Efficiency bounds for moment condition models with mixed identification strength

Prosper Dovonon[*], Yves F. Atchadé[†] and Firmin Doko Tchatoka[‡]

Concordia University, Boston University and The University of Adelaide

February 25, 2024

Abstract

Moment condition models with mixed identification strength are models that are point identified but with estimating moment functions that are allowed to drift to 0 uniformly over the parameter space. Even though identification fails in the limit, depending on how slow the moment functions vanish, consistent estimation is possible. Existing estimators such as the generalized method of moment (GMM) estimator exhibit a pattern of nonstandard or even heterogeneous rate of convergence that materializes by some parameter directions being estimated at a slower rate than others. This paper derives asymptotic semiparametric efficiency bounds for regular estimators of parameters of these models. We show that GMM estimators are regular and that the so-called two-step GMM estimator – using the inverse of estimating function's variance as weighting matrix – is semiparametrically efficient as it reaches the minimum variance attainable by regular estimators. This estimator is also asymptotically minimax efficient with respect to a large family of loss functions. Monte Carlo simulations are provided that confirm these results.

Keywords: Generalized method of moments, mixed identification strength, weak identification, efficiency bounds, semiparametric models.
JEL classification: C01, C14, C36.

[*]Economics Department, Concordia University, 1455 de Maisonneuve Blvd. West, H 1155, Montreal, Quebec, Canada, H3G 1M8. Email: prosper.dovonon@concordia.ca (Corresponding author).

[†]Department of Mathematics and Statistics, Boston University, 111 Cummington Mall, Boston, USA, MA 02215. Email: atchade@bu.edu.

[‡]School of Economics and Public Policy, The University of Adelaide, 10 Pulteney Street, Adelaide, Australia, SA 5005. Email: firmin.dokotchatoka@adelaide.edu.au.

1 Introduction

Moment equality based inference methods have made possible the investigation of the empirical content of many economic models. The validity of the standard methods popularized by Hansen (1982) in his seminal paper relies upon the property of point identification which means that the moment condition model is solved at a single point. This indeed guarantees consistent estimation by the generalized method of moment (GMM). However, some empirical evidence suggest that point identification can fail leading to poor inference.

Failure of identification occurs when multiple (or a continuum of) elements in the parameter space solve the model. While it is hard in general to decide whether identification fails by screening sample mean functions, it appears in empirical applications failing identification that the evidence for this tends to be more pronounced as the sample size gets larger. This feature has led Staiger and Stock (1997) and Stock and Wright (2000) among others – in their attempt to shed some light on the behaviour of estimators under identification failure – to consider a framework that allows the moment function to drift to zero at the rate $n^{-1/2}$ uniformly over the parameter space as the sample size n grows. This is the so-called weak identification. In this framework, point identification is possible at any given sample size[1] but in the limit the moment condition becomes uninformative about the true parameter value. They find out that consistent estimators are not available for weakly identifying models.

Hahn and Kuersteiner (2002) (in linear IV setting) and Antoine and Renault (2009, 2012) (in the general GMM context) observe that when moment conditions drift uniformly to zero at a rate $n^{-\delta} : 0 \leq \delta < 1/2$, consistent estimation is possible and they derive the asymptotic distribution of the GMM estimator in such settings. This configuration includes the standard identification framework when $\delta = 0$. We refer to Andrews and Cheng (2012), Caner (2009), Han and McCloskey (2019), among others, for further account of such models. Antoine and Renault (2012) further consider the so-called moment condition model with mixed identification strength in which the components of the estimating moment function are allowed to have specific drifting rates. They establish that the GMM estimator is consistent and, even though the rate of convergence may vary in some directions in the parameter space, by suitable rotation and scaling this estimator is asymptotically normal.

Interestingly, while the rotation and rates of convergence depend on the drifting parameters δs, they show that usual inference formulas of GMM yield valid inference without the need to know δ's, the rotation or the convergence rates. This robustness of GMM inference in models with mixed identification strength motivates a growing literature on the subject. Antoine and Renault (2020) recently propose a test for weak identification useful to detect whether a moment condition model permits consistent inference. Dovonon et al. (2023) propose moment selection methods that are consistent even if the best model is one with mixed identification strength.

This paper is concerned with efficient inference in moment condition models with mixed identifi-

[1] This paper considers a triangular array structure for the data which somewhat gives sense to the fact that identification is related to sample size. Indeed, in this case, the population distribution, say P_n, of the sample is sample size related. Therefore, identification, as a property linked to the population distribution is also associated to sample size. This would not be the case in standard frameworks where the population distribution is not sample size related.

cation strength. We derive semiparametric efficiency bounds for this class of models. One of our main contributions is that the efficiency of the commonly known two-step GMM estimator (2SGMM) in models with standard identification features carries over to models with mixed identification strength. Towards the derivation of efficiency bounds, we follow a similar approach to Dovonon and Atchadé (2020), by considering the implicit family of probability density functions - with respect to the population distribution of the observations - induced by the moment condition model. This family can be written $f_n^2(\theta, h)$ where $\theta \in \Theta$ is the initial model parameter lying in the Euclidean space $\mathbb{R}^p$ and h is an infinite dimension parameter lying in the Hilbert space $L^2(P_n)$, where P_n is the probability distribution of the sample. P_n is allowed to depend on the sample size to accommodate the possibility of drifting moment functions. We then highlight the local differentiability properties of f_n that are useful to obtain efficiency bounds.

We then follow Begun, Hall, Huang and Wellner (1983) (hereafter, BHHW), and Dovonon and Atchadé (2020) by proposing a convolution theorem for the asymptotic distribution of regular estimators of θ_0, the true parameter value of the parametric component θ of the model. Nevertheless, our framework differs from theirs in two main aspects. First, the reference Hilbert space $L^2(P_n)$ is sample size dependent. Second, the rate of convergence of the existing estimators is sharp only after rotation of the parameter space and this rate is typically not the same for all components. These key differences raise additional challenges to the derivation of efficiency bounds in our context and impose that we revisit and refine some of the standard tools.

First, while the semiparametric family induced by the moment condition models is the same as in Dovonon and Atchadé (2020) (except for the dependence on n) its tangent space does show its usual orthogonality property only in the limit as n grow. More specifically, considering a relevant sequence (θ_n, h_n) of parameter values and $f_n^2(\theta_n, h_n, \cdot)$ the associated sequence of density functions, the tangent space of $f_n(\theta_n, h_n, \cdot)$ at $f_n(\theta_0, h_0, \cdot)$ is defined - in the standard framework where $P_n = P_0$ for all n - by the set of $\alpha \in L^2(P_0)$ such that

$$\|\sqrt{n}(f_n(\theta_n, h_n) - f_n(\theta_0, h_0)) - \alpha\|_{L^2(P_0)} \to 0 \quad \text{as } n \to \infty.$$

This implies the orthogonality condition $\int \alpha f_n(\theta_0, h_0) dP_0 = 0$ which holds regardless of sample size. This property is important in the literature to derive the local asymptotic normality (LAN) property of the log-likelihood ratio (see Lemma 2.1 of BHHW). With $L^2(P_n)$ allowed to vary, we have $\int \alpha f_n(\theta_0, h_0) dP_n \neq 0$ in general. However, we show that this quantity converges to 0 and establish the LAN property under this weaker condition. The LAN property in turn has been essential to derive our convolution theorem for regular estimators.

The second main difference led us to introduce a notion of regular estimator that involves possibly many rates and a rotation of the parameter space. We argue that efficiency bounds should be associated to directions of estimation in which convergence rates are sharp for existing estimators. We then define local parameters θ_n such that $\|\Lambda_n R'(\theta_n - \theta_0) - \eta\| \to 0$ as $n \to \infty$, for some $\eta \in \mathbb{R}^p$, where Λ_n is a diagonal matrix containing the rates of convergence and R is a suitable rotation matrix. While the notion of regularity formally introduced in the paper is tied to a rotation through these sequences of parameters, we show that any estimator regular for a given rotation is also regular for any other rotation. In addition, an estimator efficient for one rotation is also efficient for any other rotation.

Our main contribution is the semiparametric efficiency bounds for regular estimators of moment condition models with mixed identification strength. These bounds are obtained via a convolution theorem that we establish. We show that GMM estimators are regular in the sense mentioned above. Moreover, any GMM estimator with weighting matrix $\hat{W}$ converging in probability to the limit of the inverse variance of the estimating function evaluated at θ_0 has its asymptotic variance that is equal to the semiparametric efficiency bound. This shows that the efficiency properties of the standard two-step GMM estimator established by Chamberlain (1987) continue to hold in models with mixed identification strength. Our findings also highlight that this estimator is asymptotically minimax optimal with respect to a large family of loss functions.

The literature on efficiency in moment condition models with non standard identification features is relatively recent. Dovonon and Atchadé (2020) deal with efficiency bounds in semiparametric models with singular score functions. Kaji (2021) introduces the notion of weak efficiency inference about the so-called weakly regular parameters. While these parameters are not consistently estimable, he proposes a Rao-Blackwellization procedure that generates estimators with reduced dispersion.

Andrews and Mikusheva (2022a, 2022b) derive large sample properties of quasi-Bayes procedures under weak identification. They propose inference methods that are asymptotically correct and more desirable – especially for weakly identified models – than many alternative methods.

Our setting differs from theirs by the fact that consistent estimation is possible. Further, while GMM estimators are not admissible for weakly identified parameters, our results support that the two-step GMM estimator is admissible in the nearly-weak models that we consider.

The rest of the paper is organized as follows. Section 2 introduces the moment condition models with mixed identification strength and provides the existing results about estimation and inference. The semiparametric model induced by the moment condition model is introduced in Section 3 which also presents the main results of the paper. Section 4 shows simulation results that illustrate the efficiency of the two-step GMM estimator in models with mixed identification strength and Section 5 concludes. Lengthy proofs are relegated to the Appendix. Throughout the paper, $\|a\| = \sqrt{a'a}$ if a is a vector or $\|a\| = \sqrt{trace(a'a)}$ if a is a matrix, and $\|a\|_{L^2(P)}$ refers to the $L^2(P)$-norm of $a \in L^2(P)$.

2 Moment models with mixed identification strength: existing results

In this section, we introduce the set-up of moment condition models with mixed identification strength along with some existing results on inference about model parameters.

Let $\{Y_{ni} : i = 1, \dots, n\}$ be a triangular array of independent and identically distributed $\mathbb{R}^d$-valued random variables with common distribution P_n and described by the population moment condition

$$\mathbb{E}_{P_n}\left(\phi(Y_{ni}, \theta_0)\right) := \int \phi(y, \theta_0) P_n(dy) = 0, \tag{1}$$

where $\phi(\cdot, \cdot)$ is a known $\mathbb{R}^k$-valued function, θ_0 is the parameter value of interest which is unknown but lies in Θ, a subset of $\mathbb{R}^p$ $(k \geq p)$. '$\mathbb{E}_{P_n}(\cdot)$' denotes expectation taken under the distribution P_n of Y_{ni}.

Consistent estimation and inference about the true parameter value θ_0 hinge on the properties of the moment function $\rho : \theta \mapsto \rho_n(\theta) := \mathbb{E}_{P_n}[\phi(Y_{ni}, \theta)]$. The moment condition model $\rho_n(\theta) = 0$ is uninformative about θ_0 if all or many elements of Θ solve the model. In this case, consistent estimation is compromised. When the moment equation is solved over Θ only by θ_0, consistent estimation becomes a possibility. This is the point identification condition which is the backbone of the GMM inference theory. In the context of triangular array that is under consideration in this paper, point identification can be expressed as:

$$\liminf_{n\to\infty} \inf_{\theta\in\Theta\setminus\mathcal{N}} \|\rho_n(\theta)\| > 0, \quad \text{for any open neighborhood } \mathcal{N} \text{ of } \theta_0. \tag{2}$$

This strong/point identification property can be restrictive in models where the moment function is local to zero over Θ, that is:

$$\mathbb{E}_{P_n}[\phi(Y_{ni}, \theta)] := \frac{\rho(\theta)}{n^\delta}, \quad \rho(\theta) \in \mathbb{R}^k, \quad \delta > 0, \tag{3}$$

with $\rho(\theta) = 0$ if and only if $\theta = \theta_0$.

In this case, assuming that $\rho(\theta)$ is bounded on Θ, the identification condition (2) fails. Especially,

$$\sup_{\theta\in\Theta} \|\rho_n(\theta)\| = O(n^{-\delta})$$

so that in the limit as n grows, the moment condition $\rho_n(\theta) = 0$ becomes uninformative about θ_0. This identification framework is labelled as weak or nearly weak by Antoine and Renault (2009).

Although under the local-to-zero property (3) the model (1) is uninformative about θ_0 in the limit, it is known that consistent estimation is possible. This depends on the possibility to estimate $\mathbb{E}_{P_n}(\phi(Y_{ni}, \theta))$ faster than the latter can vanish over the parameter set. In that respect, it is found that when $0 \leq \delta < 1/2$, consistent estimation is possible while this is ruled out when $\delta \geq 1/2$. This connection between δ and the possibility of consistent estimation justifies its consideration as identification strength of the related moment restriction. The smaller δ is, the stronger is the associated restriction.

While (3) considers that all the restrictions have the same strength, one may consider cases where each moment restriction is allowed to have its own strength leading to the following specification:

$$\mathbb{E}_{P_n}\left(\phi(Y_{ni}, \theta)\right) = \mathbb{L}_n^{-1}\rho(\theta), \tag{4}$$

with $\rho(\theta) = 0 \Leftrightarrow \theta = \theta_0$, where $\mathbb{L}_n$ is a (k, k)-diagonal matrix with j-th diagonal element equal to n^{δ_j}, $\delta_j \geq 0$, and n the sample size.

The moment condition model in (4) is referred to as a moment condition model with mixed identification strength. The restriction strengths δ_j's are typically unknown and this family of models encompasses the standard model when $\delta_j = 0$ for all j. Although $\delta < 1/2$ is, in general, essential to claim consistency, not all the δ_j's in (4) need to be smaller than 1/2 for consistency to be granted. For instance, if there is a subset of moment restrictions with related δ_j's smaller than 1/2 and such that the corresponding sub-vector of ρ, say $\rho_{\square}$, is identifying (e.g., $\rho_{\square}(\theta) = 0 \Leftrightarrow \theta = \theta_0$), then consistent estimation is possible regardless of the magnitude of the identification strength associated to the other moment restrictions. We will refer to moment restrictions with related $\delta = 0$ as being strong, those with $\delta \in]0, 1/2[$ as semi-strong and those with $\delta \geq 1/2$ as weak.

Moment condition models with mixed identification strength have been the object of study by Antoine and Renault (2009, 2012, 2020, 2021), Caner (2009), and more recently Dovonon et al. (2023). The main purpose of these studies is to propose inference methods in standard moment condition models that are robust to some forms of mixed identification strength.

This paper is concerned with efficiency bounds for the estimation of θ_0 in models with mixed identification strength. For convenience, we shall focus on a simpler model with $\mathbb{L}_n$ including only two possibly different values of δ_j so that we have the following partition of the moment function with $0 \leq \delta_1 \leq \delta_2 < 1/2$[2]:

$$\phi := (\phi_1', \phi_2')' \in \mathbb{R}^{k_1} \times \mathbb{R}^{k_2}, \ \rho := (\rho_1', \rho_2')' \in \mathbb{R}^{k_1} \times \mathbb{R}^{k_2} : \quad \mathbb{E}_{P_n}(\phi_j(Y_{ni}, \theta)) = \frac{\rho_j(\theta)}{n^{\delta_j}}, \quad j = 1, 2, \tag{5}$$

with $[\rho(\theta) = 0 \Leftrightarrow \theta = \theta_0]$.

It is worth clarifying that the moment condition model of interest is given by (1) while (5) presents some auxiliary properties typically unknown to the econometrician/practitioner about the behaviour of the moment function over the parameter set. Note that the properties in (5) include for $\delta_1 = \delta_2 = 0$, the standard framework where the model is point identified and the moment function does not drift to 0 uniformly over Θ. In (5), since $\mathbb{E}_{P_n}(\phi_1(Y_{ni}, \theta))$ vanishes on Θ more slowly than $\mathbb{E}_{P_n}(\phi_2(Y_{ni}, \theta))$, the former defines the strongest set of moment restrictions if $\delta_1 < \delta_2$.

Examples of moment condition models with mixed identification strength are presented in Dovonon et al. (2023), Antoine and Renault (2012), and Han and McCloskey (2019). We present below the linear IV model with nearly weak instruments which also is object of simulation in Section 4.

Example 1. (***Linear IV Model with Nearly Weak Instruments***). *This example relates to linear regression models with endogenous regressors for which available instrumental variables are possibly weak. Moreover, the set of instruments may be partitioned in two groups, each with a specific magnitude of partial correlation with the endogenous regressor(s). As we can see below, such setting leads to a moment condition model with identification property as in* (5).

Specifically, consider the random sample: $\{w_i := (y_i, x_i, z_i) \in \mathbb{R} \times \mathbb{R}^p \times \mathbb{R}^k : i = 1, \dots, n\}$. *Assume that:*

$$y_i = x_i'\theta_0 + u_i, \tag{6}$$

$$x_i = \Pi_{1n} z_{1i} + \Pi_{2n} z_{2i} + v_i, \tag{7}$$

$$\text{with} \ : \mathbb{E}(z_i u_i) = 0, \ \mathbb{E}(z_i v_i) = 0, \ \text{and, for each } n, Rank(\mathbb{E}(z_i x_i')) = p, \tag{8}$$

where, for $j = 1, 2$, $\Pi_{jn} = n^{-\delta_j} C_j$; $\quad C_j \in \mathbb{R}^p \times \mathbb{R}^{k_j}$; $0 \leq \delta_1 \leq \delta_2$; $z_i = (z_{1i}', z_{2i}')'$; *and* $k_1 + k_2 = k$.

[2]The main results derived in this paper stay valid in the more general cases where $\mathbb{L}_n$ features more than 2 identification strengths. They are also valid in cases where the model includes weak and/or uninformative restrictions ($\delta_j \geq 1/2$), so long as enough strong and/or semi-strong restrictions are included to ensure consistent and asymptotically normal estimation.

In this representation, δ_j captures the strength of the instruments z_j through the magnitude of its partial correlation with the endogenous variables. Clearly, θ_0 solves the moment restriction:

$$\mathbb{E}(z_i(y_i - x_i'\theta)) = 0. \tag{9}$$

Furthermore, assuming - to simplify - that the sets of instruments z_{1i} and z_{2i} are orthogonal and letting:

$$\Delta_{11} := \mathbb{E}(z_{1i}z_{1i}'); \quad \Delta_{22} := \mathbb{E}(z_{2i}z_{2i}'); \quad \rho_1(\theta) := \Delta_{11}C_1'(\theta_0 - \theta); \quad \textit{and} \quad \rho_2(\theta) := \Delta_{22}C_2'(\theta_0 - \theta),$$

we have:

$$\mathbb{E}(z_i(y_i - x_i'\theta)) := \begin{pmatrix} \mathbb{E}(z_{1i}(y_i - x_i'\theta)) \\ \mathbb{E}(z_{2i}(y_i - x_i'\theta)) \end{pmatrix} = \begin{pmatrix} n^{-\delta_1}\rho_1(\theta) \\ n^{-\delta_2}\rho_2(\theta) \end{pmatrix}.$$

This shows that the linear IV model in (6)-(8) *yields a moment condition model with mixed identification strength. Thanks to the rank condition in this model specification, we can also verify that*

$$\rho(\theta) := (\rho_1(\theta)', \rho_2(\theta)')' = 0 \Leftrightarrow \theta = \theta_0. \quad \square$$

Example 2. (***Optimal Prediction***). *This example focuses on nonlinear prediction functions. Specifically, consider the random sample $\{w_i := (y_i, x_i, z_i) \in \mathbb{R}^3 : i = 1, \dots, n\}$, where x_i is independent of z_i for all i. Our interest lies in determining the optimal projection of y_i onto z_i and $h_n(\gamma, x_i)$, where $h_n(\cdot)$ is a function depending on the sample size n and known up to some parameter $\gamma \in [0,1]$. Let $\hat{\alpha}$ and $\hat{\gamma}$ be determined such that $\hat{y}_i = \hat{\alpha}z_i + h_n(\hat{\gamma}, x_i)$ minimizes $\frac{1}{n}\sum_{i=1}^n (y_i - \hat{y}_i)^2$.*

The process of finding these values is equivalent to minimizing $\frac{1}{n}\sum_{i=1}^n (y_i - \alpha z_i - h_n(\gamma, x_i))^2$ with respect to α and γ. The first-order condition of this minimization problem implies the following moment conditions:

$$\mathbb{E}[z_i(y_i - \alpha z_i - h_n(\gamma, x_i))] = 0, \tag{10}$$

$$\mathbb{E}[h_{n\gamma}(\gamma, x_i)(y_i - \alpha z_i - h_n(\gamma, x_i))] = 0, \tag{11}$$

where $h_{n\gamma}(\gamma, x_i) = \partial h_n(\gamma, x_i)/\partial\gamma$.

