Specification Testing for Conditional Moment Restrictions under Local Identification Failure*

Prosper Dovonon† and Nikolay Gospodinov‡

January 31, 2022

Abstract

In this paper, we study the asymptotic behavior of the specification test in conditional moment restrictions model under first-order local identification failure with dependent data. More specifically, we obtain conditions under which the conventional specification test for conditional moment restrictions remains valid when first-order local identification fails but global identification is still attainable. In the process, we obtain some novel intermediate results that include extending the first- and second-order local identification framework to models defined by conditional moment restrictions, characterizing the rate of convergence of the GMM estimator and the limiting representation for degenerate U-statistics under strong mixing dependence. Simulation and empirical results illustrate the properties and the practical relevance of the proposed testing framework.

Keywords: GMM, conditional moment restrictions, test for overidentifying restrictions, local and global identification, first-order local identification failure, second-order local identification, U-statistics, strong mixing dependence.

JEL classification: C01, C1, C14, G12.

---

*The views expressed here are those of the authors and do not necessarily reflect those of the Federal Reserve Bank of Atlanta or the Federal Reserve System.

†Department of Economics, Concordia University, 1455 de Maisonneuve Blvd. West, H-1155, Montreal, Quebec, Canada, H3G 1M8. Email: prosper.dovonon@concordia.ca.

‡Research Department, Federal Reserve Bank of Atlanta, 1000 Peachtree Street N.E., Atlanta, GA 30309-4470. Email: nikolay.gospodinov@atl.frb.org.

1 Introduction

While economic models are designed to be only partial and incomplete representations of real economic phenomena, it is still highly desirable to quantify the degree of model misspecification and the directions along which the model performance is unsatisfactory. Even if the model is rejected by the data, it can still be useful for policy analysis but the standard inference and model comparison procedures should be adjusted for the underlying model uncertainty. For these reasons, it is now common practice to first subject the candidate models to specification testing before committing to a particular analytical and inference framework.

There are at least two characteristics of economic models that make the development of fully robust and reliable specification testing procedures more challenging. First, economic models are typically defined by a set of *conditional* moment restrictions. The routine approach is to resort to the law of iterated expectations and reduce the conditional restrictions to unconditional moment restrictions that are then used to design the proper estimation and testing framework. When this is done in ad hoc manner, this approach could result in loss of efficiency and even in inconsistency of the estimator (see, for example, Dominguez and Lobato, 2004). On the other hand, a transformation that preserves the information in the conditional moment restrictions leads to modified tests based on a continuum of moment conditions (see Bierens, 1982; Bierens and Ploberger, 1987; de Jong and Bierens, 1994; Carrasco and Florens, 2000; Kitamura, Tripathi and Ahn, 2004; among others). A common feature of all these tests is that they rely on root-n consistent estimators which are readily available in models that are first-order locally identified.

Second, it is often the case that the moment restriction model is *locally under-identified.* In linear models, for example, the lack of first-order local identification – rank deficiency of the Jacobian matrix of the moment conditions – implies global identification failure which typically renders the standard specification tests invalid under both the null and alternative hypotheses as the power of the test, in certain contexts, is bounded by its size (Gospodinov *et al.*, 2017). The intuition behind this result is that it is sometimes possible to recast the optimal specification test as a reduced rank test which highlights the difficulty of determining if the reduced rank is induced by correct specification or identification failure. In nonlinear models, however, first-order identification is no longer a necessary condition for global identification. There are prominent examples in which the model fails the first-order local identification property but identifies the true value locally at second order. Such examples include models with common conditionally heteroskedastic features (Dovonon and Renault, 2013), nonstationary panel AR(1) model with individual fixed effects (Dovonon and Hall, 2018; Dovonon *et al.*, 2020), and unit root moving average Gaussian models with higher-moment overidentifying restrictions (Gospodinov and Ng, 2015). The implications of the second-order local identification for unconditional moment restriction models are a slower rate of convergence of the estimator and overrejection of the standard specification test (Dovonon and Renault, 2013).

This paper builds on these two strands of literature to obtain conditions under which the conventional specification tests with conditional moment conditions remain valid under first-order local identification failure. To this end, we formalize the concepts of point identification and first-order local identification failure in conditional moment restriction models. Similarly to point-identified unconditional models, the first-order local identification failure allows only for a reduced number of directions of the parameter vector to be identified while identification of the remaining directions is obtained via a second-order expansion of the moment conditions. We then proceed with characterizing the limiting behavior of the estimator and the specification test in models with an expanding set of moment conditions when first-order local identification fails but global identification is still attainable.

Our main contributions can be summarized as follows. First, we extend the test for validity of conditional moment restrictions (de Jong and Bierens, 1994; Donald *et al.*, 2003) to moment condition models that are first-order degenerate. We establish our results in a two-step generalized method of moments (GMM) framework with general forms of moment condition models and dependent data. By contrast, the limiting behavior of the test proposed by de Jong and Bierens (1994) is obtained in the context of nonlinear regression models with cross-sectional data. We should note that the extension to dependent data and characterizing the limiting behavior of the GMM estimator and the specification test in this context is non-trivial. We outline the conditions under which the specification test with an increasing number of unconditional moment restrictions is robust to the type of singularity arising from first-order local identification failure. More specifically, we extend the notion of second-order local identification to the setting of models defined with conditional moment restrictions. The limiting behavior of the GMM estimator and the specification test are studied under the identification pattern where point identification hold, first-order local identification fails while local identification is maintained at second-order.

We show that the GMM estimator based on the expanding moment restrictions estimate the directions of the parameters that are locally identified at first order at a standard $\sqrt{n}$-rate while the remaining directions are estimated at a slower rate. Interestingly, this rate is faster than the $n^{1/4}$-rate in second-order identified models with a fixed number of moment restrictions (Dovonon and Renault, 2020). In this paper, the expanding number of moment restrictions enhances the identification signal and accelerates the rate of convergence. We also derive the asymptotic distribution of the GMM estimator in the scalar case which highlights the highly non-standard limiting behavior of the estimator. Despite this non-standard asymptotic setup, we show that the test for validity of conditional moment restrictions is characterized by a standard normal limit even when the first-order local identification condition is compromised. Another important intermediate result that we develop in the paper is a central limit theorem (CLT) for degenerate U-statistics with linear kernels of increasing dimension under strong mixing dependence. The CLT is novel and of independent interest. Establishing the asymptotic normality of the test for overidentifying restrictions draws heavily on this CLT.

The rest of the paper is structured as follows. Section 2 introduces the main setup and notation. It also presents the notions of first- and second-order local identification along with alternative characterizations in the context of conditional moment restrictions. Section 2 ends with two motivating examples. The asymptotic properties of the GMM estimator are studied in Section 3. Section 4 proposes a CLT for U-statistics under strong mixing dependence and establishes the asymptotic normality of the specification test statistic under the null hypothesis. This section also shows that this test is consistent against all alternatives. Section 5 reports simulation results for the empirical size and power properties of the proposed specification test. This section also provides an empirical application that illustrates the presence of common conditionally heteroskedastic features in portfolio bond returns. Section 6 concludes. The proofs and some supplementary results are provided in Appendices A and B, and the Online Appendix.

Throughout the paper, for any matrix C, $\|C\|_2 = \sqrt{\lambda_{\max}(CC')}$ denotes the spectral norm, with $\lambda_{\max}(\cdot)$ the largest eigenvalue function. If C is a vector, this amounts to its Euclidean norm as well. Also, let $\lambda_{\min}(\cdot)$ denote the smallest eigenvalue function, and $\mathbb{Z}$, $\mathbb{N}$ and $\mathbb{R}^m$ signify the set of all integers, the set of natural numbers and the set of real $m \times 1$ vectors, respectively. Furthermore, $\text{Card}(S)$ denotes the cardinality of a finite set S, defined to be the number of elements in the set S, $\text{vec}(C)$ signifies column vectorization of a matrix C, $\vee$ denotes maximum, $\text{Rank}(C)$ is the column rank of a matrix C, and $\text{Diag}(c_{11}, c_{22}, ..., c_{mm})$ denotes an $m \times m$ diagonal matrix with $(c_{11}, c_{22}, ..., c_{mm})'$ on its main diagonal. Convergence in probability and convergence in distribution are denoted by $\xrightarrow{P}$ and $\xrightarrow{d}$, respectively, while the abbreviation *a.s.* stands for almost surely. Let $\{X_t : t \in \mathbb{Z}\}$ be a sequence of random variables and $\mathcal{F}_a^b$ be the σ-algebra generated by $\{X_t : -\infty \leq a \leq t \leq b \leq \infty\}$. Then, $\{X_t\}$

is said to be strong mixing or α-mixing (Andrews, 1984) if

$$\sup_{-\infty<t<\infty} \sup_{A\in\mathcal{F}_{-\infty}^{t}, B\in\mathcal{F}_{t+s}^{\infty}} |\Pr(A\cap B) - \Pr(A)\Pr(B)| = \alpha(s) \to 0 \text{ as } s\to\infty.$$

Finally, $a_n = o_P(1)$ denotes that the sequence a_n tends to zero in probability and $a_n = O_P(1)$ signifies that a_n is bounded in probability.

2 Model, identification setup and examples

In this paper, we consider a single conditional moment restriction model:

$$\mathbb{E}\left(u(y_t,\theta_0)|x_t\right) = 0 \quad a.s., \tag{1}$$

where u is a real-valued function, $\theta_0 \in \Theta \subset \mathbb{R}^p$ is the parameter of interest and $\{(x_t, y_t)\}_t$ is a sequence of $\mathbb{R}^{k_x} \times \mathbb{R}^{k_y}$-valued random vectors. Many economic equilibrium models take this conditional moment restriction form. A prominent example of the role of conditioning is the stochastic discount factor framework in asset pricing (see, for instance, Hansen, 2014).[1] In this setup, the null hypothesis of validity of the conditional moment restrictions in (1) is

$$H_0 : \Pr\{\mathbb{E}\left(u(y_t,\theta_0)|x_t\right) = 0\} = 1 \tag{2}$$

against the alternative

$$H_1 : \Pr\{\mathbb{E}\left(u(y_t,\theta)|x_t\right) = 0\} < 1, \text{ for any } \theta\in\Theta.$$

While the single conditional restriction setup covers a wide range of practically relevant models, we focus on this case merely for the sake of notational simplicity. The main results in this paper carry over to higher-dimensional conditional moment restrictions at the cost of more cumbersome notation.

