_j(Z, h_P) q_j(W_j)\}\big] = 0. \tag{F.40}$$

Thus, (F.39) and (F.40) imply $\|f - g\|_{P,2} = \|f\|_{P,2} + \|g\|_{P,2} > 0$ for any $f \in \bar{\mathcal{Q}}$, from which we conclude (F.38) holds. Since $L^\infty(W_{j^*})$ is dense in $L^2(W_{j^*})$, (F.38) further yields that there is a $\tilde{g} \in L^\infty(W_{j^*}) \setminus \bar{\mathcal{Q}}$ and for any $f \in L^\infty(P)$ and $\epsilon_n \downarrow 0$ we set

$$f_n = \begin{cases} f & \text{if } f \notin \bar{\mathcal{Q}} \\ f + \epsilon_n \tilde{g} & \text{if } f \in \bar{\mathcal{Q}} \end{cases} \tag{F.41}$$

and note $f_n \in \mathcal{D} \equiv L^\infty(P) \setminus \bar{\mathcal{Q}}$ and $\|f_n - f\|_{P,2} = o(1)$. We conclude that $\mathcal{D}$ is dense in $L^\infty(P)$ with respect to $\|\cdot\|_{P,2}$ and hence also in $L^2(P)$ since $L^\infty(P)$ is a dense subset of $L^2(P)$ under $\|\cdot\|_{P,2}$.

While we have so far avoided stating an explicit formulation for Ω_f^* for ease of exposition, it is now necessary to characterize it for all $f \in \mathcal{D}$. To this end, we follow Ai and Chen (2012) by setting $\varepsilon_J(Z,h) \equiv \rho_J(Z,h)$ and recursively defining

$$\varepsilon_s(Z,h) \equiv \rho_s(Z,h) - \sum_{j=s+1}^{J} \Gamma_{s,j}(W_j)\varepsilon_j(Z,h) \tag{F.42}$$

for $1 \leq j \leq J-1$, and where for any $1 \leq s < j \leq J$, the function $\Gamma_{s,j}(W_j)$ is given by

$$\Gamma_{s,j}(W_j) \equiv E[\rho_s(Z,h_P)\varepsilon_j(Z,h_P)|W_j]\{\Sigma_j(W_j)\}^{-1} \tag{F.43}$$

$$\Sigma_j(W_j) \equiv E[\{\varepsilon_j(Z,h_P)\}^2|W_j] \tag{F.44}$$

and we note Assumptions 4.3(i)(iv)(v) and simple calculations together imply

$$P(\eta \leq \Sigma_j(W_j) \leq M) = 1 \tag{F.45}$$

for all $1 \leq j \leq J$ and some $\eta, M \in (0,+\infty)$. We further set $\Sigma_f \equiv \text{Var}\{f(X)\}$ and define

$$\Sigma_0 \equiv \text{Var}\{f(X) - \sum_{j=1}^{J} \Lambda_j(W_j)\varepsilon_j(Z,h_P)\} \tag{F.46}$$

$$\Lambda_j(W_j) \equiv E[f(X)\varepsilon_j(Z,h_P)|W_j]\{\Sigma_j(W_j)\}^{-1} \tag{F.47}$$

and note: (i) $\Lambda_j(W_j) \in L^2(W_j)$ by (F.44), (F.45), $f \in \mathcal{D} \subset L^\infty(P)$, and Jensen's inequality; (ii) $\Sigma_0 > 0$ since $f \notin \bar{\mathcal{Q}}$ and $\bar{\mathcal{Q}}$ is closed; and (iii) by direct calculation

$$\Sigma_0 = \Sigma_f - \sum_{j=1}^{J} E[\{\Lambda_j(W_j)\}^2\Sigma_j(W_j)]. \tag{F.48}$$

Next, we define the maps $a_j(W_j,\cdot) : \mathbf{H} \to L^2(W_j)$ for any $1 \leq j \leq J$ to be given by

$$a_j(W_j,h) \equiv E[\varepsilon_j(Z,h)|W_j]. \tag{F.49}$$

We further note that result (F.45) and Assumption Assumption 4.3(v) imply by arguing inductively that $\Gamma_{s,j}(W_j) \in L^\infty(W_j)$. Hence, it can be shown from definition (F.42) and Assumption 4.3(ii) that the maps $a_j(W_j,\cdot)$ are Fréchet differentiable at h_P and we denote their derivatives by $\nabla a_j(W_j,h_P) : \mathbf{H} \to L^2(W_j)$. Therefore, the Fisher norm of a $s \in \mathbf{H}$ is

$$\|s\|_w^2 \equiv \sum_{j=1}^{J} E[\{\Sigma_j(W_j)\}^{-1}\{\nabla a_j(W_j,h_P)[s]\}^2] + \{\Sigma_0\}^{-1}\{E[\sum_{j=1}^{J}\Lambda_j(W_j)\nabla a_j(W_j,h_P)[s]]\}^2 \tag{F.50}$$

(see eq. (4) in Ai and Chen (2012)), and we note $\|s\|_w < \infty$ for any $s \in \mathbf{H}$ since $\nabla a_j(W_j,h_P)[s] \in L^2(W_j)$, $\{\Sigma_j(W_j)\}^{-1} \in L^\infty(W_j)$ by (F.45), and as argued $\Lambda_j(W_j) \in L^\infty(W_j)$. Letting $\mathcal{W}$ denote the closure of $\mathbf{H}$ under $\|\cdot\|_w$, we then obtain

$$\{\Omega_f^*\}^{-1} = \inf_{s\in\mathcal{W}} \Big\{\{\Sigma_0\}^{-1}\{1+\sum_{j=1}^J E[\Lambda_j(W_j)\nabla a_j(W_j,h_P)[s]]\}^2 + \sum_{j=1}^J E[\{\Sigma_j(W_j)\}^{-1}\{\nabla a_j(W_j,h_P)[s]\}^2]\Big\} \quad \text{(F.51)}$$

by Theorem 2.1 in Ai and Chen (2012).

It is convenient for our purposes, however, to exploit the structure of our problem to further simplify the characterization in (F.51). To this end, note that (F.50) and the Cauchy-Schwarz inequality imply that the objective in (F.51) is continuous under $\|\cdot\|_w$. Hence, since $\mathcal{W}$ is the completion of $\mathbf{H}$ under $\|\cdot\|_w$, it follows from (F.51) that

$$\{\Omega_f^*\}^{-1} = \inf_{s\in\mathbf{H}} \Big\{\{\Sigma_0\}^{-1}\{1+\sum_{j=1}^J E[\Lambda_j(W_j)\nabla a_j(W_j,h_P)[s]]\}^2 + \sum_{j=1}^J E[\{\Sigma_j(W_j)\}^{-1}\{\nabla a_j(W_j,h_P)[s]\}^2]\Big\}. \quad \text{(F.52)}$$