Suppose that $x_i \sim \mathcal{U}[-\pi, \pi]$, with $\pi = 4\arctan(1)$, $z_i \sim N(0, \sigma^2)$ independently with x_i, $\mathbb{E}(y_i|x_i) = cx_i$ for some constant $c \neq 0$, and $h_n(\gamma, x_i) = n^{-\delta}[\sin(\gamma x_i) - \cos(\gamma x_i)]$ for some $\delta \in [0, 1/2)$. Let

$$\begin{aligned} \phi_1(w_i, \alpha, \gamma) &= z_i\big(y_i - \alpha z_i - n^{-\delta}[\sin(\gamma x_i) - \cos(\gamma x_i)]\big) \\ \phi_2(w_i, \alpha, \gamma) &= n^{-\delta}x_i[\sin(\gamma x_i) + \cos(\gamma x_i)]\big(y_i - \alpha z_i - n^{-\delta}[\sin(\gamma x_i) - \cos(\gamma x_i)]\big). \end{aligned} \tag{12}$$

Using the independence between z_i and x_i, it is straightforward to see that

$$\mathbb{E}[\phi_1(w_i, \alpha, \gamma)] = \mathbb{E}[z_i(y_i - \alpha z_i)] \quad \textit{and} \quad \mathbb{E}[\phi_2(w_i, \alpha, \gamma)] = n^{-\delta}\mathbb{E}\big[x_i y_i[\sin(\gamma x_i) + \cos(\gamma x_i)]\big].$$

Therefore, by setting $\phi(w_i, \alpha, \gamma) = \big[\phi_1(w_i, \alpha, \gamma), \phi_2(w_i, \alpha, \gamma)\big]'$, the moment conditions (10)-(11) *can be written as:*

$$\begin{aligned} \mathbb{E}[\phi(w_i, \alpha, \gamma)] &= \mathbb{L}_n^{-1}\rho(\alpha, \gamma), \ \ \mathbb{L}_n = \begin{pmatrix} 1 & 0 \\ 0 & n^\delta \end{pmatrix}, \ 0 \leq \delta < \frac{1}{2}, \quad \textit{and} \\ \rho(\alpha, \gamma) &= \big(\mathbb{E}[z_i(y_i - \alpha z_i)], \ c\mathbb{E}\big[x_i^2[\sin(\gamma x_i) + \cos(\gamma x_i)]\big]\big)'. \end{aligned}$$

Under the above assumptions, we can verify that

$$\rho(\alpha,\gamma)=0 \Leftrightarrow \alpha=\alpha_0=\sigma^{-2}\mathbb{E}(z_i y_i),\ \gamma=\gamma_0\approx 0.8296. \quad \square$$

We now review the existing results on inference about the model parameter θ_0. We emphasize those that are useful to us in the next section on the derivation of efficiency bounds. Let the GMM estimator $\hat{\theta}_n$ be defined by

$$\hat{\theta}_n=\arg\min_{\theta\in\Theta}\bar{\phi}_n(\theta)'W_n\bar{\phi}_n(\theta), \tag{13}$$

where $\bar{\phi}_n(\theta):=n^{-1}\sum_{i=1}^n\phi(Y_{ni},\theta)$ and W_n is a sequence of almost surely symmetric positive definite matrices converging in probability to W, a symmetric positive definite matrix.

Consistency of $\hat{\theta}_n$ for θ_0 is ensured under Assumption A.1 in Appendix A while Assumptions A.1, A.2, and A.3 present sufficient conditions for the asymptotic normality of this estimator. The asymptotic normality of $\hat{\theta}_n$ is established by Antoine and Renault (2009, 2012) under the condition that the Jacobian of $\rho(\theta)$ at θ_0 is full column rank. The rate of convergence of $\hat{\theta}_n$ depends on how fast the strongest moment function $\mathbb{E}_{P_n}(\phi_1(Y_{ni},\theta))$ vanishes and on the rank s_1 of the Jacobian matrix of $\rho_1(\theta)$ at θ_0. If this rank is smaller than p, the dimension of θ_0, then the remaining moment restrictions determine the rate of convergence of the $s_2:=p-s_1$ remaining directions of the parameter. To introduce this asymptotic distribution, we rely on the following notation.

We let $s_1=\text{Rank}\left(\frac{\partial\rho_1}{\partial\theta'}(\theta_0)\right)$.

- If $0<s_1<p$, define $R=(R_1\vdots R_2)$ a (p,p)-matrix such that $RR'=I_p$ and R_2 is a $(p,p-s_1)$-matrix with columns spanning the null space of $\frac{\partial\rho_1}{\partial\theta'}(\theta_0)$ and define:

$$J=\begin{pmatrix}\frac{\partial\rho_1}{\partial\theta'}(\theta_0)R_1 & 0\\ 0 & \frac{\partial\rho_2}{\partial\theta'}(\theta_0)R_2\end{pmatrix}\quad\text{and}\quad \Lambda_n=\begin{pmatrix}n^{\frac{1}{2}-\delta_1}I_{s_1} & 0\\ 0 & n^{\frac{1}{2}-\delta_2}I_{s_2}\end{pmatrix}. \tag{14}$$

- If $s_1=p$, set

$$J=\left(\frac{\partial\rho_1'}{\partial\theta}(\theta_0)\vdots 0\right)',\quad \Lambda_n=n^{\frac{1}{2}-\delta_1}I_p,\quad\text{and}\quad R=I_p.$$

- If $s_1=0$, set

$$J=\left(0\vdots\frac{\partial\rho_2'}{\partial\theta}(\theta_0)\right)',\quad \Lambda_n=n^{\frac{1}{2}-\delta_2}I_p,\quad\text{and}\quad R=I_p.$$

- Finally, if $\delta_1=\delta_2=\delta$, set

$$J=\frac{\partial\rho(\theta_0)}{\partial\theta'},\quad \Lambda_n=n^{\frac{1}{2}-\delta}I_p,\quad\text{and}\quad R=I_p.$$

Under Assumptions A.1, A.2 and A.3 in Appendix A, we can claim, following Antoine and Renault (2009, 2012) that, under P_n,

$$\Lambda_nR^{-1}(\hat{\theta}_n-\theta_0)\xrightarrow{d}N(0,\Omega(W)),\quad\text{with}\quad \Omega(W):=(J'WJ)^{-1}J'W\Sigma WJ(J'WJ)^{-1}, \tag{15}$$

where Σ is the asymptotic variance of $\sqrt{n}\bar{\phi}_n(\theta_0)$, under P_n.

As is standard in GMM theory, the asymptotic distribution of the GMM estimator depends on the probability limit W of the weighting matrix. Antoine and Renault (2009) show that the asymptotic variance $\Omega(W)$ is minimal for the choice $W = \Sigma^{-1}$. They show (see p.S151) how feasible estimators with asymptotic variance $\Omega(\Sigma^{-1}) = (J'\Sigma^{-1}J)^{-1}$ can be obtained. Interestingly, the proposed procedure is the same as that of the two-step GMM estimator in standard models. They also show that standard formulas for inference based on the two-step GMM are valid in the context of moment condition models with mixed identification strength. This highlights some robustness of the two-step GMM inference procedure to the identification pattern in (5) under the conditions in Assumptions A.1, A.2 and A.3. We shall reiterate that there is no need to know s_1, R, nor the rates of convergence in Λ_n to build asymptotically valid inference about θ_0 using the two-step GMM estimator.

In the next section, we derive asymptotic semiparametric efficiency bounds for the estimation of θ_0 in the moment condition model (1) under the mixed identification strength property in (5). We show in substance that the minimum variance $\Omega(\Sigma^{-1})$ corresponds to the semiparametric efficiency variance-bound for estimators that are regular in a sense that we will make precise.

3 Efficiency bounds

This section derives the asymptotic efficiency bound for the estimation of θ_0 in the moment condition model (1) characterized by the mixed identification strength property in (5). For this purpose, we rely on the technique introduced by Dovonon and Atchadé (2020). Their approach consists in: obtaining the semiparametric family implicitly induced by (1) in the form $\{f^2(\theta, h, \cdot) : (\theta, h) \in \mathcal{V}\}$, where $f^2(\theta, h, \cdot)$ is the probability density function of Y with respect to a reference measure, and indexed by θ in Θ and h lying in a Hilbert space. This semiparametric model is then used to obtain an efficiency bound in the direction of θ by relying on a similar approach to BHHW (1983).

There are two main differences between their set-up and the models of interest in this paper. First, the population distribution P_n of the data is allowed to depend on the sample size n and, second, common estimators of θ_0 display a mixture of rates of convergence and eliciting the directions of sharp rate requires some rotation of the parameter space.

Adapting the existing methods to derive efficiency bounds to this configuration proves to be challenging. Under the triangular arrays framework implied by the sample dependence of P_n, the induced family of densities also depends on n. We propose an extension of the notion of tangent space and refine the local asymptotic normality theory used by BHHW to accommodate such families of semiparametric models. We also propose an adaptation of the notion of regular estimators to accommodate our setting where sharp rates are up to a rotation of parameter space.

3.1 (Semi)parametric representation of moment condition models

Consider again the row-wise independent and identically distributed triangular array $\{Y_{n1}, \dots, Y_{nn}\}$ of $\mathbb{R}^d$-valued random vectors and common distribution P_n. Let $L^2(P_n)$ denote $L^2(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), P_n)$, a Hilbert space of real-valued functions on $\mathbb{R}^d$. Following Dovonon and Atchadé (2020), we next characterize the semiparametric family induced by the moment condition (1) in the form of density functions with respect to P_n. This allows to handle random variables with finite, discrete or continuous support in a unified manner. Our approach contrasts with Chamberlain (1987) who mainly considers random variables with finite support and provides extensions to continuous variables through an approximation theory. The main difference between the current set-up and Dovonon and Atchadé (2020) is that the reference measure P_n in the former depends on n to accommodate triangular arrays, while it is fixed in the latter.

We let $\nabla_\theta^{(j)}\phi(y,\theta)$ denote the j-th order differential of the map $\theta \mapsto \phi(y,\theta)$ evaluated at θ with the convention that $\nabla_\theta^{(0)}\phi(y,\theta) = \phi(y,\theta)$ and we make the following assumption.

Assumption 1.

(i) There exists a neighbourhood Θ of θ_0, a $L^2(P_n)$-neighbourhood $\mathcal{N}$ of $f_{n,0} \equiv 1$, and a finite constant $C > 0$, such that for P_n-almost all $y \in \mathbb{R}^d$, $\theta \mapsto \phi(y,\theta)$ is r-times continuously differentiable on Θ and, for all $f \in \mathcal{N}$,

$$\int \sup_{\theta\in\Theta} \left\| \nabla_\theta^{(j)}\phi(y,\theta) \right\| f^2(y) P_n(dy) \le C,$$

for $j = 0, \dots, r$.

(ii) The matrix $\Sigma_n = \int \phi(y,\theta_0)\phi(y,\theta_0)' P_n(dy)$ is positive definite.

Assumption 1 imposes some uniform dominance condition on $\nabla_\theta^{(j)}\phi(y,\theta)$ to ensure that this function is well-behaved. Note also that, when Y_{ni} is distributed as P_n, $f_{n,0}(y) = 1$ is the density of Y_{ni} with respect to P_n. This assumption imposes, in particular, that the relevant functions are integrable with respect to any density function in a certain neighbourhood of $f_{n,0}$. The second part of the assumption is quite standard.

Towards the introduction of the implicit model, further notation is needed. We equip $L^2(P_n)$ with the inner product $\langle u, v\rangle = \int u(y)v(y)P_n(dy) := \mathbb{E}_{P_n}(u(Y)v(Y))$. More generally, for $u : \mathbb{R}^d \to \mathbb{R}^{s\times r}$ and $v : \mathbb{R}^d \to \mathbb{R}^{q\times r}$, $\langle u, v\rangle = \mathbb{E}_{P_n}(u(Y)v(Y)')$, where the expectation of any matrix is understood to be component-wise.

Let $\varphi(y) = (\varphi_1(y), \dots, \varphi_k(y))' = \Sigma_n^{-1/2}\phi(y,\theta_0)$, and $\varphi_{k+1}(y) = 1$. For all $\theta \in \Theta$, let $\varphi_\theta(y) = \Sigma_n^{-1/2}\phi(y,\theta)$. Further, let $\bar\varphi = (1,\varphi')' = (\varphi_{k+1},\varphi')'$ and $\bar\varphi_\theta = (1,\varphi_\theta')' = (\varphi_{k+1},\varphi_\theta')'$. Thanks to the moment condition (1), the elements of $\bar\varphi$ are orthonormal elements of $L^2(P_n)$. By separability of $L^2(P_n)$, $\bar\varphi$ can be extended to have an orthonormal basis $\{\varphi_j : j \ge 1\}$ of $L^2(P_n)$ and let $\mathcal{E}$ denote the closed span of the subspace $L^2(P_n)$ generated by $\{\varphi_j : j \ge k+2\}$. Note that the elements of the basis $\{\varphi_j : j \ge 1\}$ ultimately depend on n but we do not stress this in the notation for simplicity.

We introduce the map $\mathcal{M}$ defined on $\Theta \times \mathcal{E} \times L^2(P_n)$ taking values in $L^2(P_n)$ such that for any $(\theta, h, f) \in \Theta \times \mathcal{E} \times L^2(P_n)$,

$$\mathcal{M}(\theta,h,f) := \frac{1}{2}\left\langle f^2, \varphi_\theta\right\rangle \varphi + \frac{1}{2}\left(\int f^2(y)P_n(dy) - 1\right)\varphi_{k+1} + \sum_{j=k+2}^{\infty}\langle \varphi_j, f-h\rangle \varphi_j. \tag{16}$$

By construction, the set of solutions of the equation $\mathcal{M}(\theta, h, f) = 0$ collects all the combinations $(\theta, f) \in \Theta \times L^2(P_n)$ consistent with the moment condition model. That is, all (θ, f) such that $\int \phi(y,\theta) f^2(y) P_n(dy) = 0$. To see this, note that, for any (θ, h, f), $\mathcal{M}(\theta, h, f) = 0$ if and only if

$$\int \phi(y,\theta) f^2(y) P_n(dy) = 0, \quad \int f^2(y) P_n(dy) = 1, \quad \text{and} \quad \langle \varphi_j, f - h \rangle = 0, \quad \forall j \geq k+2.$$

This means that the triplets (θ, h, f) that set $\mathcal{M}$ to zero are those in which: f^2 is a density function with respect to P_n; θ is a solution to the moment condition model with Y having f^2 as density function with respect to P_n; and h is the projection of f on the directions $\{\varphi_j : j \geq k+2\}$ of the considered basis.

Conversely, if $(\theta, f) \in \Theta \times L^2(P_n)$ is such that $f^2(y)$ is a density function with respect to P_n and $\int \phi(y,\theta) f^2(y) P_n(dy) = 0$, then $\mathcal{M}(\theta, \text{proj}_{\mathcal{E}}(f), f) = 0$, where $\text{proj}_{\mathcal{E}}$ is the orthogonal projection operator on the subspace $\mathcal{E}$.

Letting $h_0 = 0_{\mathcal{E}}$, we have $\mathcal{M}(\theta_0, h_0, f_{n,0}) = 0$. Lemma 2.1 of Dovonon and Atchadé (2020) shows that under Assumption 1, $\mathcal{M}$ is r-times continuously differentiable and for any $g \in L^2(P_n)$,

$$\nabla_f \mathcal{M}(\theta_0, h_0, f_{n,0}) \cdot g = \langle g, \bar{\varphi} \rangle \, \bar{\varphi} + \sum_{j \geq q+2} \langle \varphi_j, g \rangle \varphi_j = g.$$

It follows that $\nabla_f \mathcal{M}(\theta_0, h_0, f_{n,0})$ is an isomorphism of $L^2(P_n)$ and the implicit function theorem allows us to claim that there exists a neighbourhood $\mathcal{V}$ of (θ_0, h_0), a neighbourhood $\mathcal{U}$ of $f_{n,0}$ and a r-times continuously differentiable function f_n: $\mathcal{V} \to \mathcal{U}$ such that $f_n(\theta_0, h_0) = f_{n,0}$ and for all $(\theta, h) \in \mathcal{V}$,

$$\mathcal{M}(\theta, h, f_n(\theta, h)) = 0.$$

The family of functions $\{f_n(\theta, h, \cdot) : (\theta, h) \in \mathcal{V}\}$ defines the semiparametric model induced by the moment condition (1). This family is further characterized by Proposition B.1 in Appendix B which follows readily from Lemma 2.2 of Dovonon and Atchadé (2020).

3.2 Efficiency bounds for the (semi)parametric representation

To obtain semiparametric efficiency bounds for the estimation of θ_0 in model (1), we focus on the family of semiparametric density functions $\{f_n(\theta, h, \cdot) : (\theta, h) \in \mathcal{V}\}$ induced by the moment condition model as established by Proposition B.1. Our goal from this point consists in obtaining a bound for the parametric component θ in this induced semiparametric model and then show that this bound is sharp.

Of interest to us is the approach of BHHW (1983) to derive efficiency bounds for parameters of semiparametric models represented by a family of density functions depending on both a finite and an infinite dimension parameters. This approach consists in collecting all the elements $\alpha \in L^2(\mu)$ – where μ is a dominating measure with respect to which the family of model densities are expressed – and all the sequences θ_n, h_n converging to θ_0 and h_0 such that:

$$\|\sqrt{n}(f_n(\theta_n, h_n) - f_{n,0}) - \alpha\|_{L^2(\mu)} \longrightarrow 0, \text{ as } n \to \infty.$$

As maintained in their paper, if the function f_n does not vary with n, such α's necessarily belong to the tangent set of $f_n(\theta_n, h_n)$ at (θ_0, h_0) and therefore, satisfy $\int \alpha f_n d\mu = 0$. This property is used to establish that the log-likelihood ratio of the sample under the distributions $f_n(\theta_n, h_n)$ and $f_{n,0}$ is asymptotically normal. In turn, the local asymptotic normality (LAN) property of the log-likelihood ratio is used to derive efficiency bounds for a class of regular estimators.

Although our semiparametric model of interest fits with that analysed by BHHW, a key difference resides in the fact that our model involves density functions with respect to a dominating measure P_n that varies with the sample size. We first re-examine the result of BHHW in light of this difference and we propose a refined version of the LAN property established by their Lemma 2.1 that accommodates our settings.

We propose the extension in a context more general than needed for us by considering sequences of sigma-finite measures instead of probability measures. Let $(\mathfrak{X}, \mathscr{C})$ be a measurable space and $\mu_n, n \geq 0$ a sequence of sigma-finite measures on $(\mathfrak{X}, \mathscr{C})$. Let f_n^2, $n \geq 0$ and g_n^2, $n \geq 0$ be two sequences of density functions on $\mathfrak{X}$ with respect to μ_n. Let $L^2(\mu_n)$ denote $L^2(\mathfrak{X}, \mathscr{C}, \mu_n)$. By definition, $f_n, g_n \in L^2(\mu_n)$ and $\|f_n\|_{\mu_n} = 1$ and $\|g_n\|_{\mu_n} = 1$; where $\|h\|_\mu^2 = \int h^2 d\mu$.

Let $X_{n1}, \dots, X_{nn}$ be a row-wise independent and identically distributed triangular array of $\mathfrak{X}$-valued random variables. Define the likelihood ratio L_n by:

$$L_n = \log \left\{ \prod_{i=1}^{n} g_n^2(X_{ni}) \Big/ \prod_{i=1}^{n} f_n^2(X_{ni}) \right\}. \tag{17}$$

We have the following result:

Theorem 3.1. ***(Local asymptotic normality.)*** *If g_n and f_n defined above are such that, for $\alpha_n \in L^2(\mu_n)$,*

$$\|\sqrt{n}(g_n - f_n) - \alpha_n\|_{\mu_n} \to 0, \quad \text{as} \quad n \to \infty, \tag{18}$$

then:

(i) $\nu_n := \int f_n \alpha_n d\mu_n \to 0$ *as* $n \to \infty$.