Consistent estimation of θ_0 using (1) requires point identification; that is, for all $\theta \in \Theta$,

$$\rho(x_t,\theta) := \mathbb{E}(u(y_t,\theta)|x_t) = 0, a.s. \Leftrightarrow \theta = \theta_0. \tag{3}$$

Moreover, inference about θ_0 hinges on the sharpness of the slope of the function $\theta \mapsto \rho(x,\theta)$ at θ_0. The local behavior of this function determines the rate of convergence of the estimation of θ_0. The standard approach to inference relies on a local identification condition that can be expressed as

$$\mathbb{E}\left(\rho_\theta(x,\theta_0)'\rho_\theta(x,\theta_0)\right) \text{ is non singular,} \tag{4}$$

[1] For a comprehensive recent discussion of these issues in the context of asset pricing models, we refer the reader to Antoine *et al.* (2020).

where $\rho_\theta(x,\theta) := \mathbb{E}(\nabla_\theta u(y,\theta)|x)$ with $\nabla_\theta u(y,\theta_0) = \partial u(y_i,\theta)/\partial\theta'|_{\theta=\theta_0}$. Following the literature on unconditional moment restriction models (Sargan, 1983; Dovonon and Renault, 2013; Dovonon and Hall, 2018; among others), we shall refer to this condition as *first-order local identification* condition for conditional moment restriction models. This connection between the identification setups for unconditional and conditional restriction models is formalized in the next subsection. As pointed out in the introduction, this paper considers a framework where the conditional moment model is point identified but there is a failure of the first-order local identification condition.

2.1 Identification in conditional moment restriction models

Note that, for any k and any vector of instruments $z_t = g(x_t) \in \mathbb{R}^k$, function of x_t, the conditional moment restriction in (1) implies the unconditional moment restriction:

$$\mathbb{E}\left(z_t \cdot u(y_t,\theta_0)\right) = 0. \tag{5}$$

According to Sargan (1983), Dovonon and Renault (2013), and Lee and Liao (2018), among others, the unconditional moment restriction (5) locally identifies θ_0 at first order if

$$\text{Rank}(\mathbb{E}(z_t \cdot \nabla_\theta u(y_t,\theta_0)) = p \tag{6}$$

whereas lack of first-order local identification occurs when

$$\text{Rank}(\mathbb{E}(z_t \cdot \nabla_\theta u(y_t,\theta_0)) < p. \tag{7}$$

Therefore, it is reasonable to conjecture that the conditional moment restriction (1) identifies θ_0 locally at first order if and only if there exists a set of instruments z_i such that (6) holds. Relatedly, first-order local identification fails if and only if (7) holds regardless of the choice of instruments. The following proposition describes this property in terms of degeneracy of the expected Jacobian of $u(y_t,\theta)$ at θ_0.

Proposition 2.1 *The following two statements are equivalent:*

(i) For any k and any $\mathbb{R}^k$-valued measurable function g, $\text{Rank}\left(\mathbb{E}\left(z_t \cdot \nabla_\theta u(y_t,\theta_0)\right)\right) < p$*, where $z_t = g(x_t)$ and assuming that the moment exists.*

(ii) There exists at least one linear combination of the elements of $\mathbb{E}\left(\nabla_\theta u(y_t,\theta_0)|\, x_t\right)$ that is almost surely nil.

A proof of an alternative formulation of this proposition (Proposition 2.2(ii)) is provided in Appendix B. The characterization in (ii) validates (4) as the first-order local identification condition in the model (1). Further, this highlights some similarities with the first-order local identification failure in parametric models as studied by Rotnitzky *et al.* (2000). In this setting, first-order local identification failure amounts to linear dependence of the elements of the score function of the model, evaluated at the true parameter value.

For further characterization of point identification and first-order local identification failure, let us consider the separable Hilbert space $L^2(P) := L^2(\mathbb{R}^{k_x}, \mathcal{B}(\mathbb{R}^{k_x}), P)$ of square P-integrable real-valued functions defined on $\mathbb{R}^{k_x}$, where P is the common probability distribution of x_t's - that are assumed to be stationary - and $\mathcal{B}(\mathbb{R}^{k_x})$ is the Borel σ-algebra of $\mathbb{R}^{k_x}$. Let $(g_l)_{l\in\mathbb{N}}$ be a countable basis (not necessarily orthonormal) of $L^2(\mathbb{R}^{k_x}, \mathcal{B}(\mathbb{R}^{k_x}), P)$. Also, let $g^{(k)} := (g_1, \ldots g_k)'$ and $\alpha_l(\theta) = \langle g_l(\cdot), \rho(\cdot, \theta)\rangle := \mathbb{E}(g_l(x)\rho(x, \theta))$. Then, we have the following proposition.

Proposition 2.2

(i) – *If the conditional moment restriction* (1) *satisfies the point identification property in* (3) *and if there exists $k_0 \in \mathbb{N}$ such that $\alpha_l \equiv 0$ for all $l \geq k_0$, then, for all $k \geq k_0$,*

$$\forall \theta \in \Theta, \quad \mathbb{E}\left(g^{(k)}(x) \cdot u(y, \theta)\right) = 0 \Leftrightarrow \theta = \theta_0. \tag{8}$$

– *More generally, assume that (a) Θ is compact, (b) $(g_l)_l$ is orthonormal, (c) $\theta \mapsto \mathbb{E}[\rho(x, \theta)]^2$ is continuous on Θ, and, (d) $\lim_k \left[\sup_{\theta\in\Theta} \sum_{i\geq k} \alpha_i(\theta)^2\right] = 0$.*
Then, if model (1) *satisfies the point identification property in* (3) *and for every k there exists $\theta_k \in \Theta \setminus \{\theta_0\}$ such that: $\mathbb{E}[g^{(k)}(x)u(y, \theta_k)] = 0$, we have*

$$\theta_k \to \theta_0, \text{ as } k \to \infty.$$

(ii) If $\mathbb{E}(\nabla_\theta u(y, \theta_0)|x) \in (L^2(P))^p$, then the conditional moment restriction (1) *fails to satisfy the first-order local identification condition if and only if there exist $0 \leq r < p$ and $k_0 \in \mathbb{N}$ such that, for all $k \geq k_0$,*

$$\text{Rank}\left(\mathbb{E}\left(g^{(k)}(x) \cdot \nabla_\theta u(y, \theta_0)\right)\right) = r < p.$$

Part (i) of Proposition 2.2 aims to establish an equivalence between point identification of conditional moment restriction (1) - as expressed by (3) - and point identification of the unconditional moment condition $\mathbb{E}(g^{(k)}(x)u(y, \theta)) = 0$ for large enough k. Note that point identification of the unconditional model for any value of k implies point identification of the conditional model. Part (i) of the proposition establishes the converse for estimating functions $\rho(x,\theta)$ that have finite number of nonzero components in a given basis. This is typically the case for polynomial functions in x. In more general cases, the second part of (i) establishes, under mild conditions, that all the solutions to $\mathbb{E}(g^{(k)}(x)u(y,\theta)) = 0$ eventually collapse to θ_0 as k grows. Regarding condition (d), it is worth mentioning that $x \mapsto \rho(x,\theta) \in L^2(P)$ ensures that, for each $\theta \in \Theta$, $\sum_{i\geq k} \alpha_i^2 \to 0$ as k grows. Condition (d) imposes uniformity to simplify the proof.

Part (ii) is concerned with first-order local identification as it relates local identification failure in conditional and unconditional models. When k is large enough, the expected Jacobian

$$G^{(k)} := \mathbb{E}(g^{(k)}(x)\nabla_\theta u(y,\theta_0))$$

reaches the maximum rank $r < p$. Since the moment restrictions are inclusively increasing with k, this also implies that the null space $G^{(k)}$ and the range of its transpose are fixed. This stability of range and null space will be key to second-order local identification that will be imposed on the moment restrictions in order to characterize the limiting behavior of estimators and specification tests. Indeed, the main consequence of first-order local identification failure, while global identification holds, is that a certain number $(r < p)$ of directions of the parameter vector are identified through first-order expansions of the moment function.

Next, following Dovonon and Hall (2018) and Dovonon and Renault (2020), we focus on configurations that allow the identification of the remaining directions via a second-order expansion. Let $k \geq k_0$ such that $\text{Rank}(G^{(k)}) = r < p$, R_1 be a (p,r)-matrix with columns spanning the range of $G^{(k)\prime}$, and R_2 denote a $(p,p-r)$-matrix with columns spanning the null space of $G^{(k)}$. We say that the moment restriction (1) identifies θ_0 at second order if, for all $u \in \mathbb{R}^r$ and $v \in \mathbb{R}^{p-r}$, we have[2]

$$\left(G^{(k)}R_1u + \left(v'R_2'\mathbb{E}\left(g_l^{(k)}(x)\nabla_{\theta\theta}u(y,\theta_0)\right)R_2v\right)_{1\leq l\leq k}\right) = 0 \Leftrightarrow ((u,v) = (0,0)), \tag{9}$$

where $\nabla_{\theta\theta}u(y,\theta_0) := (\partial^2/\partial\theta\partial\theta')u(y,\theta)$, evaluated at θ_0.