Next, note that we may be view $(\nabla a_1(W_1,h_P),\ldots,\nabla a_J(W_J,h_P))$ as a map from $\mathbf{H}$ onto the product space $\bigotimes_{j=1}^J L^2(W_j)$, and we denote the range of this map by

$$\mathcal{A} \equiv \{\{r_j\}_{j=1}^J \in \bigotimes_{j=1}^J L^2(W_j) : \text{ for some } s\in\mathbf{H}, r_j = \nabla a_j(W_j,h_P)[s] \text{ for all } 1\le j\le J\} \quad \text{(F.53)}$$

and let $\bar{\mathcal{A}}$ denote the closure of $\mathcal{A}$ in the product topology. Result (F.52) then implies

$$\begin{aligned}\{\Omega_f^*\}^{-1} &= \inf_{\{r_j\}\in\bar{\mathcal{A}}} \{\{\Sigma_0\}^{-1}\{1+\sum_{j=1}^J E[\Lambda_j(W_j)r_j(W_j)]\}^2 + \sum_{j=1}^J E[\{\Sigma_j(W_j)\}^{-1}\{r_j(W_j)\}^2]\} \\ &= \min_{\{r_j\}\in\bar{\mathcal{A}}} \{\{\Sigma_0\}^{-1}\{1+\sum_{j=1}^J E[\Lambda_j(W_j)r_j(W_j)]\}^2 + \sum_{j=1}^J E[\{\Sigma_j(W_j)\}^{-1}\{r_j(W_j)\}^2]\} \quad \text{(F.54)}\end{aligned}$$

where attainment follows from $\bar{\mathcal{A}}$ being a vector space since $(\nabla a_1(W_1,h_P),\ldots,\nabla a_J(W_J,h_P))$ is linear and $\mathbf{H}$ is a vector space, the criterion in (F.54) being convex and diverges to infinity as $\sum_j \|r_j\|_{P,2}\uparrow\infty$, and Proposition 38.15 in Zeidler (1984). In particular, note that if $\{r_j^*\}\in\bar{\mathcal{A}}$ is the minimizer of (F.54), then for any $\{\delta_j\}\in\bar{\mathcal{A}}$

$$\sum_{j=1}^J E[\delta_j(W_j)\{\{\Sigma_j(W_j)\}^{-1}r_j^*(W_j) + \Sigma_0^{-1}\Lambda_j(W_j)\{1+\sum_{s=1}^J E[\Lambda_s(W_s)r_s^*(W_s)]\}\}] = 0. \quad \text{(F.55)}$$

Next, we aim to solve the optimization in (F.55) under the hypothesis that $\bar{\mathcal{A}} = \bigotimes_{j=1}^J L^2(W_j)$.

In that case, (F.55) must hold for all $\{\delta_j\} \in \bigotimes_{j=1}^J L^2(W_j)$ which implies

$$r_j^*(W_j) = -\Sigma_0^{-1}\{1 + \sum_{s=1}^J E[\Lambda_s(W_s)r_s^*(W_s)]\}\Lambda_j(W_j)\Sigma_j(W_j). \tag{F.56}$$

It is evident from (F.56) that $r_j^*(W_j) = -\Lambda_j(W_j)\Sigma_j(W_j)C_0$ for some $C_0 \in \mathbf{R}$ independent of j, and plugging into (F.56) we solve for C_0 and exploit (F.48) to find

$$r_j^*(W_j) = -\{\Sigma_f\}^{-1}\Lambda_j(W_j)\Sigma_j(W_j). \tag{F.57}$$

Thus combining (F.54) and (F.57), and repeatedly exploiting (F.48) we conclude

$$\begin{aligned}\{\Omega_f^*\}^{-1} &= \Sigma_0^{-1}\{1 - \{\Sigma_f\}^{-1}\sum_{j=1}^J E[\Lambda_j^2(W_j)\Sigma_j(W_j)]\}^2 + \{\Sigma_f\}^{-2}\sum_{j=1}^J E[\Lambda_j^2(W_j)\Sigma_j(W_j)] \\ &= \Sigma_0^{-1}\{1 - \{\Sigma_f\}^{-1}\{\Sigma_f - \Sigma_0\}\}^2 + \{\Sigma_f\}^{-2}\{\Sigma_f - \Sigma_0\} = \{\Sigma_f\}^{-1}, \end{aligned} \tag{F.58}$$

or equivalently $\Omega_f^* = \Sigma_f$. While (F.58) was derived while supposing $\bar{\mathcal{A}} = \bigotimes_{j=1}^J L^2(W_j)$, we note that since $\bar{\mathcal{A}} \subseteq \bigotimes_{j=1}^J L^2(W_j)$, the minimum in (F.54) is attained, and $r_j^*(W_j) = -\{\Sigma_f\}^{-1}\Lambda_j(W_j)\Sigma_j(W_j)$ is the unique minimizer on $\bigotimes_{j=1}^J L^2(W_j)$, we must have

$$\Omega_f^* = \Sigma_f \text{ if and only if } \{-\Sigma_f^{-1}\Lambda_j(W_j)\Sigma_j(W_j)\}_{j=1}^J \in \bar{\mathcal{A}}. \tag{F.59}$$

Since result (F.59) holds for all $f \in \mathcal{D}$ and $\bar{\mathcal{A}}$ is a vector space, (F.47) implies

$$\Omega_f^* = \Sigma_f \ \forall f \in \mathcal{D} \text{ if and only if } \{E[f(X)\epsilon_j(Z, h_P)|W_j]\}_{j=1}^J \in \bar{\mathcal{A}} \ \forall f \in \mathcal{D}. \tag{F.60}$$

Also note that if $\|f_n - f\|_{P,2} = o(1)$, then by the Cauchy-Schwarz inequality we obtain

$$\begin{aligned}&\lim_{n\to\infty} E[\{E[f_n(X)\varepsilon_j(Z, h_P)|W_j] - E[f(X)\varepsilon_j(Z, h_P)|W_j]\}^2] \\ &\qquad\qquad \leq \lim_{n\to\infty} E[\{f_n(X) - f(X)\}^2\Sigma_j(W_j)] = 0\end{aligned} \tag{F.61}$$

where the final equality follows from $\Sigma_j(W_j) \in L^\infty(W_j)$ by result (F.45). Therefore, since as argued $\mathcal{D}$ is a dense subset of $L^2(P)$, in addition $\bar{\mathcal{A}}$ is closed under the product topology in $\bigotimes_{j=1}^J L^2(W_j)$, and result (F.63) holds for all $1 \leq j \leq J$, we conclude

$$\Omega_f^* = \Sigma_f \ \forall f \in \mathcal{D} \text{ if and only if } \{E[f(X)\epsilon_j(Z, h_P)|W_j]\}_{j=1}^J \in \bar{\mathcal{A}} \ \forall f \in L^2(P). \tag{F.62}$$

Next, fix an arbitrary $\{g_j\}_{j=1}^J \in \bigotimes_{j=1}^J L^\infty(W_j)$ and note that result (F.45) then yields

$$f_0(X) \equiv \sum_{j=1}^J g_j(W_j)\epsilon_j(Z, h_P)\{\Sigma_j(W_j)\}^{-1} \tag{F.63}$$

belongs to $L^2(P)$ since $g_j \in L^\infty(W_j)$. Since $E[\{\epsilon_j(Z, h_P)\}^2|W_j] = \Sigma_j(W_j)$, $E[\epsilon_j(Z, h_P)\epsilon_s(Z, h_P)|W_j] =$