(ii) If in addition, $\|\alpha_n\|_{\mu_n}^2 \to a^2 < \infty$ *as* $n \to \infty$, *then, for every* $\epsilon > 0$,

$$P_{f_n}\left(\left| L_n - 2n^{-1/2} \sum_{i=1}^{n} [\alpha_n(X_{ni})/f_n(X_{ni}) - \nu_n] + \sigma^2/2 \right| > \epsilon \right) \to 0$$

as $n \to \infty$, *where, for any μ-measurable set A,* $P_f(A) = \int_A f^2 d\mu$, *and* $\sigma^2 = 4a^2$. *Furthermore, under* P_{f_n},

$$L_n \xrightarrow{d} N(-\sigma^2/2, \sigma^2)$$

as $n \to \infty$ *and the sequences* $\left\{\prod_{i=1}^n g_n^2(x_i)\right\}$ *and* $\left\{\prod_{i=1}^n f_n^2(x_i)\right\}$ *are contiguous.*

Proof: See Appendix. □

This result shows that in our context, if $\alpha_n \in L^2(P_n)$ is such that

$$\|\sqrt{n}(f_n(\theta_n, h_n) - f_{n,0}) - \alpha_n\|_{L^2(P_n)} \to 0, \tag{19}$$

we do not necessarily have $\int \alpha_n f_{n,0} dP_n = 0$ but instead

$$\lim_{n\to\infty} \int \alpha_n f_{n,0} dP_n = 0$$

and the LAN property in Theorem 3.1(ii) can be obtained from this *asymptotic* form of tangent space. We shall rely on this refinement to establish the main results in this paper.

Perhaps, at this point, it is worth addressing the fact that, for the same sequence $(f_n(\theta_n, h_n), f_{n,0})$, many sequences α_n of elements of $L^2(P_n)$ may satisfy (19). We observe, thanks to the triangle inequality, that any pair of sequences $\alpha_{1,n}$ and $\alpha_{2,n}$ that satisfy (19) are such that

$$\begin{aligned}\left|\|\alpha_{1,n}\|_{L^2(P_n)} - \|\alpha_{2,n}\|_{L^2(P_n)}\right| &\leq \|\alpha_{1,n} - \alpha_{2,n}\|_{L^2(P_n)} \\ &\leq \|\sqrt{n}(f_n(\theta_n, h_n) - f_{n,0}) - \alpha_{1,n}\|_{L^2(P_n)} + \|\sqrt{n}(f_n(\theta_n, h_n) - f_{n,0}) - \alpha_{2,n}\|_{L^2(P_n)} \to 0.\end{aligned}$$

As a result, $\|\alpha_{1,n}\|_{L^2(P_n)}$ and $\|\alpha_{2,n}\|_{L^2(P_n)}$ have the same limit inferior and the same limit superior. This property is of particular interest since α_n is related to the local asymptotic normal distribution in Theorem 3.1 only through the limit of its $L^2(P_n)$-norm if such a limit exists. Clearly, the existence of the limit for one solution of (19) implies that any other solution has the same limit. The practical consequence of this is that we can focus on any solution of (19) to develop our asymptotic efficiency theory.

Characterization of the *asymptotic* tangent space. Let us now give a more specific sense to $g_n(\cdot) := f_n(\theta_n, h_n, \cdot)$ by determining the set of all sequences of $\{(\theta_n, h_n)\}_n$ of interest and the associated α_n that guarantee (19). For this, we need to make a choice about the rate of convergence of (θ_n, h_n) to (θ_0, h_0). If all the components of θ_0 were estimable at the same rate, r_n, the standard approach consists in using that rate to characterize the local parameters (θ_n, h_n). This is the case in the theory of BHHW where $r_n = \sqrt{n}$. However, the asymptotic distribution in (15) shows that, except for the extreme cases of $s_1 = 0$ and $s_1 = p$, standard estimators of θ_0 do not converge at the same rate in all directions.

If we were to determine the local sequences θ_n based directly on the rate of convergence of the GMM estimator, it appears that information related to the directions estimable at a faster rate would be lost and, thereby compromising efficiency. The results in Section 2 on GMM estimation provide an intuition about this claim. The rate of convergence of this estimator is $n^{1/2-\delta_2}$ which is related to the directions estimable at the slowest rate. From (15), we can claim that, under P_n,

$$n^{1/2-\delta_2}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N\left(0, R_2\Omega(W)_{22}R_2'\right),$$

where $\Omega(W)_{22}$ is the lower-right (s_2, s_2)-sub-matrix of $\Omega(W)$. This is a degenerated Gaussian limit that accounts only for a subset of estimation directions by omitting the faster ones.

Because of this, it makes more sense to focus on the rotation of the parameter that disentangle the estimation directions with sharp rates. As established by (15), the first s_1 components of $\nu_0 = R^{-1}\theta_0$ are estimable at rate $n^{\frac{1}{2}-\delta_1}$ while the remaining s_2 are at rate $n^{\frac{1}{2}-\delta_2}$ and those rates are sharp. We shall consider this fact and explore sequences (θ_n) such that:

$$\Lambda_n R^{-1}(\theta_n - \theta_0) - \eta \to 0, \tag{20}$$

as $n \to \infty$ for some $\eta \in \mathbb{R}^p$, and R a rotation matrix satisfying the definition in (14).

Effectively, efficient bounds for θ_0 are explored in the case $0 < s_1 < p$ through its linear transformation $\nu_0 = R^{-1}\theta_0$. We will say that an estimator $\tilde{\theta}$ of θ_0 is asymptotically efficient if there is a rotation R as defined in (14) such that $R^{-1}\tilde{\theta}$ is an asymptotically efficient estimator of $\nu_{0,R} := R^{-1}\theta_0$. We shall see that if $\tilde{\theta}$ is asymptotically efficient for a specific rotation, it is also asymptotically efficient for any other rotation consistent with that definition.

Remark 1. *It is worth mentioning that the set of sequences* (θ_n) *determined by* (20) *is the same regardless of the choice of rotation matrix* R*. Indeed, as shown by Lemma B.1, any other rotation* $\mathbf{R}$ *consistent with the definition* (14) *satisfies* $\mathbf{R} = RA$*, with* A *a nonsingular block diagonal matrix. It follows that, if* (θ_n) *satisfies* (20) *with the rotation matrix* R*, it also satisfies* (20) *with the rotation matrix* $\mathbf{R}$ *and* η *replaced by* $A'\eta$*.*

This remark shows that the choice of rotation matrix is immaterial in the collection of local sequences (θ_n) given by (20). We reiterate that the discussion on rotation is relevant only in models where $0 < s_1 < p$. Note that the relevant sequences (θ_n) are such that for some $\eta \in \mathbb{R}^p$,

$$\text{if } s_1 = p,\ \Lambda_n R^{-1}(\theta_n - \theta_0) - \eta = n^{1/2-\delta_1}(\theta_n - \theta_0) - \eta \to 0, \quad \text{and}$$
$$\text{if } s_1 = 0,\ \Lambda_n R^{-1}(\theta_n - \theta_0) - \eta = n^{1/2-\delta_2}(\theta_n - \theta_0) - \eta \to 0 \quad (21)$$

so that no rotation is explicitly involved. Our aim is to derive an efficiency bound for the estimation of θ_0 that is valid whether $s_1 = 0$, p or $0 < s_1 < p$. For this reason, we will consider sequences defined by (20) with the understanding that this definition collapses to (21) in the extreme cases.

Regarding the non-parametric component of the model, we consider (h_n) such that

$$\|\sqrt{n}(h_n - h_0) - \beta\|_{L^2(P_n)} \to 0 \quad (22)$$

as $n \to \infty$, for some $\beta \in L^2(P_n)$. The parametric rate in the definition of (h_n) may seem arbitrary but the consequence of this choice is that the set of sequences (h_n) thus defined is small and may lead to irrelevant bounds. We shall see later that this set is actually the right one as the resulting bound will be proved sharp.

Following similar lines to BHHW, we collect all these sequences in specific sets by letting $\Theta(\theta_0, \eta)$ denote the set of all sequences (θ_n) satisfying (20) and $\Theta(\theta_0) = \bigcup_{\eta \in \mathbb{R}^p} \Theta(\theta_0, \eta)$. Similarly, $\mathcal{C}(h_0, \beta)$ denotes the collection of all sequences (h_n) such that (22) holds and $\mathcal{C}(h_0) = \bigcup_{\beta \in \mathcal{B}(h_0)} \mathcal{C}(h_0, \beta)$, where

$$\mathcal{B}(h_0) = \{\beta \in \mathcal{E} \text{ such that (22) holds for some sequence } (h_n) \text{ of elements of } \mathcal{E}\}.$$

The sequences of experiments that we shall consider are:

$$g_n^2(\cdot) := f_n^2(\theta_n, h_n, \cdot), \quad \text{with} \quad \{(\theta_n), (h_n)\} \in \Theta(\theta_0) \times \mathcal{C}(h_0). \quad (23)$$

From Proposition B.1, $(\theta, h) \mapsto f_n(\theta, h)$ is twice continuously Fréchet differentiable and this is sufficient to claim that $f_n^2(\theta, h)$ is Hellinger differentiable at (θ_0, h_0). In this case, the limit elements α_n in (18) are characterized by the Fréchet (or Hellinger) derivatives of $f_n(\theta, h)$ at (θ_0, h_0). Indeed, by the Taylor's formula, there exists a function $r_{n,\theta_0} \in L^2(P_n)$ and a bounded linear operator $A_n : L^2(P_n) \to L^2(P_n)$ such that:

$$\|g_n - f_{n,0} - r_{n,\theta_0} \cdot (\theta_n - \theta_0) - A_n \cdot (h_n - h_0)\|_{L^2(P_n)} = \|\mathbf{r}_2(\theta_n, h_n, \cdot)\|_{L^2(P_n)}, \tag{24}$$

where $\mathbf{r}_2(\theta_n, h_n, \cdot)$ is the Lagrange remainder. If

$$\sqrt{n}\|\mathbf{r}_2(\theta_n, h_n, \cdot)\|_{L^2(P_n)} = o(1) \tag{25}$$

then the leading part of $r_{n,\theta_0} \cdot (\theta_n - \theta_0) + A_n \cdot (h_n - h_0)$ in $L^2(P_n)$ would represent α_n associated to the sequence (θ_n, h_n) as in (18).

Lemma B.2 in Appendix B establishes (25) under the condition that

$$(\delta_1, \delta_2) \in \Delta := \Big\{(a, b) \in [0, 1/2[^2 \colon\ 0 \le a \le b < [(1/4 + a/2) \wedge 3/8]\Big\}.$$

This result is obtained by showing that $\sqrt{n}\|\mathbf{r}_2(\theta_n, h_n, \cdot)\|_{L^2(P_n)} = O(n^{-1/2+2\delta_2})$ if $\delta_2 < 1/4$ and if $\delta_2 \ge 1/4$, $\sqrt{n}\|\mathbf{r}_2(\theta_n, h_n, \cdot)\|_{L^2(P_n)} = O(n^{-1/2+2\delta_2-\delta_1} \vee n^{-3/2+4\delta_2})$. We get this result by deriving the magnitude of $\|\partial^2 f_n(\theta, h, \cdot)/\partial\theta_j\partial\theta_k\|_{L^2(P_n)}$ and also using the fact that $\|\theta_n - \theta_0\| = O(n^{-1/2+\delta_2})$.

Remark 2. *The efficiency bounds that we derive in the next section apply to* $(\delta_1, \delta_2) \in \Delta$*. Note that the condition* $\delta_2 < 1/4 + \delta_1/2$ *corresponds to the condition in Assumption A.3(i) under which the asymptotic distribution of GMM estimators is derived when the moment function is non-linear in* θ*. In our study, we will maintain this condition even in case of linearity since the induced family of densities appears to be non-linear in general as can be seen in* (B.1).

The condition on δ_2 *is more restrictive by ruling out values larger than or equal to 3/8. Although the results that we derive in this paper maintain this sufficient condition, they may still continue to hold for* $3/8 \le \delta_2 < 1/2$ *as illustrated by the simulations in Section 4.*

For any $(\delta_1, \delta_2) \in \Delta$, we have

$$\|\sqrt{n}(g_n - f_{n,0}) - \sqrt{n}(r_{n,\theta_0} \cdot (\theta_n - \theta_0) + A_n \cdot (h_n - h_0))\|_{L^2(P_n)} \to 0, \quad \text{as} \quad n \to \infty. \tag{26}$$

Therefore, we can define α_n satisfying (19) as any element of $L^2(P_n)$ such that:

$$\|\alpha_n - \sqrt{n}\left(r_{n,\theta_0} \cdot (\theta_n - \theta_0) + A_n \cdot (h_n - h_0)\right)\|_{L^2(P_n)} \to 0 \quad \text{as} \quad n \to \infty.$$

For the rest of our analysis, more relevant than the sequence $\{(\theta_n), (h_n)\} \in \Theta(\theta_0) \times \mathcal{C}(h_0)$ itself is its scaled limit which is some $(\eta, \beta) \in \mathbb{R}^p \times \mathcal{B}(h_0)$. The following proposition characterizes α_n in terms of η and β.

Proposition 3.2. *Let* R*,* J*, and* Λ_n *be defined as in* (14) *with* $(\delta_1, \delta_2) \in \Delta$*. Assume that:* θ_0 *satisfies* (1)*; the estimating function* $\phi(\cdot, \cdot)$ *satisfies* (5)*;* $\partial\rho(\theta_0)/\partial\theta'$ *is full column rank; Assumptions 1 and A.4 hold with* $r = 2$*. Then, with* $\{(\theta_n), (h_n)\} \in \Theta(\theta_0) \times \mathcal{C}(\theta_0)$*, the set* $\mathcal{H}_0$ *of* α_n*'s such that* (19) *holds is essentially given by*

$$\mathcal{H}_0 = \left\{\alpha_n \in L^2(P_n) : \alpha_n = -\frac{1}{2}\eta' J' \Sigma_n^{-1/2} \varphi + A_n \cdot \beta, \quad \eta \in \mathbb{R}^p, \beta \in \mathcal{B}(h_0)\right\}, \tag{27}$$

where $\varphi(\cdot) = \Sigma_n^{-1/2} \phi(\cdot, \theta_0)$ *and* $A_n = \nabla_h f_n(\theta_0, h_0, \cdot)$ *is given by Proposition B.1.*

In the statement of Proposition 3.2, by "essentially," we mean that any other solution $(\alpha_{1,n})$ of (19) satisfies $\|\alpha_n - \alpha_{1,n}\|_{L^2(P_n)} = o(1)$ for some $\alpha_n \in \mathcal{H}_0$. See comment following Theorem 3.1. Also, the fact that $\beta \in \mathcal{B}(h_0) \subset \mathcal{E}$ ensures that $A_n \cdot \beta = \beta$.

Proof: See Appendix. $\square$

Example 1. – *Linear IV, continued: In this example,* $Y_{n,i} := w_i = (z_i, y_i, x_i)$ *with distribution* P_n*. The elements of the tangent space are given by* $\alpha_n(w) = -(1/2)\eta' J' \Sigma_n^{-1/2} \varphi(w) + \beta(w)$*, with* $\eta \in \mathbb{R}^p$ *and* β *is any element of* $L^2(P_n)$ *orthogonal to* $y \mapsto (1, \varphi(w))$*;* $\varphi(w) = \Sigma_n^{-1/2}\phi(w, \theta_0)$*,* $\phi(w, \theta_0) := z(y - x'\theta_0)$*,* $\Sigma_n = Var(\phi(Y_{ni}, \theta_0))$*. Finally,* J *is defined as in* (14) *with* $\partial\rho_j(\theta_0)/\partial\theta' = -\Delta_{jj} C_j'$ *and* $s_1 = Rank(C_1)$*.*

Example 2. – *Optimal Prediction, continued: In this example,* $Y_{n,i} := w_i = (z_i, y_i, x_i)$ *with distribution* P_n*. The elements of the tangent space are given by* $\alpha_n(w) = -(1/2)\eta' J' \Sigma_n^{-1/2} \varphi(w) + \beta(w)$*, with* $\eta \in \mathbb{R}^p$ *and* β *is any element of* $L^2(P_n)$ *orthogonal to* $y \mapsto (1, \varphi(w))$*;* $\varphi(w) = \Sigma_n^{-1/2}\phi(w, \theta_0)$*, with* $\phi(w, \theta)$ *defined by* (12)*,* $\theta := (\alpha, \gamma)$*, and* $\Sigma_n = Var(\phi(Y_{n,i}, \theta_0))$*.*

To derive J*, we obtain:*

$$\frac{\partial\rho_1(\theta_0)}{\partial\theta'} = [-\mathbb{E}(z_i^2), 0] = [-\sigma^2, 0], \quad \text{and} \quad \frac{\partial\rho_2(\theta_0)}{\partial\theta'} = \big[0, -c\mathbb{E}\big(x_i^3 \sin(\gamma_0 x_i)\big)\big] \neq 0.$$

Using the definition (14)*, we obtain in this case,* $R = I_2$ *and*

$$J = \begin{pmatrix} -\sigma^2 & 0 \\ 0 & -c\mathbb{E}\big(x_i^3 \sin(\gamma_0 x_i)\big) \end{pmatrix}. \tag{28}$$

Remark 3. *As is commonly done, thanks to Proposition 3.2, we can index the sequence* $g_n^2 := f_n^2(\theta_n, h_n)$ *by its associated* $\alpha_n \in \mathcal{H}_0$*, i.e.* $\alpha_n \in \mathcal{H}_0$ *such that* (19) *holds or even by the parameter* $(\eta, \beta) \in \mathbb{R}^p \times \mathcal{B}(h_0)$*.*

Convolution results. Under the sequence of experiments g_n^2 as defined in (23) and the reference distribution $f_{n,0}^2$, the log-likelihood ratio L_n has the expression given in (17) with X_{ni} replaced by Y_{ni} and f_n by $f_{n,0}$. The LAN property of g_n^2 at (θ_0, h_0) follows from Theorem 3.1. Specifically, for any $\alpha_n \in \mathcal{H}_0$,

$$L_n \xrightarrow{d} N(-\sigma^2/2, \sigma^2),$$

under $f_{n,0}^2$ as $n \to \infty$, where $\sigma^2 = 4\lim_{n\to\infty} \|\alpha_n^2\|_{L^2(P_n)}$ if this limit exists.

This LAN property is key to the convolution result that we introduce next. We rule out cases of super-efficient estimators, by restricting ourselves to regular estimators of θ_0. The definition of regular estimator that we rely on is different from the standard one. A meaningful definition shall reflect the heterogeneity of convergence rates of standard estimators as obtained in (15). Our definition below accounts for the directions in which information about θ_0 has the potential to be maximum.

Definition 1. ***(Λ_n-Regularity)*** *An estimator* $\tilde{\theta}_n$ *of* θ_0 *is* Λ_n*-regular at* $f_{n,0}^2$ *if, for every sequence* $g_n(\cdot) := f_n(\theta_n, h_n, \cdot)$ *with* $\{(\theta_n), (h_n)\} \in \Theta(\theta_0) \times \mathcal{C}(h_0)$*,* $\Lambda_n R^{-1}(\tilde{\theta}_n - \theta_n)$ *converges in distribution under* g_n^2 *and* $f_{n,0}^2 = f_n^2(\theta_0, h_0)$ *to the same limit* S*.*

Remark 4. *Note that S in this definition may depend on R. However, the Λ_n-regularity property of a sequence of estimators $\tilde{\theta}_n$ is not associated to a particular rotation as the definition may suggest. Indeed, we can show that if Definition 1 holds for $\tilde{\theta}_n$, it continue to hold if R is replaced by a different rotation matrix, say $\mathbf{R} = RA$ (see Remark 1). In this case, the limiting distribution is $A'S$ instead.*

To introduce our main result, we observe that, since $h_0 = 0$, $\mathcal{B}(h_0)$ is a closed subspace of $L^2(P_n)$ hence, α_n's in Proposition 3.2 can also be written

$$\eta' \left(J'\Sigma_n^{-1/2}\varphi - A_n \cdot \mathbf{b} \right),$$

with $\eta \in \mathbb{R}^p$, $\mathbf{b} = (\beta_1, \dots, \beta_p) \in \mathcal{B}(h_0)^p$ and $A_n \cdot \mathbf{b} := (A_n \cdot \beta_1, \dots, A_n \cdot \beta_p)'$.