Letting $M^{(k)}$ be the matrix of orthogonal projection on the null space of $G^{(k)\prime}$ (or equivalently the orthogonal of the column span of $G^{(k)}$), Corollary 2.3 of Dovonon and Renault (2020) ensures that (9)

[2]From our discussion above, if (9) holds for a given k, it holds for all $k' \geq k$.

is equivalent to the existence of $\gamma_k > 0$ such that, for any $v \in \mathbb{R}^{p-r}$,

$$\left\| M^{(k)} \left(v' R_2' \mathbb{E} \left(g_l^{(k)}(x) \nabla_{\theta\theta} u(y, \theta_0) \right) R_2 v \right)_{1 \le l \le k} \right\|_2 \ge \sqrt{\gamma_k} \|v\|_2^2,$$

$$\text{with } \gamma_k = \inf_{\|v\|_2 = 1} \left\| M^{(k)} \left(v' R_2' \mathbb{E} \left(g_l^{(k)}(x) \nabla_{\theta\theta} u(y, \theta_0) \right) R_2 v \right)_{1 \le l \le k} \right\|_2^2. \quad (10)$$

This inequality is instrumental in deriving the rate of convergence of estimators of θ_0. It is worth mentioning that it can be shown that γ_k is a nondecreasing function of k and unlike the case of a fixed set of unconditional moment restrictions (Dovonon and Hall, 2018; Dovonon and Renault, 2020), this property has the potential to affect the rate of convergence of the estimators of θ_0, especially if γ_k diverges to ∞ as k grows.[3]

We conclude this section with some remarks regarding the second-order local identification. We define the (k, p^2)-matrix

$$H^{(k)}(\theta) = \mathbb{E} \left(g^{(k)}(x)[\text{vec}'(\nabla_{\theta\theta} u(y, \theta))] \right)$$

and $\bar{H}^{(k)}(\theta)$ its sample counterpart. By definition, it follows that

$$\left(v' R_2' \mathbb{E} \left(g_l^{(k)}(x) \nabla_{\theta\theta} u(y, \theta_0) \right) R_2 v \right)_{1 \le l \le k} = H^{(k)}(\theta) \cdot \text{vec}(R_2 v v' R_2).$$

Remark 1 *If $p = 1$, first-order local identification failure at θ_0 amounts to $G^{(k)} = 0$ for all k. Then, second-order local identification amounts to $H^{(k)}(\theta_0) \neq 0$ for some k.*

Remark 2 *If $p \ge 2$ and $\text{Rank}(G^{(k)}) = p - 1$ for all $k \ge k_0$, second-order local identification at θ_0 amounts to*

$$\text{Rank} \begin{bmatrix} G^{(k)} R_1 & H^{(k)}(\theta_0) \cdot \text{vec}(R_2 R_2') \end{bmatrix} = p,$$

for R_1 and R_2 defined as in (9).

Remark 3 *If, more specifically, $\text{Rank}(G^{(k)}) = p - 1$ for all $k \ge k_0$, and in some parameter direction, say θ_h, we have $\mathbb{E} \left(g^{(k)}(x)[\partial u(y, \theta_0)/\partial \theta_h] \right) = 0$, then second-order local identification amounts to*

$$\text{Rank} \begin{bmatrix} G^{(k)} d_h & H^{(k)}(\theta_0) \cdot \text{vec}(e_h e_h') \end{bmatrix} = p,$$

where $G^{(k)} d_h$ is the $(k, p-1)$-matrix of columns of $G^{(k)}$, except the h-th, and e_h is the p-vector of zeros with 1 in its h-th entry. Also, note that

$$H^{(k)}(\theta_0) \cdot \text{vec}(e_h e_h') = \mathbb{E}(g^{(k)}(x)[\partial^2 u(y, \theta_0)/\partial \theta_h^2]).$$

[3] A proof of the statement for a nondecreasing γ_k is available from the authors upon request.

The next subsection provides two examples that illustrate the identification properties which are outlined and discussed so far. In these examples, the conditional moment restriction model fails first-order local identification but is point identified and satisfies the second-order local identification condition.

2.2 Motivating examples

Example 1. The first example is quite stylized and its simplicity allows us to illustrate easily the main features of our identification setup. More specifically, consider the conditional moment restriction $\mathbb{E}(u_t(\theta)|x_t) = 0$ with

$$u_t(\theta) = (y_t - \theta)^2 - 1 : \quad t = 1, \ldots, n,$$

where $y_t \sim iid(0,1)$ and y_t is independent of the stationary process x_t. This setup corresponds to Remark 1 above and is characterized with first-order local identification failure even though global identification (of the true value $\theta_0 = 0$) is assured. Local identification is obtained at second order: here, $k_0 = 1$ since with the instrument $z_{t,1} = 1$, we have $\mathbb{E}\left(\nabla_{\theta\theta}(z_{t,1}u_i(\theta_0))\right) = -2 \neq 0$.

Example 2. The second example is much more realistic as it is believed to capture some important common drivers in financial asset returns (Engle and Kozicki, 1993; see also Dovonon and Renault, 2013). Let Y_t be a bivariate (stock, bond, or other financial asset) returns process, generated by the following conditionally heteroskedastic (CH) factor representation

$$Y_{t+1} = \Lambda f_{t+1} + e_{t+1}, \tag{11}$$

where Λ is a (2×2) matrix of factor loadings, $f_{t+1} := (f_{1,t+1}, f_{2,t+1})'$ is the vector of (unobserved) CH factors and e_{t+1} is the vector of idiosyncratic shocks. Letting $\mathcal{F}_t$ denote an increasing filtration to which all information available at date t is adapted, (11) is assumed to satisfy the following restrictions:

$$\mathbb{E}(e_{t+1}|\mathcal{F}_t) = 0,\ \mathrm{Var}(e_{t+1}|\mathcal{F}_t) = \Omega,\ \mathrm{Cov}(e_{t+1}, f_{t+1}|\mathcal{F}_t) = 0,$$

$$\mathbb{E}(f_{t+1}|\mathcal{F}_t) = 0,\ \mathrm{Var}(f_{t+1}|\mathcal{F}_t) = \mathrm{Diag}(\sigma^2_{1,t+1}, \sigma^2_{2,t+1}),$$

where Ω is time invariant and $\sigma^2_{i,t}$ $(i = 1, 2)$ are time-varying volatility processes.

Each return process is supposed to be characterized by CH dynamics so that each has at least one nonzero factor loading. Interest lies in configurations where both assets have commonality in their CH dynamics in the sense that they share the same source of heteroskedasticity. A common CH feature thus amounts to collinearity of γ_1 and γ_2 and is tested within (11) by investigating whether there exists a linear combination of returns that offsets the CH feature. The null of commonality of CH features can be posed as validity of the conditional moment restriction:

$$\mathbb{E}[u_{t+1}(\beta, c)|\mathcal{F}_t)] = 0, \text{ with } u_{t+1}(\beta, c) := (Y_{1,t+1} + \beta Y_{2,t+1})^2 - c, \tag{12}$$

where $\theta := (\beta, c)' \in \mathbb{R}^2$ is the model parameter vector.

Point identification is ensured in this model since, under the null, the only value of β that offsets the CH factor is: $\beta_0 = -\Lambda_{1,1}/\Lambda_{1,2}$ and $c_0 = \mathbb{E}(Y_{1,t+1} + \beta_0 Y_{2,t+1})^2$. This model also fails first-order local identification. Indeed, simple calculations yield:

$$\rho_\theta(\mathcal{F}_t, \theta_0) = \mathbb{E}(\nabla_\theta u_{t+1}(\theta_0)|\mathcal{F}_t) = \begin{pmatrix} \omega & -1 \end{pmatrix} \text{ and } \mathrm{Rank}(\mathbb{E}[\rho_\theta(\mathcal{F}_t, \theta_0)'\rho_\theta(\mathcal{F}_t, \theta_0)]) = 1 < 2,$$

with $\omega = 2(\Omega_{21} + \theta_0 \Omega_{22})$.

To establish second-order local identification, let us consider a basis, $(g_l(\mathcal{F}_t))_l$, of $L^2(P)$ such that $g_1(\mathcal{F}_t) = 1$ and $z_t := g_2(\mathcal{F}_t)$ is such that $Cov(z_t, Y_{2,t+1}^2) \neq 0$. Consider $Z_t := g^{(2)}(\mathcal{F}_t) = (1, z_t)'$. Then, we have

$$G^{(2)} := \mathbb{E}[Z_t \nabla_\theta u_{t+1}(\theta_0)] = \begin{pmatrix} 1 \\ \mathbb{E}(z_t) \end{pmatrix} \begin{pmatrix} \omega & -1 \end{pmatrix}$$

so that the range of $G^{(2)'}$ and the null space of $G^{(2)}$ are both of dimension 1, spanned by $R_1 = (\omega \ -1)'$ and $R_2 = (1 \ -\omega)'$, respectively. Furthermore,

$$\mathbb{E}(\nabla_{\theta\theta} u_{t+1}(\theta_0)) = \begin{pmatrix} 2\mathbb{E}(Y_{2,t+1}^2) & 0 \\ 0 & 0 \end{pmatrix}, \quad \text{and} \quad \mathbb{E}(z_t \cdot \nabla_{\theta\theta} u_{t+1}(\theta_0)) = \begin{pmatrix} 2\mathbb{E}(z_t Y_{2,t+1}^2) & 0 \\ 0 & 0 \end{pmatrix}.$$

Hence, for $u, v \in \mathbb{R}$, we have

$$G^{(2)} R_1 u + v^2 \begin{pmatrix} R_2' \mathbb{E}(\nabla_{\theta\theta} u_{t+1}(\theta_0)) R_2 \\ R_2' \mathbb{E}(z_t \cdot \nabla_{\theta\theta} u_{t+1}(\theta_0)) R_2 \end{pmatrix} = 0 \Leftrightarrow \begin{pmatrix} 1 \\ \mathbb{E}(z_t) \end{pmatrix} (\omega^2 + 1) u + \begin{pmatrix} \mathbb{E}(Y_{2,t+1}^2) \\ \mathbb{E}(z_t Y_{2,t+1}^2) \end{pmatrix} v^2 = 0$$

which yields $u = v = 0$.