0 whenever $s < j$, we obtain from result (F.36) that

$$E[f_0(X)\epsilon_j(Z,h_P)|W_j] = E[(\sum_{s=1}^{J} g_s(W_s)\epsilon_s(Z,h_P)\{\Sigma_s(W_s)\}^{-1})\epsilon_j(Z,h_P)|W_j]$$
$$= \sum_{s=1}^{J} E[g_s(W_s)\{\Sigma_s(W_s)\}^{-1}E[\epsilon_s(Z,h_P)\epsilon_j(Z,h_P)|W_{s\vee j}]|W_j] = g_j(W_j). \quad \text{(F.64)}$$

In particular, (F.64) holds for any $1 \leq j \leq J$, and since $\{g_j\}_{j=1}^J \in \bigotimes_{j=1}^J L^\infty(W_j)$ was arbitrary, it follows that if $\{E[f(X)\epsilon_j(Z,h_P)|W_j]\}_{j=1}^J \in \bar{\mathcal{A}}$ for all $f \in L^2(P)$, then $\bigotimes_{j=1}^J L^\infty(W_j) \subseteq \bar{\mathcal{A}}$. However, since $\bar{\mathcal{A}}$ is closed in the product topology of $\bigotimes_{j=1}^J L^2(W_j)$, we have that if $\bigotimes_{j=1}^J L^\infty(W_j) \subseteq \bar{\mathcal{A}}$, then $\bigotimes_{j=1}^J L^2(W_j) = \bar{\mathcal{A}}$, and hence (F.62) yields

$$\Omega_f^* = \Sigma_f \ \forall f \in \mathcal{D} \text{ if and only if } \bigotimes_{j=1}^J L^2(W_j) = \bar{\mathcal{A}}. \quad \text{(F.65)}$$

To conclude, we note that since, as previously argued, $\Gamma_{s,j}(W_j) \in L^\infty(W_j)$ for all $1 \leq s < j \leq J$, definitions (F.43) and (F.49) and an inductive calculation imply that

$$\nabla m_j(Z,h_P)[s] = \nabla a_j(Z,h_P)[s] + \sum_{k=j+1}^{J} E[\Gamma_{j,k}(W_k)\nabla a_k(Z,h_P)[s]|W_j] \quad \text{(F.66)}$$

with $\nabla m_J(Z,h_P)[s] = \nabla a_J(Z,h_P)[s]$. Thus, from (F.66) we conclude $\bar{\mathcal{A}} = \bigotimes_{j=1}^J L^2(W_j)$ if and only if $\bar{\mathcal{R}} = \bigotimes_{j=1}^J L^2(W_j)$ and therefore the Theorem follows from (F.65). ■

Appendix G - Sufficient Conditions for Assumption 3.1

In this Appendix we illustrate how to construct a statistic $\hat{\mathbb{G}}_n$ satisfying Assumption 3.1 in the context of models defined by nonparametric conditional moment restrictions as studied in Section 4. Concretely, we let $\{X_i = (Z_i, W_i)\}_{i=1}^n$ be a random sample from the distribution P satisfying model (26), which is restated below for the purpose of easy reference.

$$E[\rho_j(Z_i,h_P)|W_{ij}] = 0 \text{ for all } 1 \leq j \leq J \text{ for some } h_P \in \mathbf{H}. \quad \text{(G.1)}$$

The parameter h_P can be estimated via the method of sieves by regularizing through either the choice of sieve, employing a penalization, or a combination of both approaches (Chen and Pouzo, 2012). Here, we assume $h_P \in \mathcal{H} \subseteq \mathbf{H}$, and consider a sequence of sieve spaces $\mathcal{H}_k \subseteq \mathcal{H}_{k+1} \subseteq \mathcal{H}$, with $\mathcal{H}_k$ growing suitably dense in $\mathcal{H}$ as k diverges to infinity. In turn, we estimate the unknown conditional expectation by series regression. Specifically, for $\{p_{jl}\}_{l=1}^\infty$ a sequence of approximating functions in $L^2(W_j)$, we let $p_j^{l_{jn}}(w_j) \equiv (p_{j1}(w_j),\ldots,p_{jl_{jn}}(w_j))'$, set $P_{jn} \equiv (p_j^{l_{jn}}(W_{1j}),\ldots,p_j^{l_{jn}}(W_{nj}))'$, and define

$$\hat{m}_j(w_j,h) \equiv \{\sum_{i=1}^n \rho_j(Z_i,h)p_j^{l_{jn}}(W_{ji})'\}(P_{jn}'P_{jn})^- p_j^{l_{jn}}(w_j) \quad \text{(G.2)}$$

where $(P_{jn}'P_{jn})^-$ denotes the Moore-Penrose pseudoinverse of $P_{jn}'P_{jn}$. For a sequence k_n diverging to infinity with the sample size, the estimator $\hat{h}_n$ is then defined as

$$\hat{h}_n \in \arg\min_{h \in \mathcal{H}_{k_n}} \sum_{i=1}^{n} \sum_{j=1}^{J} \hat{m}_j^2(W_{ij}, h). \tag{G.3}$$

See Chen and Pouzo (2012) and references therein for sufficient conditions for the convergence rates of $\hat{h}_n$ to h_P.

For a set $\mathbf{T}$ and known function $\psi_j : \mathbf{W}_j \times \mathbf{T} \to \mathbf{R}$, we let $\psi(w,\tau) \equiv (\psi_1(w_1,\tau), \ldots, \psi_J(w_J,\tau))'$ similarly define $\rho(z,h) \equiv (\rho_1(z,h), \ldots, \rho_J(z,h))'$ and set

$$\hat{\mathbb{G}}_n(\tau) \equiv \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \{\psi(W_i,\tau)\}' \rho(Z_i, \hat{h}_n). \tag{G.4}$$

Note that the resulting process $\hat{\mathbb{G}}_n$ may be viewed as an element of $\ell^\infty(\mathbf{T})$ provided that the functions $\psi_j(W_{ij}, \cdot)$ are bounded almost surely. To see why $\hat{\mathbb{G}}_n$ might satisfy Assumption 3.1, we observe that for any $\tau \in \mathbf{T}$, $\hat{\mathbb{G}}_n(\tau)$ is an estimator of the parameter

$$\theta_P(\tau) \equiv E[\{\psi(W,\tau)\}' \rho(Z, h_P)]. \tag{G.5}$$

However, since h_P satisfies (G.1) by hypothesis, the model in fact dictates that $\theta_P(\tau) = 0$, and thus the efficient estimator for $\theta_P(\tau)$ is simply zero. As a result, $\hat{\mathbb{G}}_n(\tau)$ is an inefficient estimator of $\theta_P(\tau)$, and by Lemma 3.1 it should satisfy Assumption 3.1 provided that it is regular and asymptotically linear. Similarly, $\hat{\mathbb{G}}_n$ could be constructed so that specification tests built on it aim their power at particular violations of the model by setting $\hat{\mathbb{G}}_n(\tau)$ to be the efficient estimator of $\theta_P(\tau)$ under the maintained alternative model; see Lemma 3.2(ii) and related discussion.