Let $A_n \cdot \mathbf{b}_n^*$, with $\mathbf{b}_n^* \in \mathcal{B}(h_0)^p$, be the orthogonal projection of $-\frac{1}{2}J'\Sigma_n^{-1/2}\varphi$ onto $\{A_n \cdot \mathbf{b} : \mathbf{b} \in \mathcal{B}(h_0)^p\}$ and define (the efficient score in the direction of θ) as

$$s_n = -\frac{1}{2}J'\Sigma_n^{-1/2}\varphi - A_n \cdot \mathbf{b}_n^* \quad \text{and} \quad I_* = 4 \lim_{n\to\infty} \langle s_n, s_n \rangle$$

if this limit exists. We have the following result.

Theorem 3.3. *Let $\tilde{\theta}_n$ be an estimator of θ_0, Λ_n-regular at $f_{n,0}^2$ with limit distribution S. If the conclusion of Proposition 3.2 holds - that is: the set of α_n's such that (19) holds is given by $\mathcal{H}_0$ in (27) - and I_* exists and is nonsingular, then:*

$$S \overset{d}{=} Z + U,$$

where $Z \sim N(0, I_*^{-1})$ and is independent of the random vector U.

Proof: See Appendix. □

Theorem 3.3 states that any regular estimator of θ_0 has an asymptotic variance that is at least as large as I_*^{-1}. The next corollary gives a more explicit expression of this bound in terms of moments of the estimating function $\phi(Y, \theta)$.

Corollary 3.4. *Let R, J, and Λ_n be defined as in (14) with $(\delta_1, \delta_2) \in \Delta$. Assume that: θ_0 satisfies (1); the estimating function $\phi(\cdot, \cdot)$ satisfies (5); $\partial\rho(\theta_0)/\partial\theta'$ is full column rank; Assumptions 1 and A.4 hold with $r = 2$; and, as $n \to \infty$, $\Sigma_n := \mathbb{E}_{P_n}[\phi(Y, \theta_0)\phi(Y, \theta_0)'] \to \Sigma$ a symmetric positive definite matrix. If $\tilde{\theta}_n$ is Λ_n-regular estimator of θ_0 with limit distribution S, then*

$$S \overset{d}{=} Z + U, \tag{29}$$

where $Z \sim N(0, I_*^{-1})$, with $I_* = J'\Sigma^{-1}J$ and Z independent of U.

Proof: See Appendix. □

Example 1. – *Linear IV, continued: The limit distribution of any regular estimator of θ_0 in the model (9) is a convolution of two independent variables Z and U with $Z \sim N(0, I_*^{-1})$; $I_* = J'\Sigma^{-1}J$, J previously defined, and $\Sigma = \lim_{n\to\infty} \Sigma_n$.*

Example 2. – *Optimal Prediction, continued: In this example, the limit distribution of any regular estimator of* $\theta_0 := (\alpha_0, \gamma_0)$ *in the model* (10)-(11) *is a convolution of two independent variables* Z *and* U *with* $Z \sim N(0, I_*^{-1})$; $I_*^{-1} = J^{-1}\Sigma J^{-1}$, *with* J *given by* (28), *and* $\Sigma = \lim_{n\to\infty}\Sigma_n$ *with* $\Sigma_n = Var(\phi(Y_{n,i}, \theta_0))$, $Y_{n,i} := w_i = (y_i, x_i, z_i)'$ *with distribution* P_n.

Corollary 3.4 sets $L_b := I_*^{-1} = \left(J'\Sigma^{-1}J\right)^{-1}$ as the lowest asymptotic variance reachable by any regular estimator of θ_0. Note that this result holds regardless of the value of $s_1 = \text{Rank}(\partial\rho_1(\theta_0)/\partial\theta')$. If $s_1 = 0$ or p, then

$$J = \frac{\partial\rho(\theta_0)}{\partial\theta'} \quad \text{and} \quad L_b = \left(\frac{\partial\rho(\theta_0)'}{\partial\theta}\Sigma^{-1}\frac{\partial\rho(\theta_0)}{\partial\theta'}\right)^{-1}.$$

In the case where $0 < s_1 < p$, J is given by (14) and the bound is as given above. This is effectively the efficiency bound for the estimation of $\nu_0 = R^{-1}\theta_0$. However, this result seems to channel more information than that. From the previous discussion, any estimator $\tilde{\theta}$ of θ_0 that is Λ_n-regular for one choice of rotation stays so for any other rotation defined by (14). In addition, the convolution result above shows that being efficient in terms of one rotation implies efficiency in any other rotation. This provides some rational to the notion that, when $0 < s_1 < p$, a regular and efficient estimator $\tilde{\theta}$ is one that is Λ_n-regular for one choice of rotation and reaches the asymptotic semiparametric efficiency bound for that rotation.

One additional point that is worth mentioning is that Dovonon et al. (2023) have established that $\det[(J'\Sigma^{-1}J)^{-1}]$ is rotation invariant. Also, the asymptotic variance of a regular estimator $\tilde{\theta}$ is given by

$$I_*^{-1} + V, \quad \text{with} \quad V = Var(U).$$

We know that $\det(I_*^{-1} + V) \geq \det(I_*^{-1})$, with equality if and only if $V = 0$.[3] We can therefore relate efficiency of any regular estimator $\tilde{\theta}$ to the fact that the determinant of its asymptotic variance is equal to $\det(I_*^{-1})$ which is rotation invariant. We recall that the determinant of the variance-covariance matrix, also known as *generalized variance* is introduced by Wilks (1932) as the scalar measure of dispersion in a multivariate statistical population.

Finally, in relation to GMM estimation, from (15), the two-step GMM estimator (2SGMM) $\hat{\theta}_{n,\Sigma^{-1}}$ using the weighting matrix W_n with the inverse of $\Sigma = \lim_{n\to\infty} Var_{P_n}(\phi(Y_{ni}, \theta_0))$ as limit, is asymptotically distributed $N(0, (J'\Sigma^{-1}J)^{-1})$. This implies that the bound derived by Corollary 3.4 is sharp. Note that this choice of weighting matrix is known to be efficient in the standard GMM estimation setting (see Chamberlain, 1987) and also in singularity settings of first-order local identification failure (see Dovonon and Atchadé, 2020). Further, we will show in the next section that GMM estimators are regular and this will bring to light the efficiency status of 2SGMM among regular estimators.

Along with the convolution result in Corollary 3.4, we also derive an asymptotic minimax optimality result for a general class of loss functions. Let $\ell : \mathbb{R}^p \to \mathbb{R}_+$ be a loss function that is subconvex, i.e., $\{x : \ell(x) \leq a\}$ is closed, convex and symmetric for every $a \geq 0$. We have the following.

[3]See Magnus and Neudecker (2002, Th. 22, p.21).

Theorem 3.5. *Under the same conditions as in Corollary 3.4, if ℓ is subconvex and $\tilde{\theta}_n$ is a measurable sequence of estimator of θ_0, then*

$$\sup_{I \subset \mathcal{H}_0} \liminf_{n\to\infty} \sup_{\alpha_n \in I} \mathbb{E}_{g_{n,\alpha_n}} \ell\left(\Lambda_n R^{-1}\left(\tilde{\theta}_n - \theta_n\right)\right) \geq \mathbb{E}\ell(Z),$$

where Z is defined as in Corollary 3.4 and g^2_{n,α_n} is a sequence $f^2(\theta_n, h_n, \cdot)$ such that (19) holds. The first supremum is taken over all finite subset I of $\mathcal{H}_0$.

Using Corollary 3.4, the proof of this result follows readily by the application of Theorem 3.11.5 of van der Vaart and Wellner (1996, p.417).

Regularity of the GMM estimator. We now establish that the GMM estimator is Λ_n-regular at $f_{n,0}$. Consider the GMM estimator, $\hat{\theta}_n$, defined by (13) with a sequence of weighting matrix W_n converging in probability under P_n to W, a symmetric positive definite matrix. Equation (15) gives the asymptotic distribution of $\hat{\theta}_n$, under P_n:

$$\Lambda_n R^{-1}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N(0, \Omega(W))$$

which is valid under (1), (5) and Assumptions A.1-A.3. To claim regularity for $\hat{\theta}_n$, we will establish that

$$\Lambda_n R^{-1}(\hat{\theta}_n - \theta_n) \xrightarrow{d} N(0, \Omega(W)), \quad \text{under } g_n^2 := f_n^2(\theta_n, h_n),$$

with $\Lambda_n R^{-1}(\theta_n - \theta_0) - \eta \longrightarrow 0$ and $\sqrt{n}(h_n - h_0) - \beta \longrightarrow 0$ in $L^2(P_n)$ for some $\eta \in \mathbb{R}^p$ and $\beta \in \mathcal{E}$.

We will use the fact that the measures $\{\prod_{i=1}^n g_n^2(y_i)\}$ and $\{\prod_{i=1}^n f_{n,0}^2(y_i)\}$ are contiguous, see Theorem 3.1. That is, for each sequence of sets F_n measurable on the probability space $(\mathbb{R}^d \times \cdots \times \mathbb{R}^d, \mathcal{B}(\mathbb{R}^d \times \cdots \times \mathbb{R}^d), \mathbb{P}_n := P_n \otimes \cdots \otimes P_n)$, $\mathbb{P}_n(F_n) \to 0$, as $n \to \infty$ implies that $\mathbb{Q}_n(F_n) \to 0$, where $\mathbb{Q}_n$ has density $\prod_{i=1}^n g_n^2(y_i)$ with respect to $\mathbb{P}_n$. (The products in the definition are n-fold.) The consequence of contiguity is that any sequence of random variable of order $o_P(1)$ (respectively $O_P(1)$) under P_n are also $o_P(1)$ (respectively $O_P(1)$) under g_n^2. We establish regularity of GMM by strengthening Assumptions A.1-A.3 by the following assumption:

Assumption 2. *(a) There exists a neighbourhood $\mathcal{N}$ of θ_0 and a constant $C > 0$ such that, for all g in a $L^2(P_n)$-neighbourhood of $f_{n,0}$,*

$$\sup_{\theta \in \mathcal{N}} \int \|\phi(y,\theta)\|^4 g^2(y) dP_n(y) \leq C.$$

(b) For any non-random sequence (θ_n) such that $\theta_n \to \theta_0$, as $n \to \infty$, $\int \phi(y,\theta_n)\phi(y,\theta_n)' dP_n \to \Sigma$, with $\Sigma := \lim_{n\to\infty} Var_{P_n}(\phi(Y_{ni}, \theta_0))$.

This assumption is useful to establish that $\sqrt{n}\bar{\phi}_n(\theta_n)$ converges in distribution to $N(0, \Sigma)$, under g_n^2. Part (a) requires fourth moments for the estimating function $\phi(y, \theta)$ under distributions near the reference distribution P_n. We can interpret Part (b) as a continuity assumption. If P_n were fixed in n, it would follow from the continuity of $\theta \mapsto \phi(y, \theta)$ for P_n-almost all y and some dominance condition.

Proposition 3.6. *Assume θ_0 satisfies (1) and the estimating function $\phi(\cdot, \cdot)$ satisfies (5). If Assumptions A.1-A.3 and 2 hold, then the GMM estimator $\hat{\theta}_n$ is Λ_n-regular.*

Proof: See Appendix. □

This proposition establishes the regularity of GMM estimators. Asymptotic normality and the conclusion of Corollary 3.4 imply that the asymptotic distribution of $\hat{\theta}_n$ is the convolution of independent random variables Z and U with $Z \sim N(0, (J'\Sigma^{-1}J)^{-1})$ and $U \sim N(0, \Omega(W) - (J'\Sigma^{-1}J)^{-1})$. The choice $W = \Sigma^{-1}$ yields $U \equiv 0$ making 2SGMM, $\hat{\theta}_{n\Sigma^{-1}}$, asymptotically semiparametrically efficient among the family of regular estimators.

This result also establishes that, although the moment function is asymptotically vanishing uniformly over the parameter space, 2SGMM is not *inadmissible* in the sense of Andrews and Mikusheva (2022a). It is important to note that their statement that GMM is inadmissible in weakly identified models hinges crucially on the concentration measure $\sqrt{n}\|\rho_n(\theta)\|_\infty$ being bounded. Our result does not contradict theirs since $\sqrt{n}\|\rho_n(\theta)\|_\infty$ is unbounded in our framework.

Further comments

1. This paper focuses mainly on moment condition models with identification properties outlined by (5). Nevertheless, the efficiency results derived would still hold in some relaxed version of this statement. This is the case if in (5), for $j = 1, 2$,

$$\mathbb{E}_{P_n}(\phi_j(Y_{ni}, \theta)) = n^{-\delta_j}\rho_j(\theta) \text{ is replaced by } \mathbb{E}_{P_n}(\phi_j(Y_{ni}, \theta)) = n^{-\delta_j}\rho_j(\theta) + r_{jn}(\theta),$$

   with $\|r_{jn}(\theta)\|$ uniformly $o(n^{-\delta_j}\|\rho_j(\theta)\|)$, that is, there exists a sequence $k_n \to 0$ such that: for any $\theta \in \Theta$, $\|r_{jn}(\theta)\| \leq k_n \cdot n^{-\delta_j}\|\rho_j(\theta)\|$.

   In this framework, one shall maintain that, in a neighbourhood $\mathcal{N}_{\theta_0}$ of θ_0, and for $j = 1, 2$:

$$\mathbb{E}_{P_n}\left(\frac{\partial \phi_j(Y_{ni}, \theta)}{\partial \theta'}\right) = n^{-\delta_j}\frac{\partial \rho_j(\theta)}{\partial \theta'} + o(n^{-\delta_2})$$

   and for $s = 1, \ldots, k_j$,

$$\mathbb{E}_{P_n}\left(\frac{\partial^2 \phi_{js}(Y_{ni}, \theta)}{\partial \theta \partial \theta'}\right) = n^{-\delta_j}\frac{\partial^2 \rho_{js}(\theta)}{\partial \theta \partial \theta'} + o(n^{-\delta_j}).$$

2. A general characteristic of moment condition models with mixed identification strength is that the rate of convergence of common estimators is quite related to the magnitude of the components of the moment function. In that respect, a linear one-to-one transformation of the model may result in a change in the convergence rate structure of the estimator including, e.g., the number of directions of faster convergence rate. In spite of this, the results of this paper show that efficiency of the 2SGMM is warranted for any specific version of of the moment function considered. Note that in standard (strong) identification framework, the 2SGMM associated to a moment condition model is asymptotically equivalent to that of any linear one-to-one transform of that model.

3. In a more recent paper, Antoine and Renault (2021) consider a more general family of moment condition models in which point identification is characterized by:

$$\text{There exists } \nu \in [0, 1/2[\text{ such that } : \forall \epsilon > 0, \liminf_{n\to\infty} n^{\nu} \inf_{\|\theta-\theta_0\|>\epsilon} \|\rho_n(\theta)\| > 0. \tag{30}$$

They establish consistency of GMM estimators under this condition and their asymptotic normality under further regularity assumptions. We observe that the models with mixed identification strength considered in this paper fit with this property while being only a subset. There are models significantly different from those studied in this paper – such as *moment condition models with additively separable moment functions* considered by Stock and Wright (2000) – that fit (30) as well. There are no obvious reasons for us to believe that the bounds derived in this paper extend to all models with the property in (30). A more careful study may be needed to exhibit efficient estimators in this framework.

4. An interesting extension that we plan for future research is to investigate the meaning of efficiency when $\delta \geq 1/2$. Kaji (2021) focuses on this range and adopts a slightly different approach than ours. He considers a candidate limit measure P that aligns with the model at $n = \infty$ and, in particular, is uninformative about the model parameter. Then, he explores sequences of relevant measures P_n (each) identifying the parameter value and converging to P. While the parameter (considered as a function of the data distribution P_n) is not continuous, efficiency can then be built on some underlying parameter that is estimable and informative about the structural unidentified parameter.

   In our analysis, though, we consider P_n, the probability distribution of the data at a given n, as the reference probability measure and obtain the tangent space at P_n of local relevant measures. Then, we develop a convolution theory based on the limit of this (sample size dependent) tangent space. We have been able to obtain positive results when it comes to analyzing the case where $0 \leq \delta < 1/2$. Such an extension will help shed some light on how the efficiency properties transition through $\delta = 1/2$ and will also allow a meaningful connection to the work of Kaji (2021).

4 Simulations

We analyze the finite sample performance of the two-step GMM estimator of θ_0 in the moment condition model (1) in the presence of moment restrictions with nonstandard or mixed identification strength. We focus on the following linear IV model with conditional heteroskedasticity and two endogenous variables, as it offers a suitable framework for this exercise:

$$\begin{cases} y_i = x_{1i}\theta_1 + x_{2i}\theta_2 + u_i, \quad i = 1, \ldots, n \\ x_{1i} = z_{1i}\pi_{1n} + v_{1i}, \quad x_{2i} = z_{2i}\pi_{2n} + z_3\pi_{3n} + v_{2i}, \\ u_i = \sigma_\varepsilon^{-1}(x_{1i}^2\varepsilon_i - \mu_{x\varepsilon}), \quad \varepsilon_i = \rho v_{1i} + \rho v_{2i} + \eta_i, \\ \sigma_\varepsilon^2 = Var(x_{1i}^2\varepsilon_i), \quad \mu_{x\varepsilon} := \mathbb{E}(x_{1i}^2\varepsilon_i) = 2\rho\sqrt{2}, \end{cases} \tag{31}$$

where $\pi_{1n} = 1.48n^{-\delta_1}$, $\pi_{2n} = \pi_{3n} = 1.48n^{-\delta_2}$; $y_i \in \mathbb{R}$ is the ith observation on the dependent variable; $x_{1i} \in \mathbb{R}$ and $x_{2i} \in \mathbb{R}$ are observations on two possibly endogenous regressors; θ_1 and θ_2 are unknown scalar structural parameters; z_{1i}, z_{2i}, z_{3i} are instrumental variables, whose strengths are δ_1, δ_2 and δ_2 respectively [see Dovonon et al. (2023)]; u_i is a structural disturbance and (v_{1i}, v_{2i}) are reduced-form disturbances. The variance σ_ε^2 of $x_{1i}^2\varepsilon_i$ is explicitly given by $\sigma_\varepsilon^2 = 3\pi_{1n}^4 + 6\pi_{1n}^2 + 84\pi_{1n}^2\rho^2 + 732\rho^2 + 15$. The expression of the structural errors u_i in (31) clearly illustrates the presence of conditional heteroskedasticity in this IV model. The true values of θ_1 and θ_2 are set at $\theta_{01} = \theta_{02} = 0.1$, and $(v_1, v_2, \eta, z_1, z_2, z_3)' \sim N(0, \mathbb{V})$, where

$$\mathbb{V} = \begin{pmatrix} V & 0 & 0 \\ 0 & I_2 & 0 \\ 0 & 0 & V_z \end{pmatrix}, \qquad V = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}, \qquad V_z = \begin{pmatrix} 1 & \rho_z \\ \rho_z & 1 \end{pmatrix}.$$

In (31) ρ measures the correlation between ε_i and v_{ji}, $j = 1, 2$, and is kept fixed across observations. Note from the above parametrization that ρ also determines the degree of endogeneity in the model (i.e., the correlation between the structural error u_i and the reduced-form errors v_{1j}, $j = 1, 2$) when the sample n goes to infinity. We set ρ to 0.5 and 0.0925. For $\rho = 0.5$, $\text{Corr}(u_i, v_{ji})$ tends to 0.533 as n grows, while for $\rho = 0.0925$, $\text{Corr}(u_i, v_{ji})$ tends to 0.301, for both $j = 1$ and 2. Therefore, $\rho = 0.5$ corresponds to relatively high endogeneity in the model, while $\rho = 0.0925$ implies moderate endogeneity in the model. Throughout the experiments, following Dovonon et al. (2023), we consider cases where z_1, z_2 and z_3 have equal strength – $\delta_1 = \delta_2 \in \{0, 0.1, 0.2, 0.3, 0.4, 0.45, 0.5\}$ – and cases where they have mixed strength – $(\delta_1, \delta_2) \in \{(0, 0.2), (0, 0.3), (0, 0.4), (0.1, 0.2)(0.1, 0.3), (0.3, 0.4)\}$. We set the sample size n to 100, 500, 1000, 5000, 8000, and 10000.