3 Asymptotic properties of the GMM estimator

In order to establish the limiting behavior of the test of $H_0 : \Pr\{\mathbb{E}(u(y_t, \theta_0)|x_t) = 0\} = 1$ under second-order local identification, we first need to derive the asymptotic properties of the associated GMM estimators within this identification framework. In the light of Proposition 2.1, we investigate this null hypothesis by proposing a test for the sequence of unconditional moment restrictions:

$$\mathbb{E}(g_l(x_t) u(y_t, \theta_0)) = 0, \quad l = 1, \ldots, k; \quad t = 1, \ldots, n, \tag{13}$$

where $k = k(n)$ with $k(n) \to \infty$ as $n \to \infty$. In the standard identification setting, de Jong and Bierens (1994) argue that this testing approach delivers a consistent test for H_0 against arbitrary forms of alternative hypotheses. To enhance power, the sequence of functions $(g_l)_l$ is chosen as an enumeration of some series expansion that does not depend on θ.

As discussed by de Jong and Bierens (1994), the conditioning random variable x can be considered bounded since

$$\mathbb{E}(Y|x) = \mathbb{E}(Y|\Psi(x))$$

for any one-to-one function Ψ that maps $\mathbb{R}^{k_x}$ into a compact subset $D \subset \mathbb{R}^{k_x}$. In that respect, if x is not initially a bounded random variable, one can consider $g(\Psi(x))$, instead of $g(x)$. Examples of bounded transformations $\Psi(\cdot)$ include the *component-wise* `arc` tangent function, i.e. $x \mapsto \arctan(x) = (\arctan(x_1), \ldots, \arctan(x_{k_x}))'$, while choices of enumeration of weight functions $(g_l)_l$ include polynomial, trigonometric, and Flexible Fourier Form families (see de Jong and Bierens, 1994). We refer to Andrews (1991) and Gallant (1981) for more details on these families. In our simulations, we employ $g_l(x) = \cos(lx)$, $l = 1, \ldots$.

As previously, we set $g^{(k)} = (g_1, \ldots, g_k)' \in \mathbb{R}^k$ and write (13) in a more compact form as:

$$\mathbb{E}\left(g^{(k)}(x_t)u(y_t, \theta_0)\right) = 0.$$

Let the GMM estimator be given by:

$$\hat{\theta} = \arg\min_{\theta \in \Theta} \bar{f}_k(\theta)'\hat{W}_k\bar{f}_k(\theta), \tag{14}$$

where

$$\bar{f}_k(\theta) = \frac{1}{\sqrt{n}}\sum_{t=1}^{n} f_k(x_t, y_t, \theta), \quad f_k(x_t, y_t, \theta) = g^{(k)}(x_t)u(y_t, \theta_0),$$

and $\hat{W}_k$ is a sequence of symmetric, positive definite weighting matrices. The specification test that we introduce next is expressed as a function of the so-called two-step efficient GMM estimator which uses the weighting matrix

$$\hat{W}_k = \hat{V}_k^{-1}, \quad \hat{V}_k = \frac{1}{n}\sum_{t=1}^{n} f_k(x_t, y_t, \tilde{\theta})f_k(x_t, y_t, \tilde{\theta})'. \tag{15}$$

The preliminary (first-step) GMM estimator $\tilde{\theta}$ used in (15) is commonly obtained by setting $\hat{W}_k = W_{k,0}$. We shall require that for all k, all the eigenvalues of $W_{k,0}$ lie between two positive constants c_1 and c_2. Since $W_{k,0}$ is often set to I_k, this condition is not restrictive. Furthermore, we will derive our results under the condition that the sequence $(f_k(x_t, y_t, \theta_0))_i$ is serially uncorrelated. This is ensured by our maintained assumption that

$$\mathbb{E}(u(y_t, \theta_0)|\mathcal{F}_t) = 0, \text{ where } \mathcal{F}_t = \sigma(x_t, u(y_{t-1}, \theta_0), x_{t-1}, u(y_{t-2}, \theta_0), \ldots). \tag{16}$$

In fact, under this condition and for any k, $(f_k(x_t, y_t, \theta_0))_t$ is a martingale difference sequence with respect to its natural filtration. The second weighting matrix of the two-step GMM estimator therefore boils down to the inverse of $\hat{V}_k$, where $\hat{V}_k$ is a sum of outer product of $f_k(x_t, y_t, \tilde{\theta})$ as defined by (15).

Our goal is to establish the asymptotic distribution of the test of (2)

$$\hat{Z} = \frac{1}{\sqrt{2k}} \left(\bar{f}_k(\hat{\theta})' \hat{V}_k^{-1} \bar{f}_k(\hat{\theta}) - k \right) \tag{17}$$

under first-order local identification failure. Characterizing the limiting distribution of $\hat{Z}$ requires that we determine the limiting behavior of $\hat{\theta}$ under (i) an expanding set of moment conditions ($k(n) \to \infty$ as $n \to \infty$) and (ii) second-order local identification.

3.1 Assumptions

This section collects and presents the main set of assumptions that allows us to characterize the limiting behavior of the GMM estimator in the conditional moment setup under local identification failure.

Assumption 1 (Data dependence structure) *We assume that $u(y_t, \theta_0)$ is stationary and satisfies the dependence structure in* (16) *and $(y_t, x_t)_{t \in \mathbb{Z}}$ is a strong mixing process with dependence measure $\alpha(s) = O(\rho^s)$ for some $0 < \rho < 1$.*

Assumption 1 ensures that the conditional moment restriction (1) holds. It also implies that $f_k(x_i, y_i, \theta_0)$ is a martingale difference sequence with respect to its natural filtration, and is strong mixing with geometrically decreasing dependence coefficient. The mixing property is useful to deal with the serial correlation of functions of f_k. Although this dependence structure may appear restrictive, it encompasses a large class of time series representations that are useful in applications. This includes a wide range of linear and nonlinear processes. Carrasco and Chen (2002) and Francq and Zakoian (2006) establish the geometric strong mixing dependence property for conditionally heteroskedastic processes such as GARCH and stochastic volatility processes. More recently, Fryzlewicz and Subba Rao (2011) demonstrate that time-varying ARCH processes also share this property.

Assumption 2 (Identification setup) *There exists $k_0 \geq 1$ such that:*

(i) ***(Point identification)*** *The parameter space Θ is a compact subset of $\mathbb{R}^p$ and for all $k \geq k_0$,*

$$\forall \theta \in \Theta, \quad \mathbb{E}\left(u(y,\theta)g^{(k)}(x)\right) = 0 \Leftrightarrow \theta = \theta_0.$$

(ii) ***(First-order local identification failure)*** *For all $k \geq k_0$,*

$$\text{Rank}\left(G^{(k)}\right) = r < p.$$

(iii) ***(Second-order local identification)*** *For any $k \geq k_0$ and for R_1 and R_2 defined as in* (9)*, we have: for all $u \in \mathbb{R}^r$, all $v \in \mathbb{R}^{p-r}$ and all $k \geq k_0$,*

$$\left(G^{(k)}R_1 u + \left(v'R_2'\mathbb{E}\left(g_l^{(k)}(x)(\nabla_{\theta\theta}u)(\theta_0)\right)R_2 v\right)_{1\leq l\leq k} = 0\right) \Rightarrow ((u,v) = (0,0)). \tag{18}$$

Assumption 2 characterizes the identification framework for our analysis. Assumption 2(i) is necessary for establishing the consistency of GMM estimators. As we argued above in Proposition 2.2, this condition is equivalent, under mild regularity conditions, to point identification of the conditional moment model. Assumption 2(ii) imposes first-order local identification failure. This condition is shown to be equivalent to a lack of first-order local identification in conditional moment restriction models. Second-order local identification is presented in Assumption 2(iii).

Theorem A.1 in Appendix A shows that the GMM estimator is consistent under Assumptions 1, 2(i) and the following Assumption 3.

Assumption 3 (Consistency of GMM estimators) *Assume that:*

(i) $\mathbb{E}\left(g_l(x_i)^2 u(y_i,\theta_0)^2\right) < \Delta < \infty$*, for some $\Delta > 0$; $\theta \mapsto \mathbb{E}[g_l(x)u(y,\theta)]$ is continuous for each $l = 1,\ldots,k_0$, with k_0 as defined in Assumption 2(i); and*

$$\sup_{\theta\in\Theta}\left\|\frac{1}{n}\sum_{i=1}^{n} g^{(k_0)}(x_i)u(y_i,\theta) - \mathbb{E}[g^{(k_0)}(x_i)u(y_i,\theta)]\right\|_2 \xrightarrow{P} 0.$$

(ii) There exists a nonrandom sequence W_k of (k,k)-symmetric positive definite matrices and $\Delta > 0$ such that, with $\bar{\lambda}_k := \lambda_{\max}(W_k)$, we have

$$\bar{\lambda}_k \,/ \lambda_{\min}(W_k) \leq \Delta < \infty, \quad \lambda_{\max}(\hat{W}_k)\Big/\bar{\lambda}_k = 1 + o_P(1), \quad \lambda_{\min}(\hat{W}_k)\Big/\lambda_{\min}(W_k) = 1 + o_P(1).$$

The existence of the second moment of $g_l(x)u(y,\theta_0)$ in Assumption 3(i) allows us to control the order of magnitude of quantities such as $\|\bar{f}_k(\theta_0)\|_2^2$. Under this condition, $\|\bar{f}_k(\theta_0)\|_2^2 = O_P(k)$. The last part in Assumption 3(i) is the usual uniform law of large numbers. Primitive conditions for this to hold for dependent data can be found in Domowitz and White (1982) and Pötscher and Prucha (1989).