Denote $m(W,h) \equiv (m_1(W_1,h), \ldots, m_J(W_J,h))'$. We assume that the maps $m_j(W_j,h) \equiv E[\rho_j(Z,h)|W_j]$, $j = 1, ..., J$, are Fréchet differentiable at h_P with derivative $\nabla m_j(W_j, h_P) : \mathbf{H} \to L^2(W_j)$ (i.e., we impose Assumption 4.1(ii) as in Section 4). Recall that the linear map $\nabla m(W, h_P) : \mathbf{H} \to \bigotimes_{j=1}^{J} L^2(W_j)$ is given by

$$\nabla m(W,h_P)[h] \equiv \big(\nabla m_1(W_1,h_P)[h], \ldots, \nabla m_J(W_J,h_P)[h]\big)', \tag{G.6}$$

and its range space equals

$$\mathcal{R} \equiv \Big\{ f \in \bigotimes_{j=1}^{J} L^2(W_j) : f = \nabla m(W,h_P)[h] \text{ for some } h \in \mathbf{H} \Big\}, \tag{G.7}$$

which is closed under addition, and its norm closure (in $\bigotimes_{j=1}^{J} L^2(W_j)$), denoted $\bar{\mathcal{R}}$, is a vector subspace of $\bigotimes_{j=1}^{J} L^2(W_j)$. With some abuse of notation, for any $(f_1, \ldots, f_J) = f \in \bigotimes_{j=1}^{J} L^2(W_j)$ we let $\|f\|_{P,2}^2 = \sum_{j=1}^{J} \int f_j^2 dP$ and we observe $\bigotimes_{j=1}^{J} L^2(W_j)$ is a Hilbert space under $\|\cdot\|_{P,2}$ and its corresponding inner product. Therefore, since $\bar{\mathcal{R}}$ is a closed subspace of $\bigotimes_{j=1}^{J} L^2(W_j)$ we obtain from Theorem 3.4.1 in Luenberger (1969) that

$$\bigotimes_{j=1}^{J} L^2(W_j) = \bar{\mathcal{R}} \oplus \bar{\mathcal{R}}^{\perp}. \tag{G.8}$$

For any $f \in \bigotimes_{j=1}^J L^2(W_j)$ we let $\Pi_{\bar{\mathcal{R}}} f$ and $\Pi_{\bar{\mathcal{R}}^\perp} f$ denote the projection of f under $\|\cdot\|_{P,2}$ onto $\bar{\mathcal{R}}$ and $\bar{\mathcal{R}}^\perp$ respectively. We emphasize that the projection of $(f_1, \ldots, f_J) = f \in \bigotimes_{j=1}^J L^2(W_j)$ onto $\bar{\mathcal{R}}$ *need not equal* a coordinate by coordinate projection of f. Finally, recall that by Theorem 4.1, P is locally just identified if and only if $\bar{\mathcal{R}} = \bigotimes_{j=1}^J L^2(W_j)$, or if and only if $\bar{\mathcal{R}}^\perp = \{0\}$.

We in addition impose the following assumptions to study the process $\hat{\mathbb{G}}_n$.

Assumption G.1. *(i)* $\mathcal{F} \equiv \{f = \{\{\Pi_{\bar{\mathcal{R}}^\perp}\psi(\cdot,\tau)\}'\rho(\cdot,h) : (\tau,h) \in \mathbf{T} \times \mathcal{H}\}$ *is* $P-$*Donsker; (ii)* $\|\Pi_{\bar{\mathcal{R}}^\perp}\psi(w,\tau)\|$ *is bounded on* $\bigotimes_{j=1}^J \mathbf{W}_j \times \mathbf{T}$*; (iii)* $\mathcal{H}_k \subseteq \mathcal{H}$ *for all* k*.*

Assumption G.2. *(i)* $\sum_{j=1}^J \|\rho_j(\cdot,\hat{h}_n) - \rho_j(\cdot,h_P)\|_{P,2} = o_p(1)$*; (ii)* $E[\|m(W_i,\hat{h}_n) - m(W_i,h_P) - \nabla m(W_i,h_P)[\hat{h}_n - h_P]\|] = o_p(n^{-1/2})$*, (iii)* $\frac{1}{n}\sum_{i=1}^n \{\Pi_{\bar{\mathcal{R}}}\psi(W_i,\tau)\}'\rho(Z_i,\hat{h}_n) = o_p(n^{-1/2})$ *uniformly in* $\tau \in \mathbf{T}$*.*

Assumption G.1(i) ensures that the empirical process indexed by $f \in \mathcal{F}$ converges in distribution in $\ell^\infty(\mathbf{T})$, Assumption G.1(ii) demands that the weights in the linear combinations of moments be bounded, and Assumption G.1(iii) implies that $\hat{h}_n \in \mathcal{H}$ with probability one. Assumption G.2 imposes high level conditions on $\hat{h}_n$ that are transparent in their role played in the proof, though they can be verified under lower level requirements on the sieve bases, the sieve approximation errors, and the smoothness of the map $m(W,\cdot)$ near h_P. In particular, Assumption G.2(i) imposes that $\rho(\cdot,\hat{h}_n)$ be consistent for $\rho(\cdot,h_P)$ in $\bigotimes_{j=1}^J L^2(P)$. Assumption G.2(ii) demands the rate of convergence of $\hat{h}_n$ to be sufficiently fast to enable us to obtain a suitable expansion of $m(W_i,\hat{h}_n)$ around h_P. Both Assumption G.2(i) and G.2(ii) can be verified under lower level conditions by employing the results in Chen and Pouzo (2012). Finally, Assumption G.2(iii) intuitively follows from $\hat{h}_n$ satisfying (G.3) and $\nabla\hat{m}(W_i,\hat{h}_n)$ approximating $\nabla m(W_i,h_P)$;[2] see Ai and Chen (2003) and Chen and Pouzo (2009) for related arguments.

We next establish the asymptotic behavior of $\hat{\mathbb{G}}_n$.

Lemma G.1. *Let Assumptions 4.1(i)(ii), and G.1 and G.2 hold. Then:*

$$\hat{\mathbb{G}}_n(\tau) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \{\Pi_{\bar{\mathcal{R}}^\perp}\psi(W_i,\tau)\}'\rho(Z_i,h_P) + o_p(1) \tag{G.9}$$

uniformly in $\tau \in \mathbf{T}$*, and* $\hat{\mathbb{G}}_n \xrightarrow{L} \mathbb{G}_0$ *in* $\ell^\infty(\mathbf{T})$ *for some tight Gaussian measure* $\mathbb{G}_0$*.*

Lemma G.1 establishes the asymptotic linearity of $\hat{\mathbb{G}}_n$ as a process in $\ell^\infty(\mathbf{T})$. Since the influence function of $\hat{\mathbb{G}}_n$ obeys a functional central limit theorem by Assumption G.1(i), the conclusion that $\hat{\mathbb{G}}_n$ converges to a tight Gaussian process is immediate from result (G.9). Therefore, given Lemma G.1, the main requirement remaining in verifying $\hat{\mathbb{G}}_n$ satisfies Assumption 3.1 is showing that the influence function of $\hat{\mathbb{G}}_n(\tau)$ is orthogonal to the scores of the model for any $\tau \in \mathbf{T}$. However, the latter claim is immediate from the characterization of $\bar{T}(P)^\perp$ derived in Theorem 4.1.