We evaluate the performance of the baseline GMM estimator, which is the two-step GMM in (13) using the optimal weighting matrix $\hat{W}_n^{opt} = \left(\frac{1}{n}\sum_{i=1}^n \hat{u}_i^2 z_i z_i'\right)^{-1}$, where $\hat{u}_i$ represents the 2SLS residuals and $z_i = (z_{1i}, z_{2i}, z_{3i})'$. We compare its performance with two non-optimal GMM estimators of θ: (1) the 2SLS estimator obtained by setting $W_n = \left(\frac{1}{n}\sum_{i=1}^n z_i z_i'\right)^{-1}$ in (13); and (2) the naive GMM estimator obtained with $W_n = I_k$ in (13). To assess and compare these estimators, we use performance measures, including the component-wise mean squared error (MSE) and the generalized variance (gVAR), quantified by the determinant of the MSE matrix.

Tables 1-2 display performance ratios (naive estimator to optimal GMM and 2SLS estimator to optimal GMM) for various sample sizes under different levels of identification strength. Table 1 represents scenarios with relatively high endogeneity ($\rho = 0.5$), while Table 2 reflects moderate endogeneity ($\rho = 0.0925$). The results consistently show that, for both performance metrics (MSE and gVAR), across all levels of endogeneity ($\rho \in 0.0925, 0.5$) and sample sizes, the benchmark optimal two-step GMM outperforms both the 2SLS estimator and the naive GMM estimator with $W_n = I_k$. This superiority of the optimal two-step GMM is particularly pronounced in smaller samples but tends to stabilize as the sample size increases. These findings align with our theoretical results. Notably, even in the case of $\delta_1 = \delta_2 = 0.5$ under which GMM is inconsistent, the optimal two-step GMM estimator is favored based on the presented ratios. A possible explanation of this result is that the exact finite n density draws information from the data even when $\pi_{jn} = O\left(\frac{1}{\sqrt{n}}\right)$ but is insufficient to deliver a consistent estimator (see e.g., Phillips, 1980).

5 Concluding remarks

This paper is concerned with efficient estimation in moment condition models with mixed identification strength. These models are point identifying at any given sample size but their moment function drifts to zero uniformly over the parameter space as the sample size grows. This feature makes identification somewhat weak since the moment function becomes uninformative in the limit. When the moment function does not drift to zero too fast, consistent estimation is possible and GMM estimators are shown to be asymptotically normally distributed.

The purpose of this paper is to derive semiparametric efficiency bounds for parameter estimation in these models. We rely on the approach of Dovonon and Atchadé (2020) that we refine to account for the fact that the sampling process follows a drifting distribution P_n that depends on the sample size, n, instead of a fixed distribution as commonly considered in the literature.

We show that the asymptotic minimum variance bound for the estimation by regular estimators is given by $(J'\Sigma^{-1}J)^{-1}$, where J is given by (14) in Section 2. This bound corresponds to the asymptotic variance of the GMM estimator using a weighting matrix W_n converging to Σ^{-1}, where Σ is the limit variance under P_n of the estimating function evaluated at θ_0. This is the the well-known two-step GMM estimator. We establish that this estimator is regular and also asymptotically minimax efficient with respect to a large class of loss functions. Our result extends that of Chamberlain (1987) to the class of moment condition models with mixed identification strength.

One possible extension that we plan for future work is to consider models describing weakly dependent data. Hallin et al. (2015) have developed a framework useful to study such models in the parametric framework. An extension of their approach to semiparametric models can be an interesting contribution. The main challenge that we foresee for moment condition models with dependent data is related to the formulation of the dynamics in the data generating process that shall be general enough to accommodate a relevant class of models while being explicit enough to fit with the framework of Hallin et al. (2015).

Acknowledgements

The paper has benefited from many comments of the Co-Editor Serena Ng, an Associate Editor and two anonymous referees. The authors also thank Antoine Bertille, Peter C.B. Phillips, Werner Ploberger and Bas Werker for helpful comments. We also thank participants of the 2019 Africa Meeting of The Econometric Society in Morocco, the 2022 CIREQ conference in honour of Éric Renault (Montreal), the 2023 Asian Meeting of The Econometric Society at NTU Singapore, the 2023 NBER Time Series conference in Montreal, the 2023 Canadian Econometric Study Group meeting in Hamilton and seminar participants at Monash University for helpful discussions. This research is supported by the Social Sciences and Humanities Research Council of Canada and by the Australian Research Council [grant number DP200101498].

Table 1: Relative performance of the Optimal Two-Step GMM: $\rho = 0.5$

		Ratio Naive to Optimal GMM						Ratio 2SLS to Optimal GMM					
IV strength $\downarrow$	$n \rightarrow$	100	500	1000	5000	8000	10000	100	500	1000	5000	8000	10000
$\delta_1 < \delta_2$		MSE-ratio ($\hat{\theta}_1$)											
0	0.2	1.374	1.161	1.104	1.029	1.020	1.017	1.387	1.164	1.106	1.029	1.020	1.017
0	0.3	1.381	1.159	1.105	1.030	1.021	1.018	1.389	1.161	1.107	1.030	1.021	1.018
0	0.4	1.383	1.157	1.100	1.029	1.020	1.017	1.391	1.159	1.102	1.029	1.020	1.017
0.1	0.2	1.513	1.205	1.130	1.038	1.023	1.022	1.529	1.213	1.134	1.040	1.024	1.023
0.1	0.3	1.433	1.245	1.136	1.037	1.025	1.020	1.455	1.252	1.140	1.039	1.026	1.021
0.3	0.4	1.356	1.314	1.249	1.085	1.056	1.057	1.338	1.317	1.248	1.095	1.065	1.067
		MSE-ratio ($\hat{\theta}_2$)											
0	0.2	1.617	1.298	1.211	1.083	1.064	1.059	1.604	1.294	1.211	1.081	1.064	1.057
0	0.3	1.502	1.284	1.264	1.100	1.075	1.062	1.486	1.276	1.255	1.097	1.072	1.061
0	0.4	1.428	1.329	1.210	1.088	1.072	1.071	1.405	1.302	1.198	1.079	1.063	1.063
0.1	0.2	1.710	1.393	1.296	1.111	1.074	1.073	1.691	1.386	1.291	1.109	1.072	1.070
0.1	0.3	1.636	1.374	1.282	1.115	1.094	1.080	1.624	1.369	1.277	1.113	1.091	1.078
0.3	0.4	1.534	1.397	1.293	1.115	1.086	1.085	1.473	1.362	1.274	1.104	1.079	1.079
		gVAR-ratio											
0	0.2	4.930	2.271	1.787	1.240	1.178	1.159	4.937	2.270	1.792	1.237	1.177	1.156
0	0.3	4.303	2.214	1.949	1.283	1.205	1.169	4.259	2.195	1.929	1.277	1.198	1.166
0	0.4	3.898	2.364	1.773	1.252	1.195	1.185	3.816	2.279	1.743	1.232	1.178	1.168
0.1	0.2	6.650	2.818	2.144	1.329	1.206	1.203	6.636	2.826	2.146	1.330	1.204	1.196
0.1	0.3	5.503	2.922	2.118	1.338	1.258	1.213	5.588	2.937	2.119	1.337	1.253	1.211
0.3	0.4	4.379	3.374	2.606	1.464	1.317	1.315	3.923	3.215	2.528	1.460	1.319	1.325
$\delta_1 = \delta_2$		MSE-ratio ($\hat{\theta}_1$)											
0	0	1.376	1.156	1.106	1.029	1.020	1.018	1.385	1.159	1.107	1.029	1.020	1.018
0.1	0.1	1.544	1.231	1.134	1.036	1.026	1.020	1.540	1.239	1.139	1.037	1.027	1.021
0.2	0.2	1.596	1.303	1.186	1.050	1.033	1.027	1.610	1.316	1.195	1.056	1.037	1.031
0.3	0.3	1.382	1.338	1.278	1.090	1.057	1.048	1.347	1.338	1.291	1.098	1.067	1.057
0.4	0.4	1.324	1.236	1.209	1.147	1.079	1.071	1.282	1.129	1.169	1.086	1.060	1.058
0.45	0.45	1.223	1.233	1.206	1.077	1.244	1.141	1.096	1.147	1.068	1.024	1.030	1.022
0.5	0.5	1.243	1.128	1.167	1.108	1.085	1.078	1.127	1.081	1.064	1.024	1.016	1.009
		MSE-ratio ($\hat{\theta}_2$)											
0	0	1.600	1.297	1.238	1.083	1.064	1.067	1.597	1.294	1.237	1.081	1.064	1.064
0.1	0.1	1.790	1.334	1.275	1.124	1.097	1.074	1.767	1.329	1.271	1.122	1.095	1.072
0.2	0.2	1.730	1.438	1.246	1.111	1.094	1.087	1.686	1.431	1.243	1.108	1.093	1.084
0.3	0.3	1.509	1.367	1.277	1.113	1.082	1.088	1.442	1.352	1.263	1.109	1.077	1.084
0.4	0.4	1.421	1.286	1.243	1.135	1.097	1.093	1.341	1.237	1.211	1.109	1.077	1.076
0.45	0.45	1.289	1.259	1.283	1.117	1.107	1.079	1.227	1.202	1.174	1.048	1.019	1.041
0.5	0.5	1.125	1.124	1.193	1.069	1.060	1.086	1.070	1.090	1.090	1.021	1.021	1.021
		gVAR-ratio											
0	0	4.844	2.249	1.874	1.241	1.178	1.180	4.892	2.252	1.877	1.238	1.177	1.174
0.1	0.1	7.603	2.695	2.091	1.355	1.267	1.200	7.383	2.708	2.094	1.353	1.264	1.198
0.2	0.2	7.690	3.510	2.183	1.362	1.277	1.247	7.417	3.550	2.207	1.370	1.285	1.251
0.3	0.3	4.348	3.348	2.662	1.473	1.306	1.300	3.775	3.274	2.658	1.481	1.321	1.312
0.4	0.4	3.694	2.539	2.264	1.689	1.400	1.372	3.088	1.967	2.019	1.453	1.305	1.296
0.45	0.45	2.481	2.418	2.385	1.435	1.784	1.526	1.801	1.911	1.566	1.172	1.152	1.134
0.5	0.5	1.974	1.607	1.798	1.432	1.315	1.362	1.487	1.391	1.342	1.102	1.077	1.063

Table 2: Relative performance of the Optimal Two-Step GMM: $\rho = 0.0925$

		Ratio Naive to Optimal GMM						Ratio 2SLS to Optimal GMM					
IV strength $\downarrow$	$n \rightarrow$	100	500	1000	5000	8000	10000	100	500	1000	5000	8000	10000
$\delta_1 < \delta_2$		MSE-ratio ($\hat{\theta}_1$)											
0	0.2	1.128	1.066	1.056	1.024	1.017	1.015	1.131	1.067	1.057	1.024	1.018	1.015
0	0.3	1.140	1.068	1.053	1.025	1.017	1.016	1.142	1.069	1.054	1.025	1.017	1.017
0	0.4	1.150	1.075	1.058	1.024	1.018	1.015	1.152	1.075	1.059	1.024	1.018	1.016
0.1	0.2	1.259	1.138	1.121	1.044	1.028	1.024	1.262	1.141	1.123	1.045	1.029	1.025
0.1	0.3	1.217	1.162	1.107	1.043	1.031	1.025	1.223	1.165	1.109	1.045	1.032	1.026
0.3	0.4	1.549	1.228	1.182	1.059	1.044	1.035	1.524	1.225	1.174	1.063	1.046	1.039
		MSE-ratio ($\hat{\theta}_2$)											
0	0.2	1.180	1.096	1.060	1.027	1.017	1.021	1.175	1.094	1.058	1.026	1.017	1.020
0	0.3	1.216	1.093	1.073	1.027	1.024	1.016	1.205	1.088	1.068	1.024	1.021	1.012
0	0.4	1.186	1.109	1.065	1.024	1.023	1.026	1.162	1.089	1.057	1.018	1.015	1.019
0.1	0.2	1.371	1.210	1.153	1.061	1.044	1.044	1.360	1.205	1.150	1.059	1.043	1.042
0.1	0.3	1.320	1.193	1.151	1.063	1.050	1.048	1.309	1.187	1.146	1.060	1.048	1.045
0.3	0.4	1.524	1.279	1.188	1.078	1.052	1.053	1.462	1.243	1.171	1.069	1.046	1.046
		gVAR-ratio											
0	0.2	1.772	1.365	1.254	1.107	1.070	1.074	1.766	1.364	1.250	1.105	1.070	1.071
0	0.3	1.922	1.364	1.276	1.109	1.085	1.067	1.893	1.354	1.266	1.102	1.080	1.059
0	0.4	1.860	1.421	1.270	1.098	1.085	1.086	1.791	1.372	1.251	1.087	1.068	1.072
0.1	0.2	2.978	1.894	1.669	1.225	1.152	1.143	2.947	1.891	1.667	1.223	1.150	1.140
0.1	0.3	2.579	1.923	1.624	1.229	1.172	1.155	2.561	1.912	1.616	1.227	1.169	1.149
0.3	0.4	5.070	2.467	1.973	1.304	1.205	1.187	4.504	2.320	1.889	1.290	1.197	1.179
$\delta_1 = \delta_2$		MSE-ratio ($\hat{\theta}_1$)											
0	0	1.128	1.072	1.057	1.023	1.017	1.015	1.131	1.073	1.057	1.023	1.018	1.015
0.1	0.1	1.222	1.143	1.116	1.042	1.030	1.026	1.227	1.146	1.118	1.044	1.031	1.026
0.2	0.2	1.350	1.226	1.150	1.052	1.038	1.030	1.334	1.230	1.154	1.055	1.040	1.032
0.3	0.3	1.766	1.223	1.193	1.063	1.053	1.039	1.463	1.226	1.186	1.065	1.055	1.042
0.4	0.4	1.280	1.059	1.215	1.074	1.046	1.050	1.222	1.098	1.132	1.048	1.036	1.040
0.45	0.45	1.712	1.148	1.371	1.027	1.014	1.144	1.437	1.099	1.037	1.019	1.012	1.025
0.5	0.5	1.043	1.111	1.107	1.382	1.014	1.058	1.039	1.069	1.050	1.017	1.011	1.052
		MSE-ratio ($\hat{\theta}_2$)											
0	0	1.186	1.101	1.066	1.020	1.020	1.014	1.184	1.100	1.063	1.017	1.018	1.012
0.1	0.1	1.303	1.197	1.146	1.063	1.054	1.035	1.297	1.193	1.143	1.062	1.052	1.033
0.2	0.2	1.363	1.244	1.187	1.078	1.052	1.044	1.352	1.241	1.184	1.075	1.050	1.042
0.3	0.3	1.490	1.251	1.192	1.082	1.055	1.050	1.399	1.235	1.185	1.079	1.052	1.046
0.4	0.4	1.470	1.145	1.170	1.088	1.054	1.057	1.366	1.154	1.144	1.069	1.043	1.046
0.45	0.45	1.483	1.185	1.187	1.059	1.035	1.061	1.337	1.156	1.104	1.036	1.022	1.032
0.5	0.5	1.023	1.122	1.130	1.124	1.070	1.032	1.017	1.078	1.057	1.014	1.020	1.046
		gVAR-ratio											
0	0	1.789	1.394	1.269	1.088	1.076	1.058	1.793	1.393	1.264	1.083	1.074	1.055
0.1	0.1	2.534	1.872	1.635	1.229	1.179	1.125	2.531	1.868	1.631	1.228	1.177	1.124
0.2	0.2	3.390	2.327	1.862	1.285	1.193	1.157	3.254	2.329	1.865	1.287	1.193	1.157
0.3	0.3	6.748	2.344	2.019	1.323	1.233	1.189	4.156	2.291	1.975	1.320	1.232	1.186
0.4	0.4	3.370	1.543	2.020	1.363	1.215	1.234	2.678	1.635	1.677	1.256	1.168	1.184
0.45	0.45	5.510	1.850	2.523	1.177	1.104	1.475	3.418	1.614	1.313	1.113	1.073	1.118
0.5	0.5	1.253	1.553	1.565	2.134	1.177	1.100	1.184	1.327	1.232	1.065	1.064	1.162

A Assumptions

Assumption A.1. *(i)* $\rho := (\rho_1', \rho_2')' \in \mathbb{R}^{k_1} \times \mathbb{R}^{k_2}$ *is continuous on the compact parameter set* $\Theta \subset \mathbb{R}^p$ *such that,* $\forall \theta \in \Theta,\ \rho(\theta) = 0 \Leftrightarrow \theta = \theta_0$.

(ii) $\sup_{\theta\in\Theta} \sqrt{n} \left\| \bar{\phi}_n(\theta) - \mathbb{E}_{P_n}(\phi(Y_{in}, \theta)) \right\| = O_{P_n}(1)$, *with* $\bar{\phi}_n(\theta) = n^{-1} \sum_{i=1}^{n} \phi(Y_{in}, \theta)$.

Assumption A.2. *(i)* θ_0 *is interior to* Θ *and* $\phi(y, \theta)$ *is continuously differentiable on* Θ *for* P_n*-almost all* y.

(ii) $\sqrt{n}\bar{\phi}_n(\theta_0) \xrightarrow{d} N(0, \Sigma)$*, under* P_n.

(iii) $\frac{\partial \rho(\theta_0)}{\partial \theta'} = \left(\frac{\partial \rho_1'(\theta_0)}{\partial \theta} \vdots \frac{\partial \rho_2'(\theta_0)}{\partial \theta} \right)'$ *is full column rank and, for* $j = 1, 2$,

$$\mathbb{E}_{P_n}\left(\frac{\partial \phi_j(Y_{in}, \theta_0)}{\partial \theta'} \right) = n^{-\delta_j} \frac{\partial \rho_j(\theta_0)}{\partial \theta'}, \quad \text{and} \quad \sqrt{n} \sup_{\theta\in\mathcal{N}_{\theta_0}} \left\| \frac{\partial \bar{\phi}_{n,j}(\theta)}{\partial \theta'} - \mathbb{E}_{P_n}\left(\frac{\partial \phi_j(Y_{in}, \theta)}{\partial \theta'} \right) \right\| = O_{P_n}(1),$$

where $\mathcal{N}_{\theta_0}$ *is a neighbourhood of* θ_0.

Assumption A.3. *(i)* $\phi_1(y, \theta)$ *is linear in* θ *or* $\delta_2 < \frac{1}{4} + \frac{\delta_1}{2}$.

(ii) $\theta \mapsto \phi(Y_{in}, \theta)$ *is twice continuously differentiable* P_n*-almost everywhere in a neighbourhood* $\mathcal{N}_{\theta_0}$ *of* θ_0 *and, with* $j = 1, 2$,

$$\forall s : 1 \le s \le k_j, \qquad n^{\delta_j} \frac{\partial^2 \bar{\phi}_{n,js}}{\partial \theta \partial \theta'}(\theta) \xrightarrow{P_n} H_{js}(\theta),$$

uniformly over $\mathcal{N}_{\theta_0}$*, where* $H_{js}(\theta)$*'s are* (p, p)*-matrix functions of* θ *and* $\bar{\phi}_{n,js}$ *is the s-th entry of* $\bar{\phi}_{n,j}$.

Assumption A.4. *There exists a neighbourhood* $\mathcal{N}_{\theta_0}$ *of* θ_0 *and a* $L^2(P_n)$*-neighbourhood* $\mathcal{N}_1$ *of* $f_{n,0} := 1$ *such that, with* $j, k = 1, \ldots, p$*, and* $\mathbb{L}_n = \begin{pmatrix} n^{\delta_1} I_{k_1} & 0 \\ 0 & n^{\delta_2} I_{k_2} \end{pmatrix}$,

$$\mathbb{E}_{P_n}\left(\frac{\partial \phi(Y, \theta)}{\partial \theta_j} \right) = \mathbb{L}_n^{-1} \frac{\partial \rho(\theta)}{\partial \theta_j}, \qquad \mathbb{E}_{P_n}\left(\frac{\partial^2 \phi(Y, \theta)}{\partial \theta_j \partial \theta_k} \right) = \mathbb{L}_n^{-1} \frac{\partial^2 \rho(\theta)}{\partial \theta_j \partial \theta_k},\ \forall \theta \in \mathcal{N}_{\theta_0},$$

$$\sup_{\theta\in\mathcal{N}_{\theta_0}} \mathbb{E}_{P_n}\left(\|\phi(Y, \theta)\|^4 \right) = O(1), \qquad \sup_{\theta\in\mathcal{N}_{\theta_0}} \mathbb{E}_{P_n}\left(\|\partial \phi(Y, \theta)/\partial \theta_j\|^4 \right) = O(1),$$

$$\sup_{\theta\in\mathcal{N}_{\theta_0}, f\in\mathcal{N}_1} \int \left\| \frac{\partial \phi(y, \theta)}{\partial \theta_j} \right\|^2 f^2(y) dP_n(y) = O(1), \qquad \sup_{\theta\in\mathcal{N}_{\theta_0}, f\in\mathcal{N}_1} \int \left\| \frac{\partial^2 \phi(y, \theta)}{\partial \theta_j \partial \theta_k} \right\|^2 f^2(y) dP_n(y) = O(1).$$

Assumptions A.1, A.2, and A.3 are useful to establish consistency and asymptotic normality of GMM estimators for moment condition models with mixed identification strength. The second part of Assumption A.3(i) is required for models that are non linear in the parameters and amounts in our setup to Assumption 6*(i) of Antoine and Renault (2012).