The first condition in Assumption 3(ii) is not restrictive as it merely rules out the possibility that W_k is ill-conditioned. In standard problems where k is fixed, $\hat{W}_k$ is assumed to converge in probability to W_k. With increasing k, such a convergence needs to be formalized. It turns out that the convergence of the extreme eigenvalues of $\hat{W}_k$ to those of W_k is sufficient to establish consistency of GMM. The conditions in Assumption 3(ii) are trivially fulfilled by estimation procedures using non-random weighting matrix with eigenvalues bounded away from 0 and from above. This is the case for the first-step GMM estimator introduced above which uses $W_{k,0}$ as weighting matrix which is then consistent. We also show that under Assumption A.1 in Appendix A, these conditions continue to hold for the two-step efficient GMM estimator, and this estimator is consistent as well (see Corollary A.2 in Appendix A).

To summarize, Assumptions 1, 2(i), 3 already ensure the consistency of the GMM estimator in point-identified, conditional restriction model under first-order local identification failure (see Theorem A.1 in Appendix A). However, for deriving the asymptotic distribution of the test for correct model specification, we need to characterize the rate of convergence of the GMM estimator.

For this reason, we proceed with introducing additional conditions that will allow us to establish the rate of convergence of the GMM estimator in conditional moment restriction models under local identification failure. In what follows, we define $\bar{H}^{(k)}(\theta) = n^{-1}\sum_{i=1}^{n} g^{(k)}(x_i)[\text{vec}'(\nabla_{\theta\theta}u(y_i,\theta))]$ and $\bar{G}^{(k)}(\theta) = n^{-1}\sum_{i=1}^{n} g^{(k)}(x_i)\nabla_{\theta}u(y_i,\theta)$ to be the sample counterparts of $H^{(k)}(\theta)$ and $G^{(k)}(\theta)$, respectively, that were introduced above. Also, let $H^{(k)} := H^{(k)}(\theta_0)$, $D_1 = G^{(k)}R_1$, $\bar{D}_2 = \sqrt{n}\bar{G}^{(k)}(\theta_0)R_2$, $M^{(k)} = I_k - W_k^{1/2}D_1(D_1'W_kD_1)^{-1}D_1'W_k^{1/2}$, $\gamma_k = \inf_{\|v\|=1}\|M^{(k)}W_k^{1/2}H^{(k)}\text{vec}(R_2vv'R_2')\|_2^2$, $\bar{\lambda}_k := \lambda_{\max}(W_k)$ and $\lambda_k := \lambda_{\min}(W_k)$, where W_k is defined as in Assumption 3(ii).

We first state a condition (Condition C below) for an $\mathbb{R}^m$-valued random function $U_i(\theta)$ that proves useful in obtaining the order of magnitude of quantities such as $(1/n)\sum_{i=1}^{n} g^{(k)}(x_i)U_i(\bar{\theta}) - \mathbb{E}(g^{(k)}(x_i)U_i(\theta_0))$, where $\bar{\theta}$ converges in probability to θ_0 (see Lemma B.1 in Appendix B).

Condition C. *For an $\mathbb{R}^m$-valued random function $U_i(\theta)$, there exists a neighborhood $\mathcal{N}$ of θ_0 such that:*

$$\sup_{\theta\in\mathcal{N}}\left\|\frac{1}{n}\sum_{i=1}^{n} g^{(k)}(x_i)U_i(\theta)' - \mathbb{E}\left(g^{(k)}(x_i)U_i(\theta)'\right)\right\|_2 = O_P\left(\sqrt{\frac{k}{n}}\right),$$

and for each $l \in \{1,\ldots,k\}$, and $r \in \{1,\ldots,m\}$, the map $\theta \mapsto \mathbb{E}\left(g_l(x_i)U_{i,r}(\theta)\right)$ is Lipschitz continuous on $\mathcal{N}$ with coefficient $c > 0$, i.e.,

$$\forall\theta_1,\theta_2 \in \mathcal{N},\ \|\mathbb{E}\left(g_l(x_i)U_{i,r}(\theta_1)\right) - \mathbb{E}\left(g_l(x_i)U_{i,r}(\theta_2)\right)\|_2 \leq c\|\theta_1 - \theta_2\|_2.$$

The first condition is warranted if the functional central limit theorem applies. The Lipschitz property follows if the considered expectation functions are continuous on a compact set containing a neighborhood of θ_0. The common Lipschitz constant may appear restrictive although such a constant exists if we assume that $\sup_{\theta\in\mathcal{N}}\mathbb{E}(|g_l(x)|\|\partial U_{i,r}(\theta)/\partial\theta'\|_2) \leq \Delta < \infty$. We now present the final assumption that is necessary to derive the rate of convergence of the GMM estimator.

Assumption 4 (Orders of magnitude) *Assume that:*

(i) *θ_0 lies in the interior of Θ and $\theta \mapsto u(y,\theta)$ is twice continuously differentiable in a neighborhood of θ_0 for each y and the maps $\theta \mapsto \nabla_\theta u(y,\theta)$ and $\theta \mapsto \mathrm{vec}(\nabla_{\theta\theta}u(y,\theta))$ satisfy Condition C.*

(ii) *There exist $\alpha_1, \alpha_2 > 0$ such that: $\alpha_1\sqrt{k} \leq \|D_1\|_2 \leq \alpha_2\sqrt{k}$, $\alpha_1\sqrt{k} \leq \|H^{(k)}\|_2 \leq \alpha_2\sqrt{k}$, $\alpha_1 k \leq \gamma_k \leq \alpha_2 k$, $\lambda_{\max}(D_1'D_1)/\lambda_{\min}(D_1'D_1) = O(1)$, and $\|\bar{D}_2\|_2 = O_P(\sqrt{k})$.*

(iii) $S_k := \gamma_k^{-1/2}H^{(k)\prime}W_k^{1/2}M^{(k)}W_k^{1/2}\bar{f}_k(\theta_0) = O_P(1)$.

(iv) *$\bar{\lambda}_k \leq \Delta < \infty$, for some $\Delta > 0$ and $\sqrt{k}\|\hat{W}_k - W_k\|_2 = o_P(1)$.*

Note that D_1 and $H^{(k)}$ are non-zero matrices and the condition on their spectral norms in Assumption 4(ii) follows if each has at least one column with a number of non-zero elements that is proportional to k. The magnitude of γ_k follows from the fact that (a) it is a nondecreasing sequence in k, and (b) it is of order $O_P(\|H^{(k)}\|_2^2)$. The requirement that the ratio of the extreme eigenvalues of $D_1'D_1$ be bounded preserves this nonsingular matrix from being ill-conditioned. the condition on the $(k, p-r)$-matrix $\bar{D}_2$ is not particularly restrictive since each component of this matrix is $O_P(1)$ by virtue of the central limit theorem.

In Assumption 4(iii), the order of magnitude of S_k follows if $\lambda_{\max}(W_k^{1/2}V_kW_k^{1/2}) \leq \Delta < \infty$, which is the case, e.g., if V_k and W_k have bounded eigenvalues or if $W_k = V_k^{-1}$. To see this, note that if $\bar{\lambda}_k$

is bounded, then, for any unit vector c, we have:

$$c'\text{Var}(S_k)c = \gamma_k^{-1} c' H^{(k)\prime} W_k^{1/2} M^{(k)} W_k^{1/2} V_k W_k^{1/2} M^{(k)} W_k^{1/2} H^{(k)} c$$
$$\leq \Delta \gamma_k^{-1} c' H^{(k)\prime} W_k^{1/2} M^{(k)} W_k^{1/2} H^{(k)} c \leq \Delta \lambda_{\max}(W_k) \gamma_k^{-1} \|H^{(k)}\|_2^2 = O(1)$$

which is sufficient to claim that $S_k = O_P(1)$ since $\mathbb{E}(S_k) = 0$.

Lastly, Assumption 4(iv) imposes that the eigenvalues of W_k are bounded and $\hat{W}_k$ is sufficiently near W_k as n grows. Note that these two conditions are fulfilled by the GMM estimator with nonrandom matrix having bounded eigenvalues such as $W_{k,0}$. This conditions are also satisfied for the two-step efficient GMM estimator as we argue in the next subsection.

3.2 Limiting behavior of the GMM estimator

Given the set of assumptions stated above, we now proceed to establishing the rate of convergence of the GMM estimator which, in turn, will be useful to characterize the asymptotic distribution of the specification test. In the standard case of a fixed number of moment restrictions (i.e., k is fixed), the GMM estimator is known to converge at a sharp rate of $n^{1/4}$ although a faster rate in some regions of the sample space is possible (Dovonon and Renault, 2013). This mixture of rates is essential for deriving the asymptotic distribution of the GMM overidentification test statistic as a mixture of chi-squared random variables. We show a similar rate behavior for the GMM estimator in the current context under local identification failure although the original rate needs to be adjusted in order to reflect the increasing number of moment restrictions.

The next theorem states the rate of convergence of the parameter vector . Recall that R_1 denotes a (p, r)-matrix with columns spanning the range of $G^{(k)\prime}$ and R_2 is a $(p, p-r)$-matrix with columns spanning the null space of $G^{(k)}$, where $\text{Rank}(G^{(k)}) = r < p$.