Proof of Lemma G.1: We first note that Assumption G.2(iii) allows us to conclude

$$\hat{\mathbb{G}}_n(\tau) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \{\Pi_{\bar{\mathcal{R}}^\perp}\psi(W_i,\tau)\}'\rho(Z_i,\hat{h}_n) + o_p(1) \tag{G.10}$$

[2] Recall $\Pi_{\bar{\mathcal{R}}}\psi(W_i,\tau) = \nabla m(W_i,h_P)[v_n] + o(1)$ for some sequence $\{v_n\}_{n=1}^\infty \in \mathbf{H}$, while $\hat{h}_n$ solving (G.3) can be exploited to show $\frac{1}{n}\sum_{i=1}^n \{\nabla\hat{m}(W_i,\hat{h}_n)[v_n]\}'\rho(Z_i,\hat{h}_n) = o_p(n^{-1/2})$.

uniformly in $\tau \in \mathbf{T}$ since $\psi(W_i,\tau) = \Pi_{\mathcal{R}}\psi(W_i,\tau) + \Pi_{\mathcal{R}^\perp}\psi(W_i,\tau)$. Moreover, by the Cauchy Schwarz inequality, and Assumptions G.1(ii) and G.2(i) we obtain that

$$E[(\Pi_{\mathcal{R}^\perp}\psi(W_i,\tau)'\{\rho(Z_i,\hat{h}_n) - \rho(Z_i,h_P)\})^2] \\ \leq \sup_{(w,\tau)} \|\psi(w,\tau)\|^2 \times \|\rho(\cdot,\hat{h}_n) - \rho(\cdot,h_P)\|^2_{P,2} = o_p(1). \quad (G.11)$$

Thus, since $\mathcal{F} = \{f(x) = \{\Pi_{\mathcal{R}^\perp}\psi(w,\tau)\}'\rho(z,h) : (\tau,h) \in \mathbf{T}\times\mathcal{H}\}$ is P-Donsker by Assumption G.1(i) and $\hat{h}_n \in \mathcal{H}$ by Assumption G.1(iii), result (G.11) yields

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n\{\Pi_{\mathcal{R}^\perp}\psi(W_i,\tau)\}'\rho(Z_i,\hat{h}_n) - \frac{1}{\sqrt{n}}\sum_{i=1}^n\{\Pi_{\mathcal{R}^\perp}\psi(W_i,\tau)\}'\rho(Z_i,h_P) \\ = \sqrt{n}E[\{\Pi_{\mathcal{R}^\perp}\psi(W_i,\tau)\}'\{\rho(Z_i,\hat{h}_n) - \rho(Z_i,h_P)\}] + o_p(1) \quad (G.12)$$

uniformly in $\tau \in \mathbf{T}$. Furthermore, the law of iterated expectations, the Cauchy-Schwarz inequality, and Assumptions G.1(ii) and G.2(ii) together yield uniformly in $\tau \in \mathbf{T}$ that

$$\sqrt{n}E[\{\Pi_{\mathcal{R}^\perp}\psi(W_i,\tau)\}'\{\rho(Z_i,\hat{h}_n) - \rho(Z_i,h_P)\}] \\ = \sqrt{n}E[\{\Pi_{\mathcal{R}^\perp}\psi(W_i,\tau)\}'\nabla m(W_i,h_P)[\hat{h}_n - h_P]] + o_p(1) = o_p(1) \quad (G.13)$$

where in the final equality we exploited that by definition of $\mathcal{R}$ and $\mathcal{R}^\perp$ it follows that for any $h \in \mathbf{H}$ we have $E[\{\Pi_{\mathcal{R}^\perp}\psi(W_i,\tau)\}'\nabla m(W_i,h_P)[h]] = 0$. Hence, the Lemma follows from results (G.10), (G.12), and (G.13), and the class $\mathcal{F} = \{f(x) = \{\Pi_{\mathcal{R}^\perp}\psi(w,\tau)\}'\rho(z,h) : (\tau,h) \in \mathbf{T}\times\mathcal{H}\}$ being P-Donsker by Assumption G.1(i). ■

Appendix H - Examples for Section 4

In this appendix we provide additional discussions on Examples 4.1, 4.2 and 4.3 to illustrate how to employ Theorem 4.1, Lemmas 4.1 and 4.2, and Corollary 4.1 to determine whether P is locally overidentifed by the model $\mathbf{P}$ in specific applications. We also introduce a final example based on DiNardo et al. (1996).

Example 4.1. In this application Z represents the distinct elements of $(V, Y_1, \ldots, Y_J)$ and there are J moment restrictions. For any $h = (h_1,\ldots,h_J) \in \mathbf{H} = \bigotimes_{j=1}^J \mathbf{H}_j$, each $\rho_j : \mathbf{Z}\times\mathbf{H} \to \mathbf{R}$ then equals

$$\rho_j(Z,h) = Y_j - h_j(V). \quad (H.1)$$

Therefore, for any $h = (h_1,\ldots,h_J) \in \mathbf{H}$, $m_j(W_j,h) = E[Y_j - h_j(V)|W_j]$ which is affine and continuous by Jensen's inequality and $\mathbf{H}_j \subseteq L^2(V)$. Hence, $m_j(W_j,h) : \mathbf{H} \to L^2(W_j)$ is Fréchet differentiable with

$$\nabla m_j(W_j,h_P)[h] = -E[h_j(V)|W_j] \quad (H.2)$$

for any $h = (h_1,\ldots,h_J) \in \mathbf{H}$. In particular, note that the conditions of Lemma 4.2 are trivially satisfied since $\nabla m_{j,k}(W_j,h_P)[h_k] = 0$ for all $k \neq j$ and $1 \leq j \leq J$. Hence, defining

$$\mathcal{R}_j \equiv \{f \in L^2(W_j) : f = E[h_j(V)|W_j] \text{ for some } h_j \in \mathbf{H}_j\}, \quad (H.3)$$

we conclude from Lemma 4.2 that P is locally just identified if and only if $\bar{\mathcal{R}}_j = L^2(W_j)$ for all $1 \leq j \leq J$ – i.e. we may study local overidentification by examining each moment condition separately. We gain insight into the condition $\bar{\mathcal{R}}_j = L^2(W_j)$ by considering two separate cases.

Case I: We first suppose $W_j = V$ (i.e. V is exogenous in the j-th moment restriction). In this case, we may view $E[\cdot|W_j] : \mathbf{H}_j \to L^2(W_j)$ as the identity mapping, and hence $\bar{\mathcal{R}}_j = L^2(W_j)$ whenever $\mathbf{H}_j = L^2(V)$. Notice in fact that by Corollary 4.1, $\bar{\mathcal{R}}_j$ continues to satisfy $\bar{\mathcal{R}}_j = L^2(W_j)$ if we set $\mathbf{H}_j$ to be any Banach space that is dense in $L^2(W_j)$, such as the set of bounded functions, continuous functions, or differentiable functions. On the other hand, Corollary 4.1 implies $\bar{\mathcal{R}}_j \neq L^2(W_j)$ whenever $\mathbf{H}_j$ is a strict closed subspace of $L^2(V)$, which occurs, for example, when we impose a partially linear or an additively separable specification for h_j (Robinson, 1988; Stone, 1985).