Assumption A.4 is not particularly restrictive. It is useful in Proposition 3.2 to control the Lagrange remainder of the first-order Taylor expansion of the semiparametric density functions induced by the moment condition model.

B Propositions and Lemmas

The following proposition further characterizes the family of semiparametric densities functions induced by the moment condition model (1). This proposition follows readily for Lemma 2.2 of Dovonon and Atchadé (2020).

Proposition B.1. *If θ_0 satisfies* (1)*, and Assumption 1 holds with $r = 2$, then there exists a neighborhood $\mathcal{V}$ of (θ_0, h_0) in $\mathbb{R}^p \times \mathcal{E}$, where h_0 denotes the zero element of $\mathcal{E}$, a family $\{f_n(\theta, h, \cdot) : (\theta, h) \in \mathcal{V}\}$ of measurable functions on $\mathbb{R}^k$, such that $f_n(\theta_0, h_0, \cdot) = f_{n,0} := 1$, and for all $(\theta, h) \in \mathcal{V}$,*

$$\int \phi(y, \theta) f_n^2(\theta, h, y) P_n(dy) = 0, \quad \int f_n^2(\theta, h, y) P_n(dy) = 1.$$

Furthermore, the map $(\theta, h) \mapsto f_n(\theta, h, \cdot)$ is differentiable and its first partial derivatives are given by

$$\forall h_1 \in \mathcal{E}, \ \nabla_h f_n(\theta, h, \cdot) \cdot h_1 = h_1 - \langle f_{n,\theta,h} h_1, \bar{\varphi}_\theta \rangle \langle f_{n,\theta,h} \bar{\varphi}, \bar{\varphi}_\theta \rangle^{-1} \bar{\varphi},$$

and

$$\forall w \in \mathbb{R}^p, \quad \nabla_\theta f_n(\theta, h, \cdot) \cdot w = -\frac{1}{2} w' \left\langle f_{n,\theta,h}^2, \nabla_\theta \bar{\varphi}_\theta \right\rangle \langle f_{n,\theta,h} \bar{\varphi}, \bar{\varphi}_\theta \rangle^{-1} \bar{\varphi},$$

$$\text{For } j = 1, \ldots, p, \ \frac{\partial}{\partial \theta_j} f_n(\theta, h, \cdot) = -\frac{1}{2} \left\langle f_{n,\theta,h}^2, \frac{\partial}{\partial \theta_j} \bar{\varphi}_\theta \right\rangle \langle f_{n,\theta,h} \bar{\varphi}, \bar{\varphi}_\theta \rangle^{-1} \bar{\varphi}.$$

In particular, $\nabla_\theta f_n(\theta, h, \cdot)$ evaluated at (θ_0, h_0) is $\nabla_\theta f_n(\theta_0, h_0, \cdot) = -\frac{1}{2} \Gamma_n' \Sigma_n^{-1/2} \varphi$, where

$$\Gamma_n \equiv \mathbb{E}_{P_n} \left(\nabla_\theta \phi(Y, \theta_0) \right).$$

with $f_{n,\theta,h}(\cdot)$ standing for $f_n(\theta, h, \cdot)$. Furthermore, for $i, k = 1 \ldots, p$,

$$\begin{aligned}
\frac{\partial^2}{\partial \theta_k \partial \theta_j} f_n(\theta, h, \cdot) = &-\frac{1}{2} \left\langle f_{n,\theta,h}^2, \frac{\partial^2}{\partial \theta_k \partial \theta_j} \bar{\varphi}_\theta \right\rangle \langle f_{n,\theta,h} \bar{\varphi}, \bar{\varphi}_\theta \rangle^{-1} \bar{\varphi} \\
&- \left\langle \frac{\partial}{\partial \theta_k} f_{n,\theta,h} \cdot f_{n,\theta,h}, \frac{\partial}{\partial \theta_j} \bar{\varphi}_\theta \right\rangle \langle f_{n,\theta,h} \bar{\varphi}, \bar{\varphi}_\theta \rangle^{-1} \bar{\varphi} - \left\langle \frac{\partial}{\partial \theta_j} f_{n,\theta,h} \cdot f_{n,\theta,h}, \frac{\partial}{\partial \theta_k} \bar{\varphi}_\theta \right\rangle \langle f_{n,\theta,h} \bar{\varphi}, \bar{\varphi}_\theta \rangle^{-1} \bar{\varphi} \\
&- \left\langle \frac{\partial}{\partial \theta_j} f_{n,\theta,h} \cdot \frac{\partial}{\partial \theta_k} f_{n,\theta,h}, \bar{\varphi}_\theta \right\rangle \langle f_{n,\theta,h} \bar{\varphi}, \bar{\varphi}_\theta \rangle^{-1} \bar{\varphi}. \quad \text{(B.1)}
\end{aligned}$$

Lemma B.1. *Assume $\mathbf{R} = (\mathbf{R}_1 | \mathbf{R}_2)$ is a different rotation matrix than R but also consistent with the definition in* (14)*, that is, $\mathbf{R}'\mathbf{R} = I_p$ and $\mathbf{R}_2$ spans the null space of $\partial \rho_1(\theta_0)/\partial \theta'$. Then, there exists a (p, p)-matrix A such that: $\mathbf{R} = RA$, $A = \begin{pmatrix} A_1 & 0 \\ 0 & A_2 \end{pmatrix}$ and $A'A = I_k$, where A_1 and A_2 are (s_1, s_1)- and $(p - s_1, p - s_1)$-matrices, respectively.*

Lemma B.2. *Let $\mathbf{r}_2(\theta, h, \cdot)$ be the Lagrange remainder of the first-order Taylor's expansion of $(\theta, h) \mapsto f_n(\theta, h, \cdot)$ around (θ_0, h_0). Let $\{(\theta_n), (h_n)\} \in \Theta(\theta_0) \times \mathcal{C}(h_0)$ and $(\bar{\theta}_n, \bar{h}_n)$ such that:*

$$\bar{\theta}_n = t_n \theta_n + (1 - t_n) \theta_0, \quad \bar{h}_n = t_n h_n + (1 - t_n) h_0, \quad t_n \in (0, 1).$$

Assume θ_0 satisfies (1) *and $(\delta_1, \delta_2) \in \Delta := \{(a, b) \in [0, 1/2[^2 \colon 0 \le a \le b < (1/4 + a/2) \wedge 3/8\}$.*

Then, if Assumptions 1 and A.4 hold with $r = 2$, we have:

$$\sqrt{n} \|\mathbf{r}_2(\bar{\theta}_n, \bar{h}_n, \cdot)\|_{L^2(P_n)} = o(1).$$

C Proofs

Proof of Lemma B.1: Write $R = (R_1 \mid R_2)$. Since $\mathbf{R}_2$ spans the null space of $\partial\rho_1(\theta_0)/\partial\theta'$, then $\mathbf{R}_2 = R_2 A_2$, where A_2 is a $(p-s_1, p-s_1)$-nonsingular matrix. The fact that $\mathbf{R}_2'\mathbf{R}_2 = I_{p-s_1}$ ensures that $A_2'A_2 = I_{p-s_1}$. Also, the fact that $\mathbf{R}_1'\mathbf{R}_2 = 0$ implies that the columns of $\mathbf{R}_1$ lie in the span of the columns of R_2 so that $\mathbf{R}_1 = R_1 A_1$ and we can also claim that $A_1'A_1 = I_{s_1}$ and the result follows. $\square$

Proof of Lemma B.2: We have:

$$f_n(\bar{\theta}_n, \bar{h}_n, \cdot) = f_{n,0} + r_{n,\theta_0} \cdot (\bar{\theta}_n - \theta_0) + A_n \cdot (\bar{h}_n - h_0) + \mathbf{r}_2(\bar{\theta}_n, \bar{h}_n, \cdot),$$

see (24) for the definition of r_{n,θ_0} and A_n. By second-order differentiability of f_n, we have:

$$\|\mathbf{r}_2(\bar{\theta}_n, \bar{h}_n, \cdot)\|_{L^2(P_n)} \leq \Bigg\| \frac{1}{2}\sum_{j,k=1}^{p} \frac{\partial^2}{\partial\theta_j\partial\theta_k} f_n(\bar{\theta}, \bar{h}, \cdot)(\bar{\theta}_{n,j} - \theta_{0,j})(\bar{\theta}_{n,k} - \theta_{0,k}) + O\left(\|\bar{\theta}_n - \theta_0\| \|\bar{h}_n - h_0\|_{L^2(P_n)}\right) + O\left(\|\bar{h}_n - h_0\|^2_{L^2(P_n)}\right)\Bigg\|_{L^2(P_n)}.$$

Thus,

$$\sqrt{n}\|\mathbf{r}_2(\bar{\theta}_n, \bar{h}_n, \cdot)\|_{L^2(P_n)} \leq \frac{\sqrt{n}}{2}\sum_{j,k=1}^{p} \left\|\frac{\partial^2}{\partial\theta_j\partial\theta_k} f_n(\bar{\theta}, \bar{h}, \cdot)\right\|_{L^2(P_n)} \cdot \|\bar{\theta}_n - \theta_0\|^2 + o(1), \tag{C.1}$$

with $\bar{\theta} = t\bar{\theta}_n + (1-t)\theta_0$ and $\bar{h} = t\bar{h}_n + (1-t)h_0$; $t \in (0,1)$. By definition, $\|\bar{h}_n - h_0\|_{L^2(P_n)} \leq \|h_n - h_0\|_{L^2(P_n)} = O(n^{-1/2})$ and $\|\bar{\theta}_n - \theta_0\| \leq \|\theta_n - \theta_0\| = O(n^{-1/2+\delta_2})$ and, the second-order differentiability ensures that: $\|\partial^2 f(\bar{\theta}, \bar{h}, \cdot)/\partial\theta_j\partial\theta_k\|_{L^2(P_n)} = O(1)$ and it follows that:

$$\sqrt{n}\|\mathbf{r}_2(\bar{\theta}_n, \bar{h}_n, \cdot)\|_{L^2(P_n)} = O(n^{-1/2+2\delta_2}).$$

Therefore, if $\delta_2 < 1/4$, $\sqrt{n}\|\mathbf{r}_2(\bar{\theta}_n, \bar{h}_n, \cdot)\|_{L^2(P_n)} = o(1)$.

Consider the case $\delta_2 \geq 1/4$. If

$$\left\|\frac{\partial^2}{\partial\theta_j\partial\theta_k} f_n(\bar{\theta}, \bar{h}, \cdot)\right\|_{L^2(P_n)} = O(n^{-\delta_1} \vee n^{-1+2\delta_2}), \tag{C.2}$$

$$\sqrt{n}\|\mathbf{r}_2(\bar{\theta}_n, \bar{h}_n, \cdot)\|_{L^2(P_n)} = O(n^{-\delta_1} \vee n^{-1+2\delta_2}) \times O(n^{-1/2+2\delta_2}) = O(n^{-\delta_1 - 1/2 + 2\delta_2} \vee n^{-3/2+4\delta_2}) = o(1),$$

for any $(\delta_1, \delta_2) \in \Delta$. To complete the proof, let us establish (C.2). Again, take any $(\bar{\theta}_n, \bar{h}_n)$ convex combination of (θ_n, h_n) and (θ_0, h_0). We have:

$$\begin{aligned} n^{1-2\delta_2}\|f_n(\bar{\theta}_n, \bar{h}_n, \cdot) - f_{n,0}\|_{L^2(P_n)} &\leq n^{1-2\delta_2}\|r_{n,\theta_0}(\bar{\theta}_n - \theta_0) + A_n \cdot (\bar{h}_n - h_0)\|_{L^2(P_n)} + n^{1-2\delta_2}\|\mathbf{r}_2(\bar{\theta}_n, \bar{h}_n, \cdot)\|_{L^2(P_n)} \\ &\leq n^{1-2\delta_2}\|r_{n,\theta_0}(\theta_n - \theta_0) + A_n \cdot (h_n - h_0)\|_{L^2(P_n)} + n^{1-2\delta_2}\|\mathbf{r}_2(\bar{\theta}_n, \bar{h}_n, \cdot)\|_{L^2(P_n)}. \end{aligned}$$

The proof of Proposition 3.2, we establish that $\sqrt{n}\|r_{n,\theta_0}(\theta_n - \theta_0) + A_n \cdot (h_n - h_0)\|_{L^2(P_n)} = O(1)$ and it results that, for $\delta_2 \geq 1/4$,

$$n^{1-2\delta_2}\|r_{n,\theta_0}(\theta_n - \theta_0) + A_n \cdot (h_n - h_0)\|_{L^2(P_n)} = o(1).$$

We can thus claim that

$$\|n^{1-2\delta_2}(f_n(\bar{\theta}_n, \bar{h}_n, \cdot) - f_{n,0})\|_{L^2(P_n)} = O(1), \tag{C.3}$$

which also holds for $(\bar{\theta}, \bar{h})$ as it is also a convex combination of (θ_n, h_n) and (θ_0, h_0). We will use this to establish (C.2).

$\partial^2 f(\theta, h, \cdot)/\partial\theta_j\partial\theta_k$ is given by (B.1) which, using the expression of the first-order derivative of f_n, can also be written:

$$\frac{\partial^2}{\partial\theta_k\partial\theta_j} f_n(\theta, h, \cdot) = -\frac{1}{2}\left\langle f^2_{n,\theta,h}, \frac{\partial^2}{\partial\theta_k\partial\theta_j}\bar{\varphi}_\theta\right\rangle \langle f_{n,\theta,h}\bar{\varphi}, \bar{\varphi}_\theta\rangle^{-1}\bar{\varphi}$$
$$+ \frac{1}{2}\left\langle f^2_{n,\theta,h}, \frac{\partial}{\partial\theta_k}\bar{\varphi}_\theta\right\rangle \langle f_{n,\theta,h}\bar{\varphi}, \bar{\varphi}_\theta\rangle^{-1}\left\langle f_{n,\theta,h}\bar{\varphi}, \frac{\partial}{\partial\theta_j}\bar{\varphi}_\theta\right\rangle \langle f_{n,\theta,h}\bar{\varphi}, \bar{\varphi}_\theta\rangle^{-1}\bar{\varphi}$$
$$+ \frac{1}{2}\left\langle f^2_{n,\theta,h}, \frac{\partial}{\partial\theta_j}\bar{\varphi}_\theta\right\rangle \langle f_{n,\theta,h}\bar{\varphi}, \bar{\varphi}_\theta\rangle^{-1}\left\langle f_{n,\theta,h}\bar{\varphi}, \frac{\partial}{\partial\theta_k}\bar{\varphi}_\theta\right\rangle \langle f_{n,\theta,h}\bar{\varphi}, \bar{\varphi}_\theta\rangle^{-1}\bar{\varphi}$$
$$+ \frac{1}{2}\left\langle f^2_{n,\theta,h}, \frac{\partial}{\partial\theta_j}\bar{\varphi}_\theta\right\rangle \langle f_{n,\theta,h}\bar{\varphi}, \bar{\varphi}_\theta\rangle^{-1}\left\langle \bar{\varphi}\frac{\partial}{\partial\theta_k} f_{n,\theta,h}, \bar{\varphi}_\theta\right\rangle \langle f_{n,\theta,h}\bar{\varphi}, \bar{\varphi}_\theta\rangle^{-1}\bar{\varphi}. \quad \text{(C.4)}$$

We next derive the limits or bounds for each of the inner products involved in this expression but evaluated at $(\bar{\theta}, \bar{h})$. We use the notation $f_n(\theta, h, \cdot)$, $f_{n,\theta,h}$, $f_{\theta,h}$ interchangeably.

(a) Consider: $\langle f_{\bar{\theta},\bar{h}}\bar{\varphi}, \bar{\varphi}_{\bar{\theta}}\rangle$. Note that

$$\bar{\varphi}(y) = \begin{pmatrix} 1 \\ \Sigma_n^{-1/2}\phi(y, \theta_0) \end{pmatrix}, \quad \bar{\varphi}_\theta(y) = \begin{pmatrix} 1 \\ \Sigma_n^{-1/2}\phi(y, \theta) \end{pmatrix}.$$

Hence,

$$\langle f_{\bar{\theta},\bar{h}}\bar{\varphi}, \bar{\varphi}_{\bar{\theta}}\rangle = \int f_{\bar{\theta},\bar{h}}(y)\begin{pmatrix} 1 & \phi(y,\bar{\theta})'\Sigma_n^{-1/2} \\ \Sigma_n^{-1/2}\phi(y,\theta_0) & \Sigma_n^{-1/2}\phi(y,\theta_0)\phi(y,\bar{\theta})'\Sigma_n^{-1/2} \end{pmatrix} dP_n(y).$$

We have: (a.1)

$$\int f_{\bar{\theta},\bar{h}} dP_n = 1 + \int (f_{\bar{\theta},\bar{h}} - 1) dP_n.$$

But, from (C.3),

$$\left|\int (f_{\bar{\theta},\bar{h}} - 1) dP_n\right| \le \left(\int (f_{\bar{\theta},\bar{h}} - 1)^2 dP_n\right)^{1/2} = O(n^{-1+2\delta_2}).$$

Thus:

$$\int f_{\bar{\theta},\bar{h}} dP_n = 1 + O(n^{-1+2\delta_2}).$$

(a.2)

$$\int f_{\bar{\theta},\bar{h}}(y)\phi(y,\theta_0)dP_n(y) = \int \phi(y,\theta_0)dP_n(y) + \int (f_{\bar{\theta},\bar{h}}(y) - 1)\phi(y,\theta_0)dP_n(y) = \int (f_{\bar{\theta},\bar{h}}(y) - 1)\phi(y,\theta_0)dP_n(y).$$

Note that

$$\left|\int (f_{\bar{\theta},\bar{h}}(y) - 1)\phi(y,\theta_0)dP_n(y)\right| \le \left(\int (f_{\bar{\theta},\bar{h}} - 1)^2 dP_n\right)^{1/2}\left(\int \|\phi(y,\theta_0)\|^2 dP_n(y)\right)^{1/2}.$$

Thus,

$$\int f_{\bar{\theta},\bar{h}}(y)\phi(y,\theta_0)dP_n(y) = O(n^{-1+2\delta_2}).$$

(a.3)

$$\int f_{\bar{\theta},\bar{h}}(y)\phi(y,\bar{\theta})dP_n(y) = \int \phi(y,\bar{\theta})dP_n(y) + \int (f_{\bar{\theta},\bar{h}}(y) - 1)\phi(y,\bar{\theta})dP_n(y)$$
$$= O(n^{-\delta_1}) + \int (f_{\bar{\theta},\bar{h}}(y) - 1)\phi(y,\bar{\theta})dP_n(y) = O(n^{-\delta_1}) + O(n^{-1+2\delta_2}).$$

(a.4)

$$\int f_{\bar{\theta},\bar{h}}(y)\phi(y,\theta_0)\phi(y,\bar{\theta})'dP_n(y) = \int \phi(y,\theta_0)\phi(y,\bar{\theta})'dP_n(y) + \int (f_{\bar{\theta},\bar{h}}(y)-1)\phi(y,\theta_0)\phi(y,\bar{\theta})'dP_n(y).$$

Under the conditions of the lemma, $\bar{\theta} \mapsto \int \phi(y,\theta_0)\phi(y,\bar{\theta})'dP_n(y)$ is continuously differentiable in a neighbourhood of θ_0 and we write

$$\int \phi(y,\theta_0)\phi(y,\bar{\theta})'dP_n(y) = \Sigma_n + O(\|\bar{\theta}-\theta_0\|) = \Sigma_n + O(n^{-1/2+\delta_2}).$$