Theorem 3.1 *If Assumptions 1-4 hold and $k \to \infty$ as $n \to \infty$ with $k^3/n \to 0$, then*

$$\|\hat{\theta} - \theta_0\|_2 = O_P(\gamma_k^{-1/4} n^{-1/4}), \quad \|R_1'(\hat{\theta} - \theta_0)\|_2 = O_P(n^{-1/2}), \quad \textit{and} \quad \|R_2'(\hat{\theta} - \theta_0)\|_2 = O_P(\gamma_k^{-1/4} n^{-1/4}).$$

Theorem 3.1 establishes that each of the components of the GMM estimator converges at least at a nonstandard rate of $\gamma_k^{1/4} n^{1/4}$ while the standard $\sqrt{n}$-rate of convergence is possible in some directions. More specifically, the directions of the parameter that are identified at first order are $\sqrt{n}$-convergent while the directions that are second-order locally identified converge at a slower, $\gamma_k^{1/4} n^{1/4} \sim k^{1/4} n^{1/4}$, rate. Interestingly, this rate is faster than the result in Dovonon and Renault (2020) who obtain, in a configuration of fixed number of moment restrictions, a slower rate $n^{1/4}$ for the directions identified at second order. The faster rate in our context is essentially due to to the increased information brought by the growing number of moment restrictions.

This finding bears some similarities to Han and Phillips (2006) who show in the context of weak instruments, that the GMM estimator may be consistent if the number of moment instruments is allowed to increase with the sample size (see also Chao and Swanson, 2005, among others). The intuition behind this result is that the expanding number of moment conditions, if growing at an appropriate rate with the sample size, enhances the identification signal and renders a consistent estimator even with possibly irrelevant instruments. In our framework, point identification is maintained and consistent estimation is therefore possible even if the number of moment restriction does not grow. But, as Theorem 3.1 shows, the second-order local identification also reaps important benefits from the expanding set of moment restrictions as the second-order identified parameters can be estimated at a faster rate. It is worth mentioning that since achieving consistent estimation requires the number of moment restrictions to grow at a slower rate than the sample size, it will not be possible to accelerate the convergence rate of second-order identified directions to the parametric $\sqrt{n}$-rate.

Some further remarks on the rates of convergence in Theorem 3.1, specialized to the first-step and two-step GMM estimators, are warranted. The validity of the results in Theorem 3.1 for the first-step and two-step GMM estimators hinge on verifying Assumptions 3(ii) and 4(iii, iv). For the two-step GMM estimator, we also use Assumption A.1 in Appendix A and Lemma B.2 in Appendix B.

Remark 4 *Since the weighting matrix $\hat{W}_k := W_{k,0}$ for the first-step GMM estimator is nonrandom with bounded eigenvalues from above and away from zero, Assumptions 3(ii) and 4(iv) are trivially verified. Assumption 4(iii) is also satisfied if, for instance, V_k has bounded eigenvalues and this estimator, say $\tilde{\theta}$, is characterized by $\tilde{\theta} - \theta_0 = O_P(k^{-1/4}n^{-1/4})$. Note that the eigenvalues of V_k are bounded under Assumption A.1(iv) and if V_k has uniformly bounded diagonal elements. This latter condition is implied by A.1(ii).*

Remark 5 *For the two-step efficient estimator with $\hat{W}_k := \hat{V}_k^{-1}$, we assume that the smallest eigenvalue of V_k is bounded away from 0; i.e., $\lambda_{\min}(V_k) \geq \underline{\lambda} > 0$ for all k. This is a reasonable assumption since $\lambda_{\max}(V_k)$ is an increasing sequence and we shall require that $\lambda_{\min}(V_k)$ and $\lambda_{\max}(V_k)$ are of the same order of magnitude to preserve V_k from being-ill conditioned. In this case,*

$$\lambda_{\max}(V_k^{-1}) = 1/\lambda_{\min}(V_k) \leq 1/\underline{\lambda}.$$

Furthermore, Lemma B.2(ii, v) in Appendix B ensures that

$$\lambda_{\max}(\hat{V}_k^{-1})\Big/\lambda_{\max}(V_k^{-1}) = 1 + o_P(1), \text{ and } \lambda_{\min}(\hat{V}_k^{-1})\Big/\lambda_{\min}(V_k^{-1}) = 1 + o_P(1).$$

Note that in this lemma, $v_n = (k^3/n)^{1/4} = o(1)$. This shows that Assumption 3(ii) holds.

Remark 6 *Finally, from Lemma B.2(iv) in Appendix B we have that $\sqrt{k}\|\hat{V}_k^{-1} - V_k^{-1}\|_2 = O_P(k^{5/4}/n^{1/4})$ which ensures that Assumption 4(iv) holds if $k^5/n \to 0$ as $n \to \infty$.*

Although the rates of convergence that are stated in Theorem 3.1 are sufficient to derive the asymptotic distribution of the specification test, it is interesting to further investigate the large sample properties of the GMM estimators. Unfortunately, characterizing the asymptotic distribution of the GMM estimator $\hat{\theta}$ in the general case proves difficult. For this reason, we restrict our attention to the simplest case of single parameter ($p = 1$) models with a second-order local identification property.

Theorem 3.2 *Suppose that $p = 1$ and Assumptions 1-4 hold. In addition, if $k \to \infty$ as $n \to \infty$ with $k^4/n \to 0$, and $\gamma_k^{-1/2} H^{(k)\prime} W_k \bar{f}_k(\theta_0) \xrightarrow{d} Z := N(0, \sigma^2)$ for some $\sigma^2 > 0$, then:*

$$\sqrt{\gamma_k n}(\hat{\theta} - \theta_0)^2 \xrightarrow{d} 1_{\{Z \geq 0\}}(2Z).$$

Theorem 3.2 first demonstrates that the slow rate of convergence derived in Theorem 3.1 is, in fact, sharp meaning that within the assumed model and identification framework, the estimator cannot converge at a faster rate. Furthermore, the asymptotic distribution in Theorem 3.2 can be readily used to conduct inference about the true parameter value θ_0 by replacing γ_k with its sample counterpart. Note that this non-standard asymptotic distribution with an atom mass of 1/2 at the origin is similar to the one derived by Dovonon and Hall (2018) for a fixed k. As pointed out above, the characterization of the asymptotic distribution in the general case of $p > 1$ appears to be quite involved and is beyond the scope of this paper.

4 Asymptotic distribution of the specification test

The characterization of the asymptotic distribution of our specification test statistic requires a central limit theorem for degenerate U-statistics with linear kernel of the form $h_n(x_t, x_s) := f_k'(x_t) V_k^{-1} f_k(x_s)$, where $(x_t)_{t\in\mathbb{Z}}$ is a stationary and strong mixing process and $(f_k(x_t))_{t\in\mathbb{Z}}$ is a martingale difference sequence with respect to its natural filtration. More specifically, we are interested in the asymptotic distribution of U-statistics of the form:

$$U_n = \frac{1}{n} \sum_{t \neq s} \frac{f_k(x_t)' V_k^{-1} f_k(x_s)}{\sqrt{k}}, \tag{19}$$

where $V_k := \mathrm{Var}(f_k(x_t))$. The degeneracy of U_n arises from the fact that $\int h_n(x, y)dF(y) = 0$ for all x, with F denoting the marginal distribution of x_t. The asymptotic theory of degenerate U-statistics has been extensively studied in the literature. Existing results cover cases where x_t is assumed to be i.i.d. or time-dependent as well as cases where the kernel function is sample-size dependent - as above - or fixed. A CLT for i.i.d. data and sample-size dependent kernel has been developed by Hall (1984) and played a key role in the main results of de Jong and Bierens (1994). An extension of Hall's (1984) results to β-mixing processes has been provided by Fan and Li (1999). More recent contributions to this literature include Leutch (2012), who proposes a CLT for τ-dependent processes[4] and fixed kernel,[5] and Gao (2007) and Gao and Hong (2008) who consider α-mixing processes and sample-size dependent kernels. Kim *et al.* (2011) further extend these results by establishing the CLT for U-statistics under quite general conditions. Our kernel is more consistent with the formulation in Kim *et al.* (2011) but the special factorization that they require is not well-aligned with our framework which features an inner product with an increasing dimension.

For this reason, we develop a new CLT that is adapted to the form of the U-statistic in (19). Since this result may be of independent interest, we collect the conditions that are sufficient for establishing the CLT in the following assumptions.

Assumption-CLT 1 *Assume that $(x_t)_{t\in\mathbb{Z}}$ is stationary and geometric strong mixing process, $f_k(x_t)$ is an $\mathbb{R}^k$-valued measurable function of x_t such that the sequence $(f_k(x_t))_{t\in\mathbb{Z}}$ is a martingale difference with respect to the σ-algebra $\sigma(f_k(x_s) : s \leq t)$.*

Assumption-CLT 2 *Assume that $k \sim n^\alpha$ for some $\alpha \in (0, 1)$ and there exists $\epsilon > 0$, such that*

$$\sup_{k\in\mathbb{N}} \frac{1}{k} \sum_{h=1}^{k} \mathbb{E}|[V_k^{-1/2} f_k(x_t)]_h|^{4+\epsilon} < \infty,$$

where $[a]_h$ is the h-th element of the vector a.

Stationarity and mixing of $(x_t)_{t\in\mathbb{Z}}$ is already assumed above (see Assumption 1) and is restated in Assumption-CLT 1 to ensure that the results in Proposition 4.1 and Theorem 4.2 below, which could be of independent interest, are self-contained. Assumption-CLT 2 is used to obtain the limit variance of U_n because its derivation requires dealing with 4-th order moments of $f_k(x_t)$. Replacing

[4] See Dedecker and Prieur (2005) for a definition. Note that i.i.d. $\Rightarrow$ β-mixing $\Rightarrow$ α-mixing $\Rightarrow$ τ-dependence.

[5] CLT for degenerate U-statistics with fixed kernel give rise to nonstandard asymptotic distribution taking the form of a quadratic function of an infinite number of independent Gaussian variables, while sample-size dependent kernels are typically associated with the standard normal distribution.

these moments by their analogues under independence is a common approach in the literature. The remainder is then controlled by resorting to Lemma OA.1 in the online Appendix, due to Roussas and Ioannides (1987), which can be applied if the condition on the moments in Assumption-CLT 2 is satisfied. Note that this condition is not too restrictive. It imposes the existence of moments of order higher than the fourth for the normalized components of $f_k(x_t)$. The boundedness of the average of these moments means that no component dominates the others in terms of these moments.