Case II: We next consider the case where W_j is an instrument, so that $W_j \neq V$. We let $\mathbf{H}_j = L^2(V)$, the condition that the closure of the range of $E[\cdot|W_j] : L^2(V) \to L^2(W_j)$ be equal to $L^2(W_j)$ is most easily interpreted through Lemma 4.2. Note that the adjoint of $E[\cdot|W_j]$ is $E[\cdot|V] : L^2(W_j) \to L^2(V)$. Thus, Lemma 4.2 implies that $\bar{\mathcal{R}}_j = L^2(W_j)$ if and only if

$$\{0\} = \{f \in L^2(W_j) : E[f(W_j)|V] = 0\}. \tag{H.4}$$

The requirement in (H.4) is known as the distribution of (V, W_j) being L^2-complete with respect to W_j, which is an untestable property of the distribution of the data (Andrews, 2017; Canay et al., 2013). As in Case I, however, we may obtain $\bar{\mathcal{R}}_j \neq L^2(W_j)$ by restricting the parameter space for h_j. Suppose, for example, that $\mathbf{H}_j$ is a closed subspace of $L^2(V)$, such as in a partially linear or an additive separable specification. For any $f \in L^2(W_j)$, then let

$$\Pi_{\mathbf{H}_j} f \equiv \arg\min_{h \in \mathbf{H}_j} \|f - h\|_{P,2} \tag{H.5}$$

and note $\Pi_{\mathbf{H}_j} : L^2(W_j) \to \mathbf{H}_j$ is the adjoint of $E[\cdot|W_j] : \mathbf{H}_j \to L^2(W_j)$. Thus, applying Lemma 4.2 we obtain that $\bar{\mathcal{R}}_j = L^2(W_j)$ if and only if

$$\{0\} = \{f \in L^2(W_j) : \Pi_{\mathbf{H}_j} f = 0\}. \tag{H.6}$$

Condition (H.6) may be viewed as a generalization of (H.4), and can be violated even when $\mathbf{H}_j$ is infinite dimensional yet a strict subspace of $L^2(V)$. ■

Example 4.2. We will study a general nonparametric specification for the parameter space and aim to show P is nonetheless locally overidentified. To this end, let

$$C^1([0,1]) \equiv \{f : [0,1] \to \mathbf{R} : f \text{ is continuously differentiable on } [0,1]\}, \tag{H.7}$$

which is a Banach space when endowed with the norm $\|f\|_{C^1} \equiv \|f\|_\infty + \|f'\|_\infty$ for f' the derivative of f. We then set the parameter space $\mathbf{H}$ to be given by

$$\mathbf{H} = L^\infty((V,R)) \times L^2(V) \times L^2(V) \times C^1([0,1]) \times C^1([0,1]), \tag{H.8}$$

and assume $(s_P, g_{0,P}, g_{1,P}, \lambda_{0,P}, \lambda_{1,P}) = h_P \in \mathbf{H}$ – i.e. we require the $\lambda_{d,P}$ functions in (42) to be continuously differentiable. For $X = (Y, D, V, R)$ and $W_1 = (V, R)$, the moment restriction in (43) corresponds to for any $(s, g_0, g_1, \lambda_0, \lambda_1) = h \in \mathbf{H}$ setting

$$\rho_1(Z, h) = D - s(V, R). \tag{H.9}$$

In turn, for the moment restriction in (44) we let $W_2 = (V, R, D)$ and define

$$\rho_2(Z, h) = D\{Y - g_1(V) - \lambda_1(s(V, R))\} + (1 - D)\{Y - g_0(V) - \lambda_0(s(V, R))\} \tag{H.10}$$

for any $(s, g_0, g_1, \lambda_0, \lambda_1) = h \in \mathbf{H}$.[3] We note that by (H.9), $m_1(W_1, \cdot) : \mathbf{H} \to L^2(W_1)$ is affine and continuous and therefore Fréchet differentiable with derivative

$$\nabla m_1(W_1, h_P)[h] = -s(V, R) \tag{H.11}$$

for any $(s, g_0, g_1, \lambda_0, \lambda_1) = h \in \mathbf{H}$. The second restriction is Fréchet differentiable as well, and for any $(s, g_0, g_1, \lambda_0, \lambda_1) \in \mathbf{H}$, $\nabla m_2(W_2, h_P) : \mathbf{H} \to L^2(W_2)$ is given by

$$\begin{aligned}\nabla m_2(W_2, h_P)[h] = D\{&-g_1(V) - \lambda_1(s_P(V, R)) - \lambda'_{1,P}(s_P(V, R))s(V, R)\} \\ &+ (1 - D)\{-g_0(V) - \lambda_0(s_P(V, R)) - \lambda'_{0,P}(s_P(V, R))s(V, R)\}.\end{aligned} \tag{H.12}$$

To verify this claim, first note $\nabla m_2(W_{2,}, h_P) : \mathbf{H} \to L^2(W_2)$ is continuous when $\mathbf{H}$ is endowed with the product topology. Moreover, by the mean value theorem we have

$$\begin{aligned}\{\lambda_{d,P} + \lambda_d\}(s_P(V, R) + s(V, R)) - \{\lambda_{d,P} + \lambda_d\}(s_P(V, R)) - \lambda'_{d,P}(s_P(V, R))s(V, R) \\ = (\lambda'_{d,P}(\bar{s}(V, R)) - \lambda'_{d,P}(s_P(V, R)))s(V, R) + \lambda'_d(\bar{s}(V, R))s(V, R)\end{aligned} \tag{H.13}$$

for some $\bar{s}(V, R)$ a convex combination of $s_P(V, R)$ and $s_P(V, R) + s(V, R)$. Exploiting (H.10), (H.12), and (H.13) we can then obtain that

$$\|m_2(W_2, h_P + h) - m_2(W_2, h_P) - \nabla m_2(W_2, h_P)[h]\|_{P,2} = o\big(\|s\|_\infty\{1 + \sum_{d=1}^{2} \|\lambda'_d\|_\infty\}\big) \tag{H.14}$$

since $\lambda'_{d,P}$ is uniformly continuous on $[0, 1]$ and $\|s_P - \bar{s}\|_\infty \leq \|s\|_\infty$. Thus, from (H.14) we conclude $\nabla m_2(W_2, h_P)$ is indeed the Fréchet derivative of $m_2(W_2, \cdot) : \mathbf{H} \to L^2(W_2)$. In order to show that P is locally overidentified, we note that the moment restrictions (H.9) and (H.10) posses a triangular structure. Hence we aim to apply Lemma 4.2 with $s_P \in \mathbf{H}_1 = L^\infty((V, R))$ and $(g_{0,P}, g_{1,P}, \lambda_{0,P}, \lambda_{1,P}) = h_{P,2} \in \mathbf{H}_2 = L^2(V) \times L^2(V) \times C^1([0, 1]) \times C^1([0, 1])$, for which (34) then holds since $\|\lambda'_{d,P}\|_\infty < \infty$ for $d \in \{0, 1\}$. Moreover, Corollary 4.1 implies $\bar{\mathcal{R}}_1 = L^2(W_1)$ since $L^\infty(W_1)$ is dense in $L^2(W_1)$ under $\|\cdot\|_{P,2}$. Therefore, letting $S = s_P(V, R)$ and $h_2 = (g_0, g_1, \lambda_0, \lambda_1)$ for notational simplicity, we note (H.12) and Lemma 4.2 together imply that P is locally just identified by $\mathbf{P}$ if and only if

$$\mathcal{R}_2 = \big\{f \in L^2(W_2) : f(W_2) = D\{g_1(V) + \lambda_1(S)\} + (1 - D)\{g_0(V) + \lambda_0(S)\} \text{ for } h_2 \in \mathbf{H}_2\big\} \tag{H.15}$$