By the Cauchy-Schwarz inequality,

$$\begin{aligned}&\left|\int (f_{\bar{\theta},\bar{h}}(y)-1)\phi(y,\theta_0)\phi(y,\bar{\theta})'dP_n(y)\right| \\ &\qquad \le \left(\int (f_{\bar{\theta},\bar{h}}-1)^2 dP_n\right)^{1/2}\left(\int \|\phi(y,\theta_0)\|^4 dP_n\right)^{1/4}\left(\int \|\phi(y,\bar{\theta})\|^4 dP_n\right)^{1/4} = O(n^{-1+2\delta_2}).\end{aligned}$$

As a result,

$$\langle f_{\bar{\theta},\bar{h}}\bar{\varphi}, \bar{\varphi}_{\bar{\theta}}\rangle = I_{k+1} + o(1)$$

and we can also claim that

$$\langle f_{\bar{\theta},\bar{h}}\bar{\varphi}, \bar{\varphi}_{\bar{\theta}}\rangle^{-1} = I_{k+1} + o(1).$$

(b) Consider: $\left\langle f^2_{\bar{\theta},\bar{h}}, \frac{\partial^2}{\partial\theta_k\partial\theta_j}\bar{\varphi}_{\bar{\theta}}\right\rangle$

$$\left\langle f^2_{\bar{\theta},\bar{h}}, \frac{\partial^2}{\partial\theta_k\partial\theta_j}\bar{\varphi}_{\bar{\theta}}\right\rangle = \int f^2_{\bar{\theta},\bar{h}}(y)\begin{pmatrix} 0 & \frac{\partial^2}{\partial\theta_k\partial\theta_j}\phi(y,\bar{\theta})'\Sigma_n^{-1/2}\end{pmatrix} dP_n(y).$$

We can write:

$$\int f^2_{\bar{\theta},\bar{h}}(y)\frac{\partial^2}{\partial\theta_k\partial\theta_j}\phi(y,\bar{\theta})dP_n(y) = \int \frac{\partial^2}{\partial\theta_k\partial\theta_j}\phi(y,\bar{\theta})dP_n(y) + \int (f^2_{\bar{\theta},\bar{h}}(y)-1)\frac{\partial^2}{\partial\theta_k\partial\theta_j}\phi(y,\bar{\theta})dP_n(y) = (1)+(2).$$

By assumption, $(1) = \mathbb{L}_n^{-1}\frac{\partial^2\rho(\bar{\theta})}{\partial\theta_k\partial\theta_j}$. It follows that $(1) = O(n^{-\delta_1})$. It is not hard to see that

$$\|(2)\| \le \left(\int (f_{\bar{\theta},\bar{h}}-1)^2 dP_n\right)^{1/2}\left(2\sup_{\theta\in\mathcal{N}_\theta, f\in\mathcal{N}_1}\int f^2(y)\|\partial^2\phi(y,\theta)/\partial\theta_k\partial\theta_j\|^2 dP_n(y)\right)^{1/2} = O(n^{-1+2\delta_2}).$$

Thus:

$$\left\|\left\langle f^2_{\bar{\theta},\bar{h}}, \frac{\partial^2}{\partial\theta_k\partial\theta_j}\bar{\varphi}_{\bar{\theta}}\right\rangle\right\| = O(n^{-\delta_1}) + O(n^{-1+2\delta_2}) = O(n^{-\delta_1}\vee n^{-1+2\delta_2}).$$

(c) Consider: $\left\langle f^2_{\bar{\theta},\bar{h}}, \frac{\partial}{\partial\theta_j}\bar{\varphi}_{\bar{\theta}}\right\rangle$

$$\left\langle f^2_{\bar{\theta},\bar{h}}, \frac{\partial}{\partial\theta_j}\bar{\varphi}_{\bar{\theta}}\right\rangle = \int f^2_{\bar{\theta},\bar{h}}(y)\begin{pmatrix} 0 & \frac{\partial}{\partial\theta_j}\phi(y,\bar{\theta})'\Sigma_n^{-1/2}\end{pmatrix} dP_n(y)$$

and we establish as in (b) that the norm of this quantity is of order $O(n^{-\delta_1}\vee n^{-1+2\delta_2})$.

(d) Consider: $\left\langle f_{\bar{\theta},\bar{h}}\bar{\varphi}, \frac{\partial}{\partial\theta_j}\bar{\varphi}_{\bar{\theta}}\right\rangle$.

$$\left\langle f_{\bar{\theta},\bar{h}}\bar{\varphi}, \frac{\partial}{\partial\theta_j}\bar{\varphi}_{\bar{\theta}}\right\rangle = \int f_{\bar{\theta},\bar{h}}(y)\begin{pmatrix} 0 & \frac{\partial\phi(y,\bar{\theta})'}{\partial\theta_j}\Sigma_n^{-1/2} \\ 0 & \Sigma_n^{-1/2}\phi(y,\theta_0)\frac{\partial\phi(y,\bar{\theta})'}{\partial\theta_j}\Sigma_n^{-1/2}\end{pmatrix} dP_n(y).$$

We have:

$$\left\|\int f_{\bar\theta,\bar h}(y)\frac{\partial\phi(y,\bar\theta)'}{\partial\theta_j}dP_n(y)\right\| \le \left\|\int \frac{\partial\phi(y,\bar\theta)'}{\partial\theta_j}dP_n(y)\right\| + \left\|\int (f_{\bar\theta,\bar h}(y)-1)\frac{\partial\phi(y,\bar\theta)'}{\partial\theta_j}dP_n(y)\right\|$$

$$\le \int \left\|\frac{\partial\phi(y,\bar\theta)'}{\partial\theta_j}\right\| dP_n(y) + \left(\int (f_{\bar\theta,\bar h}-1)^2 dP_n\right)^{1/2}\left(\int\left\|\frac{\partial\phi(y,\bar\theta)'}{\partial\theta_j}\right\|^2 dP_n(y)\right)^{1/2} = O(1)+O(n^{-1+2\delta_2}).$$

$$\left\|\int f_{\bar\theta,\bar h}(y)\phi(y,\theta_0)\frac{\partial\phi(y,\bar\theta)'}{\partial\theta_j}dP_n(y)\right\| \le \left\|\int \phi(y,\theta_0)\frac{\partial\phi(y,\bar\theta)'}{\partial\theta_j}dP_n(y)\right\| + \left\|\int (f_{\bar\theta,\bar h}(y)-1)\phi(y,\theta_0)\frac{\partial\phi(y,\bar\theta)'}{\partial\theta_j}dP_n(y)\right\|$$

$$\le \int \|\phi(y,\theta_0)\|\left\|\frac{\partial\phi(y,\bar\theta)'}{\partial\theta_j}\right\| dP_n(y) + \left(\int (f_{\bar\theta,\bar h}-1)^2 dP_n\right)^{1/2}\left(\int \|\phi(y,\theta_0)\|^2\left\|\frac{\partial\phi(y,\bar\theta)'}{\partial\theta_j}\right\|^2 dP_n(y)\right)^{1/2}$$

$$= O(1)+O(n^{-1+2\delta_2}).$$

It follows that $\left\langle f_{\bar\theta,\bar h}\bar\varphi, \frac{\partial}{\partial\theta_j}\bar\varphi_{\bar\theta}\right\rangle = O(1)$.

(e) Consider: $\left\langle \bar\varphi\frac{\partial}{\partial\theta_j}f_{\bar\theta,\bar h}, \bar\varphi_{\bar\theta}\right\rangle$.
We know from Proposition B.1,

$$\frac{\partial}{\partial\theta_j}f_{\bar\theta,\bar h} = -\frac{1}{2}\left\langle f^2_{\bar\theta,\bar h}, \frac{\partial\bar\varphi_{\bar\theta}}{\partial\theta_j}\right\rangle \left\langle f_{\bar\theta,\bar h}\bar\varphi, \bar\varphi_{\bar\theta}\right\rangle^{-1}\bar\varphi := a'\bar\varphi.$$

Hence,

$$\left\langle \bar\varphi\frac{\partial}{\partial\theta_j}f_{\bar\theta,\bar h}, \bar\varphi_{\bar\theta}\right\rangle = \langle a'\bar\varphi\cdot\bar\varphi, \bar\varphi_{\bar\theta}\rangle = \int \begin{pmatrix} a'\bar\varphi & a'\bar\varphi\cdot\phi(y,\bar\theta)\Sigma_n^{-1/2} \\ \Sigma_n^{-1/2}a'\bar\varphi\cdot\phi(y,\theta_0) & \Sigma_n^{-1/2}a'\bar\varphi\cdot\phi(y,\theta_0)\phi(y,\bar\theta)'\Sigma_n^{-1/2}\end{pmatrix} dP_n(y).$$

We have:

$$\left|\int a'\bar\varphi dP_n\right| \le \|a\|\int(1+\|\phi(y,\theta_0)\|)dP_n(y) = O(n^{-\delta_1}\vee n^{-1+2\delta_2}),$$

where we use (c) and (a).

$$\left\|\int a'\bar\varphi\cdot\phi(y,\theta_0)dP_n(y)\right\| \le \|a\|\int(1+\|\phi(y,\theta_0)\|)\,\|\phi(y,\theta_0)\|\, dP_n(y) = O(n^{-\delta_1}\vee n^{-1+2\delta_2}).$$

$$\left\|\int a'\bar\varphi\cdot\phi(y,\bar\theta)dP_n(y)\right\| \le \|a\|\int(1+\|\phi(y,\theta_0)\|)\left\|\phi(y,\bar\theta)\right\| dP_n(y)$$

$$\le \|a\|\left(\int(1+\|\phi(y,\theta_0)\|)^2 dP_n(y)\right)^{1/2}\left(\int\left\|\phi(y,\bar\theta)\right\|^2 dP_n(y)\right)^{1/2} = O(n^{-\delta_1}\vee n^{-1+2\delta_2}).$$

$$\left\|\int a'\bar\varphi\cdot\phi(y,\theta_0)\phi(y,\bar\theta)'dP_n(y)\right\| \le \|a\|\int(1+\|\phi(y,\theta_0)\|)\,\|\phi(y,\theta_0)\|\left\|\phi(y,\bar\theta)\right\| dP_n(y)$$

$$\le \|a\|\left(\int(1+\|\phi(y,\theta_0)\|)^2\|\phi(y,\theta_0)\|^2 dP_n(y)\right)^{1/2}\left(\int\left\|\phi(y,\bar\theta)\right\|^2 dP_n(y)\right)^{1/2} = O(n^{-\delta_1}\vee n^{-1+2\delta_2}).$$

Since the eigenvalues of Σ_n are bounded from above and away from 0, we can claim that $\left\|\left\langle \bar\varphi\frac{\partial}{\partial\theta_j}f_{\bar\theta,\bar h}, \bar\varphi_{\bar\theta}\right\rangle\right\| = O(n^{-\delta_1}\vee n^{-1+2\delta_2})$.

We obtain (C.2) by applying the triangle inequality and then the Cauchy-Schwarz inequality to the terms in (C.4) . Then, the order of magnitude follows from (a), (b), (c), (d), and (e) above. □

Proof of Theorem 3.1:
(i) Writing $\varepsilon_n = \sqrt{n}(g_n - f_n) - \alpha_n$, we have $g_n = f_n + \alpha_n/\sqrt{n} + \varepsilon_n/\sqrt{n}$. Thus:

$$g_n^2 = f_n^2 + \frac{\alpha_n^2}{n} + \frac{\varepsilon_n^2}{n} + 2\frac{\alpha_n f_n}{\sqrt{n}} + 2\frac{\varepsilon_n f_n}{\sqrt{n}} + 2\frac{\alpha_n \varepsilon_n}{n}.$$

Integrating each side with respect to μ_n yields:

$$2\int \alpha_n f_n d\mu_n = -\frac{1}{\sqrt{n}}\int \alpha_n^2 d\mu_n - \frac{1}{\sqrt{n}}\int \varepsilon_n^2 d\mu_n - 2\int \varepsilon_n f_n d\mu_n - \frac{2}{\sqrt{n}}\int \alpha_n \varepsilon_n d\mu_n$$

and the result follows by the Cauchy-Schwarz inequality and the fact that $\int \alpha_n^2 d\mu_n$ is bounded, $\int \varepsilon_n^2 d\mu_n \to 0$ as $n \to \infty$ and $\int f_n^2 d\mu_n = 1$.

(ii) We establish this result by relying on Le Cam's second lemma (see Bickel et al., 1998, Lemma 2, p.500). To obtain the first and second conclusion in (ii), it suffices to show that:

(a) For all $\epsilon > 0$ and as $n \to \infty$,

$$\max_{1\le i\le n} P_{f_n}\left(\left|\frac{g_n^2(X_{ni})}{f_n^2(X_{ni})} - 1\right| > \epsilon\right) \to 0,$$

and (b) Under f_n^2,

$$W_n := 2\sum_{i=1}^{n}\left(\frac{g_n(X_{ni})}{f_n(X_{ni})} - 1\right) \xrightarrow{d} N(-\sigma^2/4, \sigma^2).$$

By the triangle inequality, (18) implies that $\|\sqrt{n}(g_n - f_n)\|_{\mu_n} - \|\alpha_n\|_{\mu_n} \to 0$ as $n \to \infty$ and as a result, $n\|g_n - f_n\|_{\mu_n}^2 \to a^2 \equiv \lim_{n\to\infty} \|\alpha_n\|_{\mu_n}^2$ and $\|g_n - f_n\|_{\mu_n} \to 0$.

To establish (a), pick $\epsilon > 0$; we have:

$$\begin{aligned}\epsilon P_{f_n}\left(\left|\tfrac{g_n^2(X_{ni})}{f_n^2(X_{ni})} - 1\right| > \epsilon\right) &\le \mathbb{E}_{f_n}\left(\left|\tfrac{g_n^2(X_{ni})}{f_n^2(X_{ni})} - 1\right|\right) = \int |g_n^2 - f_n^2| d\mu_n = \int |g_n - f_n||g_n + f_n| d\mu_n \\ &\le \left(\int (g_n - f_n)^2 d\mu_n\right)^{1/2}\left(\int (g_n + f_n)^2 d\mu_n\right)^{1/2}\end{aligned}$$

and the expected result follows since $\int (g_n + f_n)^2 d\mu_n \le 4$.

To establish (b), let

$$Z_n = \frac{2}{\sqrt{n}}\sum_{i=1}^{n}\left(\frac{\alpha_n(X_{ni})}{f_n(X_{ni})} - \nu_n\right),$$

with $\nu_n \equiv \mathbb{E}_{f_n}(\alpha_n(X_{ni})/f_n(X_{ni})) = \int \alpha_n f_n d\mu_n$. We obtain (b) by showing that under f_n^2, Z_n converges in distribution to $N(0, 4a^2)$ and that $\mathbb{E}_{f_n}(W_n - Z_n + a^2)^2 = o(1)$.

Under f_n^2, $\mathbb{E}_{f_n}(\alpha_n(X_{ni})/f_n(X_{ni}) - \nu_n) = 0$ and $Var_{f_n}(\alpha_n(X_{ni})/f_n(X_{ni}) - \nu_n) = \int \alpha_n^2 d\mu_n - \nu_n^2 \to a^2$ as $n \to \infty$. Therefore, the central limit theorem for row-wise independent and identically distributed triangular arrays ensures that under f_n^2,

$$Z_n \xrightarrow{d} N(0, \sigma^2).$$

Next, we observe that $\mathbb{E}_{f_n}(W_n - Z_n + a^2)^2 = \text{Var}_{f_n}(W_n - Z_n + a^2) + [\mathbb{E}_{f_n}(W_n - Z_n + a^2)]^2$. We have:

$$\mathbb{E}_{f_n}(W_n - Z_n + a^2) = \mathbb{E}_{f_n}(W_n) + a^2 = 2n\left(\int g_n f_n d\mu_n - 1\right) + a^2 = -n\int (g_n - f_n)^2 d\mu_n + a^2 \to 0$$

as $n \to \infty$.

$$\begin{aligned}
Var_{f_n}(W_n - Z_n + a^2) &= Var_{f_n}(W_n - Z_n) = 4Var_{f_n}\left(\sum_{i=1}^n \left\{\frac{g_n(X_{ni})}{f_n(X_{ni})} - 1 - \frac{1}{\sqrt{n}}\left(\frac{\alpha_n(X_{ni})}{f_n(X_{ni})} - \nu_n\right)\right\}\right) \\
&= 4\text{Var}_{f_n}\left(\frac{\sqrt{n}(g_n(X_{ni}) - f_n(X_{ni})) - \alpha_n(X_{ni})}{f_n(X_{ni})}\right) \\
&= 4\mathbb{E}_{f_n}\left(\frac{[\sqrt{n}(g_n(X_{ni}) - f_n(X_{ni})) - \alpha_n(X_{ni})]^2}{f_n(X_{ni})^2}\right) - 4\left[\mathbb{E}_{f_n}\left(\frac{\sqrt{n}(g_n(X_{ni}) - f_n(X_{ni})) - \alpha_n(X_{ni})}{f_n(X_{ni})}\right)\right]^2 \\
&= 4\int [\sqrt{n}(g_n - f_n) - \alpha_n]^2 d\mu_n - 4\left(\int [\sqrt{n}(g_n - f_n) - \alpha_n] f_n d\mu_n\right)^2 \\
&\leq 4\|\sqrt{n}(g_n - f_n) - \alpha_n\|^2_{\mu_n} + 4\left(\|\sqrt{n}(g_n - f_n) - \alpha_n\|_{\mu_n}\|f_n\|_{\mu_n}\right)^2 \\
&= 8\|\sqrt{n}(g_n - f_n) - \alpha_n\|^2_{\mu_n} \to 0, \quad \text{as } n \to \infty.
\end{aligned}$$

This establishes (b).

We can therefore apply Le Cam's second lemma and claim that $\log L_n - (W_n - \frac{\sigma^2}{4}) = o_{P_{f_n}}(1)$. Therefore, under f_n^2,

$$\log L_n \xrightarrow{d} N(-\sigma^2/2, \sigma^2)$$

and we can claim using Le Cam's first lemma (see van der Vaart, 1998, p.88) that $\left\{\prod_{i=1}^n g_n^2(x_i)\right\}$ and $\left\{\prod_{i=1}^n f_n^2(x_i)\right\}$ are contiguous. $\square$

Proof of Proposition 3.2: In this proof, we focus only on the case where $(\delta_1, \delta_2) \in \Delta$, $\delta_1 < \delta_2$ and $0 < s_1 < p$. All the other cases follow along the same lines. The Taylor expansion yields (24) with $r_{n,\theta_0}(\cdot) = \nabla_\theta f_n(\theta_0, h_0, \cdot)$ and $A_n = \nabla_h f_n(\theta_0, h_0, \cdot)$. From Proposition B.1, $r_{n,\theta_0}(\cdot) = -\frac{1}{2}\Gamma_n' \Sigma_n^{-1/2}\varphi(\cdot)$, with $\Gamma_n = \mathbb{E}_{P_n}\left(\frac{\partial}{\partial\theta'}\phi(Y, \theta_0)\right)$. This proposition also gives:

$$\forall u \in \mathcal{E},\ \nabla_h f_n(\theta, h, \cdot) \cdot u = u - \langle f_{n,\theta,h} u, \bar{\varphi}_\theta\rangle \langle f_{n,\theta,h}\bar{\varphi}, \bar{\varphi}_\theta\rangle^{-1} \bar{\varphi},$$

with $f_{n,\theta,h} := f_n(\theta, h, \cdot)$. At (θ_0, h_0), $\langle f_{n,\theta,h} u, \bar{\varphi}_\theta\rangle = \langle u, \bar{\varphi}\rangle = 0$, since $u \in \mathcal{E}$. Hence, $\nabla_h f_n(\theta_0, h_0, \cdot) \cdot u = u$. It follows that, since $h_n, h_0 \in \mathcal{E}$,

$$\nabla_h f_n(\theta_0, h_0, \cdot) \cdot (h_n - h_0) = h_n - h_0.$$

Recall that θ_n and h_n are defined such that: $\Lambda_n R^{-1}(\theta_n - \theta_0) - \eta \to 0$ and $\sqrt{n}(h_n - h_0) - \beta \to 0$ in $L^2(P_n)$ (see Equations (20) and (22)). For $(\delta_1, \delta_2) \in \Delta$, according to the discussion leading to the statement of the proposition, we need to find $\alpha_n \in L^2(P_n)$ such that

$$\|\alpha_n - \sqrt{n}[r_{n,\theta_0} \cdot (\theta_n - \theta_0) + A_n \cdot (h_n - h_0)]\|_{L^2(P_n)} \to 0.$$

It is obvious that $\|\sqrt{n}A_n \cdot (h_n - h_0) - A_n \cdot \beta\|_{L^2(P_n)} = \|\sqrt{n}(h_n - h_0) - \beta\|_{L^2(P_n)} \to 0$, as $n \to \infty$.