Proposition 4.1 *Under Assumptions-*CLT *1 and 2,* $\text{Var}(U_n) = 2 + o(1)$.

To obtain the asymptotic normality for U_n, we impose an additional assumption.

Assumption-CLT 3 *For some* $\beta \geq 0$*,*

$$\mathbb{E}\left(\max_{1\leq t\leq n} \|V_k^{-1/2} f_k(x_t)\|/\sqrt{k}\right) = O(\log^\beta n) \text{ and } \mathbb{E}\left(\max_{1\leq t\neq s\leq n} |f_k(x_t)' V_k^{-1} f_k(x_s)|/\sqrt{k}\right) = O(\log^\beta n).$$

The first bound in Assumption-CLT 3 is not restrictive as it holds with $\beta = 1$ provided that the moment generating function of $z_t := \|V_k^{-1/2} f_k(x_t)\|_2/\sqrt{k}$ exists. This holds regardless of the dependence structure. For instance, $\beta = 1$ if z_t has a Gamma distribution, and $\beta = 1/2$ if z_t is Gaussian. We would like to remark that $z_t = O_P(1)$ since $\mathbb{E}(z_t^2) = 1$. If z_t's are i.i.d. with common distribution F, it is known that this bound holds for a large class of F but rules out those with Paretian tail; see Pereira (1983). Similar results hold for time-dependent processes as well. We refer to Berman (1964) and Isaev *et al.* (2020) for a discussion.

The second bound in Assumption-CLT 3 is not too restrictive either. If $f_k(x_t)$ and $f_k(x_s)$ are independent, then $\mathbb{E}[f_k(x_t)' V_k^{-1} f_k(x_s)/\sqrt{k}]^2 = 1$ so that $|f_k(x_t)' V_k^{-1} f_k(x_s)|/\sqrt{k} = O_P(1)$ and, as before, we can claim that the stated bound accommodates a large class of processes. We are now ready to state the following CLT for the scaled U-statistic in (19).

Theorem 4.2 *Under Assumptions-*CLT *1, 2 and 3,*

$$\frac{U_n}{\sqrt{2}} \xrightarrow{d} N(0,1).$$

The proof of Theorem 4.2 follows similar arguments as in Kim *et al.* (2011) and is provided in the Online Appendix. We establish this CLT by showing that the moments of U_n converge to those of the normal distribution. Under Assumption-CLT 3, we show that the summands of U_n are essentially bounded by a slowly increasing function of the sample size, which turns out to be essential for controlling the difference between the moments U_n and those of its Gaussian limit.

Building on this central limit theorem, we now characterize the asymptotic distribution of the specification test statistic $\hat{Z} = \frac{1}{\sqrt{2k}}\left(\bar{f}_k(x_t, y_t, \hat{\theta})'\hat{V}_k^{-1}\bar{f}_k(x_t, y_t, \hat{\theta}) - k\right)$ under the null hypothesis that the conditional moment restriction (1) is correctly specified.

Theorem 4.3 *Suppose that Assumptions 1, 2, 3(i), 4(i ,ii), A.1(ii, iii, iv), and Assumptions-*CLT *2-3 with* $f_k(x) := f_k(x, y, \theta_0)$*, hold. Also, assume that* $k = o(n^{1/5})$ *and* $W_{k,0}$ *has bounded eigenvalues. Then, as* $n \to \infty$,

$$\hat{Z} \xrightarrow{d} N(0, 1).$$

Several remarks are warranted regarding the result in Theorem 4.3. First, it is important to underscore that the standard normal limit distribution in Theorem 4.3 is obtained in a highly non-standard setting. In particular, we have a lack of first-order local identification which, as discussed earlier, gives rise to non-standard limiting behavior of the GMM estimator. The second-order local identification, in conjunction with the expanding set of moment conditions, ensures the consistency of the estimator and determines its rate of convergence. The conditions for the consistency of the two-step GMM estimator are collected in Assumption A.1 in Appendix A and are used in establishing the limit in Theorem 4.3. While the $\hat{Z}$ test statistic is based on the efficient GMM estimator with $\hat{W}_k = \left[\frac{1}{n}\sum_{t=1}^{n} f_k(x_t, y_t, \tilde{\theta}) f_k(x_t, y_t, \tilde{\theta})'\right]^{-1}$, stating explicitly that $W_{k,0}$ has bounded eigenvalues allows us to invoke Assumption 4(iii) in order to ensure the desired rate of convergence for the preliminary GMM estimator $\tilde{\theta}$. (See Remark 4.) Also, as discussed earlier, Assumption 4(iii) holds provided that $\lambda_{\max}(W_k^{1/2} V_k W_k^{1/2}) \leq \Delta < \infty$ which is trivially satisfied by the efficient GMM estimator that sets $W_k = V_k^{-1}$.

In the conventional framework where the conditional model is point identified, the properly recentered and standardized specification test with an increasing number of moment conditions converges, under some regularity conditions, to a standard normal limit (see, for example, Carrasco and Florens, 2000; Donald *et al.*, 2003; Tripathi and Kitamura, 2003; among others). Theorem 4.3 establishes that the standard normal distribution continues to be the correct limit for the $\hat{Z}$ test statistic under the null of correct specification, provided that $k = o(n^{1/5})$ as $n \to \infty$. Unlike the regular setup, this limit is obtained within the second-order local identification framework in Assumption 2 which is characterized by first-order local identification failure. Intuitively, this is achieved by combining and balancing the benefits from the second-order local identification and the expanding number of moment conditions. Importantly, for the appropriate choice of k (as a function of n), inference for the correct specification of the conditional moment restriction model is straightforward in practice as it is based on the critical values from the standard normal distribution.

We complete our theoretical analysis by characterizing the limiting behavior of $\hat{Z}$ under the alternative hypothesis H_1, specified in (2). We will rely on the following strengthened version of Lemma 1 of de Jong and Bierens (1994).

Lemma 4.4 *Let Θ and D be compact subsets of $\mathbb{R}^p$ and $\mathbb{R}^{k_x}$, respectively. Let $u(\theta)$ be an $\mathbb{R}$-valued random function and x be a random variable with values in D. Assume that the functions $g_l(x)$ form an algebra of continuous real-valued functions on D that separates the points of D and contains the constant function. Assume further that $\theta \mapsto u(\theta)$ is continuous on Θ almost everywhere and $\mathbb{E}(\sup_{\theta\in\Theta}|u(\theta)|) < \infty$. Then, if $\Pr(\mathbb{E}(u(\theta)|x) = 0) < 1$ for each $\theta \in \Theta$,*

$$\exists k_0 \in \mathbb{N} \text{ and } \delta_0 > 0 : \quad \inf_{\theta\in\Theta} \left\| \mathbb{E}(g^{(k_0)}(x)u(\theta)) \right\|_2^2 > \delta_0.$$

The conditions imposed on the series functions $g_l(\cdot)$ are as in de Jong and Bierens (1994). The continuity and dominance conditions on $u(\theta)$ are useful to guarantee the continuity of $\theta \mapsto \mathbb{E}(g_l(x)u(y,\theta))$ for each l. Continuity of these functions and compactness of Θ are essential to claim the stated result.

This lemma can be applied readily to our conditional moment model and alternative hypothesis H_1 in (2). With appropriate choices of series functions $g_l(\cdot)$, H_1 implies that $\inf_{\theta\in\Theta} \|\mathbb{E}(g^{(k_0)}(x)u(y,\theta))\|_2 > 0$ for a fixed k_0 so that the unconditional moment restriction $\mathbb{E}(g^{(k_0)}(x)u(y,\theta)) = 0$ is misspecified. In this case, it is known that the Hansen-Sargan specification test for this unconditional restriction – albeit infeasible because k_0 is unknown – would be consistent. Theorem 4.5 shows that this result carries over to the feasible statistic $\hat{Z}$ which makes our specification test consistent against all alternatives.[6]

Theorem 4.5 *Let $\hat{V}_k(\theta) := n^{-1}\sum_{t=1}^n f_k(x_t, y_t, \theta) f_k(x_t, y_t, \theta)'$. Assume that $k^2 = o(n)$, the $g_l(\cdot)$ series are as in Lemma 4.4 and the conditions of that lemma are satisfied with $u(\theta) := u(y,\theta)$, and H_1 is true. Assume further that there exists $\bar{\lambda} > 0$ such that, with probability approaching one, $\sup_{\theta\in\Theta} \lambda_{\max}(\hat{V}_k(\theta)) \leq \bar{\lambda}k$, and $\sup_{\theta\in\Theta} |(1/n)\sum_{t=1}^n g_l(x_t)u(y_t,\theta) - \mathbb{E}(g_l(x_t)u(y_t,\theta))| = o_P(1)$ for each l. Then,*

$$\exists \delta > 0 : \quad \lim_{n\to\infty} \Pr(k^{3/2} n^{-1} |\hat{Z}| > \delta) = 1.$$

The conditions of this theorem are essentially a subset of those of the main Theorem 4.3. The purpose of the condition on the bound of $\lambda_{\max}(\hat{V}_k)$ is to facilitate the proof as we can rely on more

[6] Studying the asymptotic behavior of the test under local alternatives proves to be very involved as it requires characterizing the limiting behavior of the GMM estimator in misspecified conditional restriction models under drifting sequences and first-order local identification failure. This analysis is beyond the scope of this paper.

primitive conditions. Theorem 4.5 shows that $|\hat{Z}|$ diverges to infinity if k is such that $k^{3/2} = o(n)$. Note that in Theorem 4.3, that studies $\hat{Z}$ under H_0, we impose $k^5 = o(n)$. This shows that the proposed test is consistent and has power against all alternatives.

5 Numerical illustrations

In this section, we provide simulation evidence on the empirical size of the standard normal asymptotic approximation of the specification test. We also apply the proposed testing framework to study the presence of a common CH factor in bond portfolio returns.