[3] Technically, $\lambda_j(s(V, R))$ may not be well defined if $s(V, R) \notin [0, 1]$ since $\lambda_j \in C^1([0, 1])$. However, note $s_P(V, R) \in [0, 1]$ almost surely by (43) so for notational simplicity we ignore this issue.

is dense in $L^2(W_2)$. However, if $S = s_P(V, R)$ is not a measurable function of V (i.e. the instrument R is relevant), then $\{f \in L^2((V, S)) : f(V, S) = g(V) + \lambda(S)\}$ is not dense in $L^2((V, S))$. Hence, from (H.15) we conclude that $\bar{\mathcal{R}}_2$ is not dense in $L^2(W_2)$ and thus by Lemma 4.2 and Theorem 4.1 that P is locally overidentified. ■

Example 4.3. We study a more general version of the model introduced in the main text. In particular, we still maintain that for some U_{it} mean independent of (K_{it}, L_{it}, I_{it})

$$Y_{it} = g_P(K_{it}, L_{it}) + \omega_{it} + U_{it}. \tag{H.16}$$

However, we now let L_{it} be a possibly dynamic variable, in which case (48) becomes

$$\omega_{it} = \lambda_P(K_{it}, L_{it}, I_{it}). \tag{H.17}$$

Maintaining that ω_{it} follows an AR(1) process with coefficient π_P, and recalling that $W_i = (K_{i1}, L_{i1}, I_{i1})$ we then obtain the following two conditional moment restrictions

$$E[Y_1 - \nu_P(W)|W] = 0 \tag{H.18}$$

$$E[Y_2 - g_P(K_2, L_2) - \pi_P(\nu_P(W) - g_P(K_1, L_1))|W] = 0 \tag{H.19}$$

where $\nu_P(W) = g_P(K_1, L_1) + \lambda_P(K_1, L_1, I_1)$. Let $L^2((K_1, L_1)) = L^2((K_2, L_2))$, $h_P = (\nu_P, g_P, \pi_P)$ and the parameter space be $\mathbf{H} = L^2(W) \times L^2((K_1, L_1)) \times \mathbf{R}$. It is straightforward to verify that in this model we have for any $h = (\nu, g, \pi) \in \mathbf{H}$,

$$\nabla m_1(W, h_P)[h] = -\nu(W), \tag{H.20}$$

$$\nabla m_2(W, h_P)[h] = -E[g(K_2, L_2)|W] - \pi\lambda_P(K_1, L_1, I_1) - \pi_P(\nu(W) - g(K_1, L_1)). \tag{H.21}$$

Since the model defined by (H.18) and (H.19) has a triangular structure, we next apply Lemma 4.2 to establish that it is locally overidentified. Let $\nu_P \in \mathbf{H}_1 = L^2(W)$ and $(g_P, \pi_P) \in \mathbf{H}_2 = L^2((K_1, L_1)) \times \mathbf{R}$. Since $\pi_P < \infty$, condition (34) of Lemma 4.2 is satisfied by (H.20), (H.21), and direct calculation. Applying Corollary 4.1 to (H.20) and since $\mathbf{H}_1 = L^2(W)$ we trivially obtain that $\bar{\mathcal{R}}_1 = L^2(W)$. Therefore, by Theorem 4.1 and Lemma 4.2 we can conclude that P is locally overidentified if and only if $\bar{\mathcal{R}}_2 \neq L^2(W)$, where

$$\mathcal{R}_2 = \big\{-E[g(K_2, L_2)|W] + \pi_P g(K_1, L_1) - \pi\lambda_P(K_1, L_1, I_1) : (g, \pi) \in L^2((K_1, L_1)) \times \mathbf{R}\big\}. \tag{H.22}$$

Inspecting (H.22), a sufficient condition for P to be locally overidentified is therefore for the map $g \mapsto E[g(K_2, L_2)|W]$ to not be able to generate arbitrary functions of $W = (K_1, L_1, I_1)$. Formally, defining the spaces

$$\bar{F} \equiv \mathrm{cl}\big\{E[g(K_2, L_2)|W] - E[g(K_2, L_2)|K_1, L_1] : g \in L^2((K_1, L_1))\big\}$$

$$L^2((K_1, L_1))^{\perp} \equiv \big\{f \in L^2(W) : \int fgdP = 0 \text{ for all } g \in L^2((K_1, L_1))\big\} \tag{H.23}$$

we note $\bar{F} \subseteq L^2((K_1, L_1))^{\perp}$ by the law of iterated expectations, and therefore we may decompose $L^2((K_1, L_1))^{\perp} = \bar{F} \oplus \bar{F}^{\perp}$. A sufficient condition for P to be locally overidentified is then that the dimension of $\bar{F}^{\perp}$ is at least two. ■

We conclude by discussing an additional example based on the nonparametric analysis of changes in the wage distribution by DiNardo et al. (1996).

Example H.1. Suppose we observe $\{H_i, D_i, V_i, T_i\}_{i=1}^n$ where for each individual i, H_i denotes hourly wages, D_i is a dummy variable for union membership, V_i is a vector of covariates, and $T_i \in \{1,2\}$ indicates the time period individual i was measured in. The parameter of interest θ_P is the counterfactual τ^{th} quantile of wages that would have held in period two if unionization rates had been constant between periods, which solves

$$E\Big[\tau - 1\{H \leq \theta_P, D = 1\}\frac{\lambda_{1,P}(V)}{\lambda_{2,P}(V)} - 1\{H \leq \theta_P, D = 0\}\frac{1-\lambda_{1,P}(V)}{1-\lambda_{2,P}(V)}\Big| T = 2\Big] = 0 \tag{H.24}$$

for $\lambda_{t,P}(V)$ the unionization rate conditional on V at time t.[4] Since $\lambda_{t,P}$ satisfies

$$E[D - \lambda_{t,P}(V)|V, T = t] = 0 \text{ for } t \in \{1,2\}, \tag{H.25}$$

this setting fits model (26) with parameters $(\lambda_{1,P}, \lambda_{2,P}, \theta_P)$. Specifically, we suppose that $\lambda_{t,P} \in \mathbf{L}_t \subseteq L^{\infty}(V)$, let $\mathbf{H} = \mathbf{L}_1 \times \mathbf{L}_2 \times \mathbf{R}$, set $X = (Z, W) = (H, D, V, T)$ and define