Also,

$$r_{n,\theta_0} \cdot (\theta_n - \theta_0) = -\frac{1}{2}(\theta_n - \theta_0)'\Gamma_n'\Sigma_n^{-1/2}\varphi$$

and

$$\Gamma_n(\theta_n-\theta_0)=\left(\begin{array}{c} n^{-\delta_1}\frac{\partial\rho_1(\theta_0)}{\partial\theta'} \\ n^{-\delta_2}\frac{\partial\rho_2(\theta_0)}{\partial\theta'}\end{array}\right)R\Lambda_n^{-1}\Lambda_nR^{-1}(\theta_n-\theta_0)=n^{-1/2}\left(\begin{array}{cc} D_1R_1 & 0 \\ n^{\delta_1-\delta_2}D_2R_1 & D_2R_2\end{array}\right)\Lambda_nR^{-1}(\theta_n-\theta_0),$$

with $D_j=\frac{\partial\rho_j(\theta_0)}{\partial\theta'}$. (We use the fact that $D_1R_2=0$.) Hence,

$$\sqrt{n}\Gamma_n(\theta_n-\theta_0)=J\eta+o(1).$$

As a result, we can set

$$\alpha_n(\cdot)=-\frac{1}{2}\eta'J'\Sigma_n^{-1/2}\varphi(\cdot)+A_n\cdot\beta,\quad \eta\in\mathbb{R}^p,\ \beta\in\mathcal{E}.\quad\square$$

Proof of Theorem 3.3: This proof follows similar lines to that of Theorem 4.4 in Dovonon and Atchadé (2020). Let $S_n=\Lambda_nR^{-1}(\tilde{\theta}_n-\theta_n)$. The characteristic function of S_n under g_n^2 is

$$\begin{aligned}\mathbb{E}_{g_n}\left[\exp(iw'S_n)\right] &= \mathbb{E}_{g_n}\left[\exp(iw'\Lambda_nR^{-1}(\tilde{\theta}_n-\theta_n))\right]\\ &= \mathbb{E}_{g_n}\left[\exp(iw'\Lambda_nR^{-1}(\tilde{\theta}_n-\theta_0-(\theta_n-\theta_0)))\right]\\ &= \mathbb{E}_{g_n}\left[\exp(iw'\Lambda_nR^{-1}(\tilde{\theta}_n-\theta_0))\exp(-iw'(\eta+\varepsilon_n))\right],\end{aligned}$$

for some $\eta\in\mathbb{R}^p$ and $\varepsilon_n:=\Lambda_nR^{-1}(\theta_n-\theta_0)-\eta$ which tends to 0 as $n\to 0$. Thus,

$$\begin{aligned}\mathbb{E}_{g_n}\left[\exp(iw'S_n)\right] &= \mathbb{E}_{g_n}\left[\exp(iw'\Lambda_nR^{-1}(\tilde{\theta}_n-\theta_0))\exp(-iw'\eta)\right]+o(1)\\ &= \mathbb{E}_{f_{n,0}}\left[\exp(iw'\Lambda_nR^{-1}(\tilde{\theta}_n-\theta_0)-iw'\eta+L_n)\right]+o(1).\end{aligned}$$

This holds for any sequence $\{g_n^2(\cdot)\}$ associated to any $\alpha_n=-\frac{1}{2}\eta'J'\Sigma_n^{-1/2}\varphi-A_n\cdot\mathbf{b}$, with $\mathbf{b}=(\beta_1,\dots,\beta_p)\in\mathcal{B}(h_0)^p$ (where "associated" is meant in the sense described by Equation (19)). In particular, this holds for:

$$\alpha_n=\eta'\left(-\frac{1}{2}J'\Sigma_n^{-1/2}\varphi-A_n\cdot\mathbf{b}_n^*\right).$$

Thanks to Theorem 3.1, under $f_{n,0}^2$, $\left(\Lambda_nR^{-1}(\tilde{\theta}_n-\theta_0),\ \frac{2}{\sqrt{n}}\sum_{i=1}^n\left(\frac{\alpha_n(Y_{ni})}{f_n(Y_{ni})}-\nu_n\right)\right)$ converges in distribution coordinate-wise to $(S,\eta'Z_0)$, with: $\nu_n=\mathbb{E}_{f_{n,0}}(\alpha_n(Y_{ni})/f_{n,0}(Y_{ni}))$, $Z_0\sim N(0,I_*)$, and

$$I_*=4\lim_{n\to\infty}\left\langle -\frac{1}{2}J'\Sigma_n^{-1/2}\varphi-A_n\cdot\mathbf{b}_n^*,-\frac{1}{2}J'\Sigma_n^{-1/2}\varphi-A_n\cdot\mathbf{b}_n^*\right\rangle.$$

Therefore, by the Prohorov's theorem, there is a subsequence of $\left(\Lambda_nR^{-1}(\tilde{\theta}_n-\theta_0),L_n\right)$ that converges weakly under $f_{n,0}^2$ to $\left(S,\eta'Z_0-\frac{1}{2}\eta'I_*\eta\right)$. Along that subsequence, we can claim that:

$$\begin{aligned}&\mathbb{E}_{f_{n,0}}\exp\left(iw'\Lambda_nR^{-1}(\tilde{\theta}_n-\theta_0)-iw'\eta+L_n\right)\\ \to\ &\mathbb{E}\exp\left(iw'S-iw'\eta+\eta'Z_0-\tfrac{1}{2}\eta'I_*\eta\right)=\mathbb{E}\exp\left[iw'S+\eta'Z_0\right]\exp[-iw'\eta-\tfrac{1}{2}\eta'I_*\eta].\end{aligned}\tag{C.5}$$

Also, $\tilde{\theta}_n$ being a Λ_n-regular estimator ensures that

$$\mathbb{E}_{g_n}\exp\left[iw'\Lambda_nR^{-1}(\tilde{\theta}_n-\theta_n)\right]\to\mathbb{E}\exp(iw'S).\tag{C.6}$$

Letting $\Phi(w,v) = \mathbb{E}\exp(iw'S + iv'Z_0)$, we have

$$\Phi(w,0) = \mathbb{E}\left[\exp(iw'S + \eta' Z_0)\right]\exp\left[-iw'\eta - \frac{1}{2}\eta' I_* \eta\right].$$

The right-hand-side of this equality is analytic in η and constant on $\mathbb{R}^p$. As a result, it is constant for $\eta \in \mathbb{C}^p$. Now, choosing $\eta = -iI_*^{-1}w$, we have:

$$\Phi(w,0) = \mathbb{E}\exp\left[iw'(S - I_*^{-1}Z_0)\right]\exp\left[-\frac{1}{2}w' I_*^{-1} w\right]. \tag{C.7}$$

One can recognize in (C.7), the product of the characteristic functions of $U = S - Z$ and Z with $Z = I_*^{-1}Z_0 \sim N(0, I_*^{-1})$ independent of U. This concludes the proof. $\square$

Proof of Corollary 3.4: Under the conditions of the Corollary, the conditions of Proposition 3.2 and Theorem 3.3 are satisfied. As a result, (29) holds. It remains to show that $I_* = J'\Sigma^{-1}J$. From Theorem 3.3, $I_* = 4\langle -\frac{1}{2}J'\Sigma_n^{-1/2}\varphi - A_n \cdot \mathbf{b}_n^*, -\frac{1}{2}J'\Sigma_n^{-1/2}\varphi - A_n \cdot \mathbf{b}_n^*\rangle$, with $\mathbf{b}_n^* \in \mathcal{B}(h_0)^p$, and $A_n \cdot \mathbf{b}_n^*$ the orthogonal projection of $-\frac{1}{2}J'\varphi$ onto $\{A_n \cdot \mathbf{b} : \mathbf{b} \in \mathcal{B}(h_0)^p\}$. Recall that $\mathcal{B}(h_0)$ is a subspace of $\mathcal{E}$. Hence from Proposition B.1, along with simple derivations, we have that, for any $\beta \in \mathcal{B}(h_0)$,

$$A_n \cdot \beta := \nabla_h f_n(\theta_0, h_0, \cdot) \cdot \beta = \beta = \sum_{j \geq k+2} a_j \varphi_j,$$

where for $j \geq k+2$, $a_j = \langle \beta, \varphi_j \rangle = \int \beta \varphi_j dP_n$. The last equality follows from the fact that $\beta \in \mathcal{E}$. Hence, $A_n \cdot \beta$ is orthogonal to φ for any $\beta \in \mathcal{B}(h_0)$. Thus $\mathbf{b}_n^* = 0$ and

$$I_* = 4\lim_{n\to\infty}\left\langle \frac{1}{2}J'\Sigma_n^{-1/2}\varphi, \frac{1}{2}J'\Sigma_n^{-1/2}\varphi \right\rangle = \lim_{n\to\infty} J'\Sigma_n^{-1/2}\int \varphi\varphi' dP_n \Sigma_n^{-1/2}J = \lim_{n\to\infty} J'\Sigma_n^{-1}J = J'\Sigma^{-1}J. \quad \square$$

Proof of Proposition 3.6: Note that since, from (15) $\hat{\theta}_n - \theta_0 = O_P(n^{-1/2+\delta_2})$ under P_n, this also holds under g_n^2. By the definition, we also have $\theta_n - \theta_0 = O(n^{-1/2+\delta_2})$ so that $\hat{\theta}_n - \theta_n = O_P(n^{-1/2+\delta_2})$ under g_n^2.

The first order optimality condition for GMM is given by:

$$\frac{\partial\bar{\phi}_n(\hat{\theta}_n)'}{\partial\theta}\hat{W}\bar{\phi}_n(\hat{\theta}_n) = 0.$$

By the mean-value expansion, we write

$$\frac{\partial\bar{\phi}_n(\hat{\theta}_n)'}{\partial\theta}\hat{W}\bar{\phi}_n(\theta_n) + \frac{\partial\bar{\phi}_n(\hat{\theta}_n)'}{\partial\theta}\hat{W}\frac{\partial\bar{\phi}_n(\bar{\theta}_n)}{\partial\theta'}(\hat{\theta}_n - \theta_n) = 0, \tag{C.8}$$

where $\bar{\theta}_n \in (\hat{\theta}_n, \theta_n)$ and may differ from row to row.

From Lemma A.5 of Antoine and Renault (2009), we can claim that:

$$\sqrt{n}\frac{\partial\bar{\phi}_n(\hat{\theta}_n)}{\partial\theta'}R\Lambda_n^{-1} \xrightarrow{P} J \quad \text{and} \quad \sqrt{n}\frac{\partial\bar{\phi}_n(\bar{\theta}_n)}{\partial\theta'}R\Lambda_n^{-1} \xrightarrow{P} J, \tag{C.9}$$

both under P_n and g_n^2. Also, $\hat{W} - W \xrightarrow{P} 0$, under g_n^2. It follows that (recall that $R' = R^{-1}$):

$$n\Lambda_n^{-1}R^{-1}\frac{\partial\bar{\phi}_n(\hat{\theta}_n)'}{\partial\theta}\hat{W}\frac{\partial\bar{\phi}_n(\bar{\theta}_n)}{\partial\theta'}R\Lambda_n^{-1} \xrightarrow{P} J'WJ, \quad \text{under } g_n^2. \tag{C.10}$$

Next, we show that $\sqrt{n}\bar{\phi}_n(\theta_n) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \phi(Y_{ni}, \theta_n)$ converges in distribution to $N(0,\Sigma)$ under g_n^2. By construction, $\mathcal{M}(\theta_n, h_n, g_n) = 0$. Hence, using (16), we get $\langle g_n^2, \varphi_{\theta_n}\rangle = 0$. That is

$$\int g_n^2(y)\phi(y,\theta_n)'\Sigma_n^{-1/2}dP_n(y) = 0,$$

which implies that $\int g_n^2(y)\phi(y,\theta_n)dP_n(y)=0$, that is

$$\mathbb{E}_{g_n}\left(\phi(Y_{ni},\theta_n)\right)=0.$$

Also,

$$\begin{aligned}Var_{g_n}(\phi(Y_{ni},\theta_n)) &= \mathbb{E}_{g_n}\left(\phi(Y_{ni},\theta_n)\phi(Y_{ni},\theta_n)'\right)\\ &= \int \phi(y,\theta_n)\phi(y,\theta_n)'(g_n^2(y)-1)dP_n(y)+\int \phi(y,\theta_n)\phi(y,\theta_n)'dP_n(y).\end{aligned}$$

By Assumption 2(b), $\int \phi(y,\theta_n)\phi(y,\theta_n)'dP_n(y)=\Sigma+o(1)$. Note that:

$$\begin{aligned}\left\|\int \phi(y,\theta_n)\phi(y,\theta_n)'(g_n^2(y)-1)dP_n(y)\right\| &\leq \int \|\phi(y,\theta_n)\phi(y,\theta_n)'\|\,|g_n^2(y)-1|dP_n(y)\\ &\leq \left(\int (g_n-1)^2dP_n\int \|\phi(y,\theta_n)\|^4(g_n(y)+1)^2dP_n(y)\right)^{1/2}\\ &\leq \|g_n-1\|_{L^2(P_n)}\cdot\left(2\int \|\phi(y,\theta_n)\|^4g_n^2(y)dP_n(y)+2\int\|\phi(y,\theta_n)\|^4dP_n(y)\right)^{1/2}\\ &\leq 2\sqrt{C}\|g_n-1\|_{L^2(P_n)}:=2\sqrt{C}\|g_n-f_{n,0}\|_{L^2(P_n)}=o(1),\end{aligned}$$

where the second inequality follows from the Cauchy-Schwarz inequality and the last one follows from Assumption 2(a). Thus, $Var_{g_n}(\phi(Y_{ni},\theta_n))\to\Sigma$, as $n\to\infty$. The central limit theorem for row-wise independent and identically distributed triangular arrays ensures that:

$$\sqrt{n}\bar{\phi}_n(\theta_n)\xrightarrow{d}N(0,\Sigma),\quad \text{under } g_n^2. \tag{C.11}$$

We write (C.8) as:

$$\sqrt{n}\Lambda_n^{-1}R^{-1}\frac{\partial\bar{\phi}_n(\hat{\theta}_n)'}{\partial\theta}\hat{W}\sqrt{n}\bar{\phi}_n(\theta_n)+\sqrt{n}\Lambda_n^{-1}R^{-1}\frac{\partial\bar{\phi}_n(\hat{\theta}_n)'}{\partial\theta}\hat{W}\frac{\partial\bar{\phi}_n(\bar{\theta}_n)}{\partial\theta'}R\Lambda_n^{-1}\sqrt{n}\left[\Lambda_nR^{-1}(\hat{\theta}_n-\theta_n)\right]=0.$$

Using (C.9) and (C.10), this yields:

$$J'W\sqrt{n}\bar{\phi}_n(\theta_n)+J'WJ\left[\Lambda_nR^{-1}(\hat{\theta}_n-\theta_n)\right]=o_P(1),$$

that is:

$$\Lambda_nR^{-1}(\hat{\theta}_n-\theta_n)=-(J'WJ)^{-1}J'W\sqrt{n}\bar{\phi}_n(\theta_n)+o_P(1),$$

where the $o_P(1)$ is under g_n^2. Using (C.11), we conclude that

$$\Lambda_nR^{-1}(\hat{\theta}_n-\theta_n)\xrightarrow{d}N(0,\Omega(W)),$$

under g_n^2 and we claim that $\hat{\theta}_n$ is Λ_n-regular. $\square$

References

Andrews, D. W. K. and Cheng, X. (2012). Estimation and inference with weak, semi-strong, and strong identification. *Econometrica*, 80(5):2153–2211. https://doi.org/10.3982/ECTA9456.

Andrews, I. and Mikusheva, A. (2022a). GMM is inadmissible under weak identification. Technical report, MIT.

Andrews, I. and Mikusheva, A. (2022b). Optimal decision rules for weak GMM. *Econometrica*, 90(2):715–748. https://doi.org/10.3982/ECTA18678.

Antoine, B. and Renault, E. (2009). Efficient GMM with nearly-weak instruments. *The Econometrics Journal*, 12:S135–S171. doi: 10.1111/j.l 368-423X.2009.00279.X.

Antoine, B. and Renault, E. (2012). Efficient minimum distance estimation with multiple rates of convergence. *Journal of Econometrics*, 170(2):350–367. https://doi.org/10.1016/j.jeconom.2012.05.010.

Antoine, B. and Renault, E. (2020). Testing identification strength. *Journal of Econometrics*, 218:271–293. https://doi.org/10.1016/j.jeconom.2020.04.017.

Antoine, B. and Renault, E. (2021). GMM with nearly-weak identification. *Econometrics and Statistics*. https://doi.org/10.1016/j.ecosta.2021.10.010. To appear.

Begun, J. M., Hall, W. J., Huang, W.-M., and Wellner, J. A. (1983). Information and asymptotic efficiency in parametric-nonparametric models. *The Annals of Statistics*, 11(2):432–452. doi: 10.1214/aos/1176346151.

Bickel, P., Klaassen, C., Ritov, Y., and Wellner, J. (1998). *Efficient and adaptive estimation for semiparametric models*. New York : Springer.

Caner, M. (2009). Testing, estimation in GMM and CUE with nearly-weak identification. *Econometric Reviews*, 29(3):330–363. doi: 10.1080/07474930903451599.

Chamberlain, G. (1987). Asymptotic efficiency in estimation with conditional moment restrictions. *Journal of Econometrics*, 34(3):305–334. https://doi.org/10.1016/0304-4076(87)90015-7.

Dovonon, P. and Atchadé, Y. F. (2020). Efficiency bounds for semiparametric models with singular score functions. *Econometric Reviews*, 39(6):612–648. doi: 10.1080/07474938.2019.1701809.

Dovonon, P., Doko Tchatoka, F., and Aguessy, M. (2023). Relevant moment selection under mixed identification strength. *Econometric Theory*, pages 1–62. doi: 10.1017/S0266466622000640. To appear.

Hahn, J. and Kuersteiner, G. (2002). Discontinuities of weak instrument limiting distributions. *Economics Letters*, 75(3):325–331.

Hallin, M., van den Akker, R., and Werker, B. J. M. (2015). *On Quadratic Expansions of Log-Likelihoods and a General Asymptotic Linearity Result*, pages 147–165. Springer International Publishing, Cham. doi: 10.1007/978-3-319-12442-1_9.

Han, S. and McCloskey, A. (2019). Estimation and inference with a (nearly) singular jacobian. *Quantitative Economics*, 10(3):1019–1068. https://doi.org/10.3982/QE989.

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica*, 50(4):1029–1054. https://doi.org/10.2307/1912775.

Kaji, T. (2021). Theory of weak identification in semiparametric models. *Econometrica*, 89(2):733–763. https://doi.org/10.3982/ECTA16413.

Magnus, J. R. and Neudecker, H. (2002). *Matrix differential calculus with applications in statistics and econometrics*. Wiley.

Phillips, P. C. B. (1980). The exact distribution of instrumental variable estimators in an equation containing $n+1$ endogenous variables. *Econometrica*, 48(4):861–878. https://doi.org/10.2307/1912937.

Staiger, D. and Stock, J. (1997). Instrumental variables regression with weak instruments. *Econometrica*, 65(3):557–586. https://doi.org/10.2307/2171753.

Stock, J. H. and Wright, J. H. (2000). GMM with weak identification. *Econometrica*, 68(5):1055–1096. https://doi.org/10.1111/1468-0262.00151.

van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press. https://doi.org/10.1017/CBO9780511802256.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak convergence and empirical processes with applications to statistics*. Springer Series in Statistics. Springer New York. https://doi.org/10.1007/978-1-4757-2545-2.

Wilks, S. S. (1932). Certain generalizations in the analysis of variance. *Biometrika*, 24(3-4):471–494. https://doi.org/10.2307/2331979.