$$\rho_1(Z,h) = 1\{T=1\}(D - \lambda_1(V)) + 1\{T=2\}(D-\lambda_2(V))$$
$$\rho_2(Z,h) = 1\{T=2\}\Big(\tau - 1\{H \leq \theta, D=1\}\frac{\lambda_1(V)}{\lambda_2(V)} - 1\{H \leq \theta, D = 0\}\frac{1-\lambda_1(V)}{1-\lambda_2(V)}\Big)$$

for any $h = (\lambda_1, \lambda_2, \theta) \in \mathbf{H}$, and where $W_1 = (V, T)$ and since the second moment restriction is *unconditional* we set $W_2 = \{1\}$. It is straightforward to verify that in this model

$$\nabla m_1(W_1, h_P)[h] = -1\{T=1\}\lambda_1(V) - 1\{T=2\}\lambda_2(V) \tag{H.26}$$

for any $h = (\lambda_1, \lambda_2, \theta) \in \mathbf{H}$. For notational simplicity we let $R = (V, T, D)$ and $G_{H|R}(h|r)$ and $g_{H|R}(h|r)$ respectively denote the cdf and density of H conditional on R. Then, by direct calculation it follows that for any $h = (\lambda_1, \lambda_2, \theta) \in \mathbf{H}$ we have

$$m_2(W_2, h) = E\Big[1\{T=2\}\Big(\tau - G_{H|R}(\theta|R)1\{D=1\}\frac{\lambda_1(V)}{\lambda_2(V)} - G_{H|R}(\theta|R)1\{D=0\}\frac{1-\lambda_1(V)}{1-\lambda_2(V)}\Big)\Big],$$

and that sufficient conditions for $m_2(W_2, \cdot) : \mathbf{H} \to \mathbf{R}$ to be Fréchet differentiable are that: (i) $g_{H|R}(H|R)$ be continuously differentiable in H with almost surely bounded level and derivative in (H, R), and (ii) $P(1 - \epsilon \geq \lambda_{2,P}(V) \geq \epsilon) = 1$ for some $\epsilon > 0$. In addition, notice that this model possess the triangular structure required in Lemma 4.2 with $\mathbf{H}_1 = \mathbf{L}_1 \times \mathbf{L}_2$ and $\mathbf{H}_2 = \mathbf{R}$, while requirement (34) holds under the additional assumption that $P(P(T = t|V) \geq \epsilon) = 1$ for $t \in \{1,2\}$

[4] I.e., as in DiNardo et al. (1996), we desire the τ^{th} quantile of $G(H|D, V, T = 2)G(D|V, T = 1)G(V|T = 2)$, where for any (A, B), $G(A|B)$ denotes the distribution of A conditional on B.

and some $\epsilon > 0$. Employing the notation of Lemma 4.2, we obtain by direct calculation that

$$\mathcal{R}_1 = \left\{ f \in L^2(W_1) : f(T,V) = \sum_{t=1}^{2} 1\{T = t\}\lambda_t(V) \text{ for } (\lambda_1, \lambda_2) \in \mathbf{L}_1 \times \mathbf{L}_2 \right\}$$

$$\mathcal{R}_2 = \left\{ \theta \times E\left[1\{T = 2\}g_{H|R}(\theta_P|R)\left(1\{D = 1\}\frac{\lambda_{1,P}(V)}{\lambda_{2,P}(V)} + 1\{D = 0\}\frac{1 - \lambda_{1,P}(V)}{1 - \lambda_{2,P}(V)}\right)\right] : \theta \in \mathbf{R} \right\}.$$

Notice, however, that the expectation defining $\mathcal{R}_2$ is necessarily positive, and thus $\mathcal{R}_2 = \mathbf{R}$ – an equality that simply reflects that assuming existence of θ_P offers no additional information. Thus, Lemma 4.2 implies that P is locally just identified if and only if $\bar{\mathcal{R}}_1 = L^2(W_1)$. Equivalently, P is locally just identified iff $\mathbf{L}_t$ is dense in $L^2(V)$ for $t \in \{1, 2\}$. For example, P is locally overidentified if we restrict $\lambda_{t,P}$ through an additive separable specification. In that case, the choice of estimator for $\lambda_{t,P}$ can affect the asymptotic distribution of a plug-in estimator for θ_P. ■

References

Ai, C. and Chen, X. (2003). Efficient estimation of models with conditional moment restrictions containing unknown functions. *Econometrica*, **71** 1795–1844.

Ai, C. and Chen, X. (2012). The semiparametric efficiency bound for models of sequential moment restrictions containing unknown functions. *Journal of Econometrics*, **170** 442–457.

Aliprantis, C. D. and Border, K. C. (2006). *Infinite Dimensional Analysis – A Hitchhiker's Guide*. Springer-Verlag, Berlin.

Andrews, D. W. K. (2017). Examples of l2-complete and boundedly-complete distributions. *Journal of Econometrics*, **199** 213–220.

Bickel, P., Klaassen, C., Ritov, Y. and Wellner, J. (1993). *Efficient and Adaptive Estimation for Semiparametric Models*. Johns Hopkins University Press.

Billingsley, P. (2008). *Probability and measure*. John Wiley & Sons.

Bogachev, V. I. (1998). *Gaussian Measures*. American Mathematical Society, Providence.

Bogachev, V. I. (2007). *Measure Theory: Volume I*. Springer-Verlag, Berlin.

Canay, I. A., Santos, A. and Shaikh, A. M. (2013). On the testability of identification in some nonparametric models with endogeneity. *Econometrica*, **81** 2535–2559.

Chen, X. and Pouzo, D. (2009). Efficient estimation of semiparametric conditional moment models with possibly nonsmooth residuals. *Journal of Econometrics*, **152** 46–60.

Chen, X. and Pouzo, D. (2012). Estimation of nonparametric conditional moment models with possibly nonsmooth generalized residuals. *Econometrica*, **80** 277–321.

DiNardo, J., Fortin, N. M. and Lemieux, T. (1996). Labor market institutions and the distribution of wages, 1973-1992: A semiparametric approach. *Econometrica*, **64** 1001–1044.

LeCam, L. (1986). *Asymptotic Methods in Statistical Decision Theory*. Springer-Verlag, New York.

Lehmann, E. and Romano, J. (2005). *Testing Statistical Hypotheses*. Springer Verlag.

Lewandowski, M., Ryznar, M. and Zak, T. (1995). Anderson inequality is strict for gaussian and stable measures. *Proceedings of the American Mathematical Society*, **123** 3875–3880.

Luenberger, D. (1969). *Optimization by Vector Space Methods*. John Wiley & Sons.

Robinson, P. M. (1988). Root-n-consistent semiparametric regression. *Econometrica*, **56** 931–954.

Stone, C. J. (1985). Additive regression and other nonparametric models. *The annals of Statistics*, **13** 689–705.

van der Vaart, A. (1991). An asymptotic representation theorem. *International Statistical Review*, **59** 97–121.

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

van der Vaart, A. and Wellner, J. (1996). *Weak Convergence and Empirical Processes.* Springer Verlag.

Williams, D. (1991). *Probability with Martingales.* Cambridge University Press, New York.

Zeidler, E. (1984). *Nonlinear Functional Analysis and its Applications III.* Springer-Verlag, New York.