5.1 Simulations

We assess the finite-sample properties of the specification test $\hat{Z}$ in Monte Carlo simulations. We consider several simulation setups. The first set of simulations employs the simple design discussed in Section 2.2:

$$u_t(\theta) = (y_t - \theta)^2 - 1 : \quad t = 1, \ldots, n,$$

where $y_t \sim iidN(0, 1)$ and $x_t \sim iidN(0, 1)$ for $t = 1, \ldots, n$, and y is independent of x. The moment condition tested is

$$H_0 : \Pr\{\mathbb{E}(u|x) = 0\} = 1.$$

The null H_0 is correct and identifies the true value $\theta_0 = 0$ of the parameter θ. This design is characterized with first-order local identification failure even though global identification is ensured. Local identification is obtained at second order according to Assumption 2(iii). Here, $k_0 = 1$ since with the instrument $z_{t,1} = 1$, we have $\mathbb{E}\left(\nabla_{\theta\theta}(z_{t,1}u_i(\theta_0))\right) = -2 \neq 0$.

We present results based on 1000 Monte Carlo replications for three specification tests: the traditional J-test, the conditional moment restriction test by Smith (2007) and Tripathi and Kitamura (2003), denoted by S-TK, and the test $\hat{Z}$ based on the $N(0, 1)$ asymptotic approximation. The J-test is performed on the unconditional moment conditions $\mathbb{E}(z_i u_i(\theta)) = 0$ with $z_i' = (1, x_i^2)$. We present results for the J-test under the standard chi-squared asymptotics as well as the mixture of chi-squared distribution proposed in Dovonon and Renault (2013). The conditional moment restriction tests, S-TK and $\hat{Z}$, involve a choice of tuning parameters. The tuning parameters for the S-TK test are the biweight kernel $k(x) = \frac{15}{16}(1 - x^2)^2 1_{\{|x| \leq 1\}}$ for a kernel function, $K^{**} = \frac{1168780}{2263261}$, kernel bandwidth $b_n = n^{-1/4}$ and trimming function $1_{\{|x| \leq S_*\}}$ with $S_* = 1.96$. The test is computed for $\gamma = 1$ which corresponds to the continuously-updated GMM. Finally, the non-linear least squares estimator is used as input parameter estimator. For our $\hat{Z}$ test, the tuning parameters are $\Psi(\cdot) : \mathbb{R} \to [-\pi, \pi]$, $x \mapsto 2\arctan(x)$, series of bounded functions $g_l(\cdot) : [-\pi, \pi] \to [-1, +1]$, $x \mapsto \cos(lx)$ for $l = 1, \ldots k$, and $k = n^{1/5}$.

Table 1. Empirical rejection rates of specification tests under the null (size): simulation design 1.

n	J-test		S-TK	$\hat{Z}$ test
	$\chi^2(1)$	mixt.		
50	5.70	3.60	26.20	10.70
100	6.20	3.80	27.90	7.80
200	6.60	3.70	27.80	8.00
500	7.70	4.50	23.40	6.50
1000	7.00	3.50	19.50	6.20
5000	8.40	4.20	11.80	6.70
10000	8.30	4.70	7.30	4.90

Notes: The nominal level is 5%. 'mixt.' stands for $\frac{1}{2}\chi^2(1) + \frac{1}{2}\chi^2(2)$ with a critical value at the 5% level equal to 5.14. The J-test is performed on the unconditional moment conditions $E(z_i u_i(\theta)) = 0$ with $z_i' = (1, x_i^2)$; S-TK is the conditional moment restriction test by Smith (2007) and Tripathi and Kitamura (2003); and $\hat{Z}$ test is the conditional moment restriction test proposed in this paper.

The results for the empirical size of the tests at the 5% nominal level are reported in Table 1 as the sample size n increases. The S-TK test exhibits substantial overrejections even for fairly large sample sizes. While the standard chi-squared J-test is not asymptotically valid, the chi-squared mixture performs well with only slight underrejections. The $\hat{Z}$ test tends to overreject at small sample sizes (with 7.8% rejections at the 5% significance level) but approaches the nominal level as n grows.

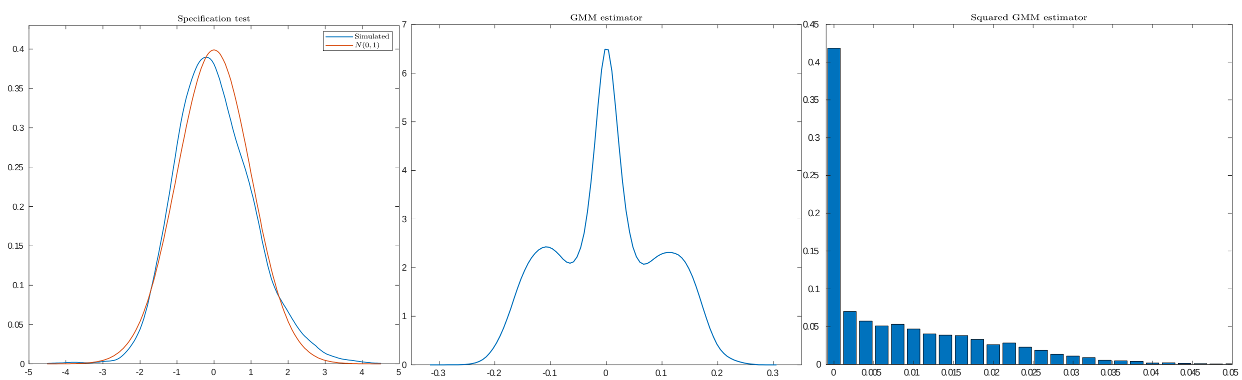

Figure 1: The left graph presents the Monte Carlo kernel density and the limit $N(0,1)$ density for the specification test. The middle and right graphs plot the Monte Carlo kernel density and histogram for the GMM estimator and the squared GMM estimator, respectively. The sample size is $n = 10,000$ and the number of Monte Carlo replications is 10,000.

To visualize the quality of the $N(0,1)$ approximation for the $\hat{Z}$ test, the left panel of Figure 1 plots the kernel density obtained from 10,000 replications and $n = 10,000$ against the $N(0,1)$ density. The simulated density appears to be very close to the limit distribution. The middle plot in Figure

1 presents the simulated kernel density[7] for the GMM estimator of θ while the right plot provides the histogram for the squared GMM estimator of θ. The histogram for the squared GMM estimator nicely illustrates the half probability mass at zero suggested by theory (Theorem 3.2). In addition, the simulated kernel density for $\hat{\theta}$ reveals some features in the shape of the distribution that are not immediately evident in the histogram for $\hat{\theta}^2$.

The second simulation design is tailored to the common CH factor discussed in Section 2.2 and focus only on the properties of the $\hat{Z}$ test. The data generating process has the form:

$$Y_{t+1} = \Lambda f_{t+1} + e_{t+1}, \tag{20}$$

where Y_{t+1} and $e_{t+1} \sim iidN(0, \kappa I_m)$ are $m \times 1$ vectors and f_{t+1} is an $(m-1) \times 1$ vector. The i-th component $f_{i,t+1}$ of f_{t+1} follows a GARCH(1,1) process:

$$f_{i,t+1} = \sigma_{i,t+1}\varepsilon_{i,t+1}, \qquad \sigma^2_{i,t+1} = \omega_{i,0} + \omega_{i,1} f^2_{i,t} + \omega_{i,2}\sigma^2_{i,t}, \tag{21}$$

with $\omega_{i,0}, \omega_{i,1}, \omega_{i,2} > 0$ and $\varepsilon_{i,t+1} \sim iidN(0,1)$. Bougerol and Picard (1992) derive the conditions for strict stationarity of GARCH processes.

We consider two cases: (i) $m = 2$, bivariate Y_{t+1} with a single common CH factor, and (ii) $m = 3$, trivariate Y_{t+1} with two common CH factors. In both cases, we set $\kappa = 0.1$ and $\omega_{i,0} = 1 - \omega_{i,1} - \omega_{i,2}$ for $i = 1, 2$. For case (i), $\Lambda = \begin{pmatrix} 1 & 0.5 \end{pmatrix}'$, $\omega_{1,1} = 0.2$ and $\omega_{1,2} = 0.6$. For case (ii), $\Lambda = \begin{pmatrix} 1 & 1 & 0.5 \\ 0 & 1 & 0.5 \end{pmatrix}'$, with the first CH factor having the same GARCH parameters as in case (i) and the second CH factor having GARCH parameters $\omega_{2,1} = 0.4$ and $\omega_{2,2} = 0.4$. We use Y_t as conditioning variables in constructing the functions $g_l(\cdot)$, $l = 1, \dots, k$.

The empirical rejection probabilities of the $\hat{Z}$ test for $n = 2,000$, $5,000$ and $10,000$ and $1,000$ Monte Carlo replications are presented in Table 2, with the empirical size of the test reported in the top panel. For both cases, (i) and (ii), the rejections of the test are close to the nominal levels. Since in these simulations $k = n^{1/5}$ (by rounding up k to the nearest integer), it appears that further improvements can be obtained if we set $k = const \cdot n^{1/5}$, where the constant $const$ is calibrated to the particular setup (values of n and m).

The third simulation design assesses the power of the $\hat{Z}$ test in the context of model (20)-(21). We again consider two cases but with the following modifications: (i) $m = 2$, bivariate Y_{t+1} with two CH factors, and (ii) $m = 3$, trivariate Y_{t+1} with three CH factors. In both cases, we set Λ to be the identity matrix. For case (i), $\omega_{1,1} = 0.2$ and $\omega_{1,2} = 0.6$, and $\omega_{2,1} = 0.4$ and $\omega_{2,2} = 0.4$. For

[7] For the kernel density estimation, we use the `ksdensity` function in MATLAB.

case (ii), the first two CH factors have the same GARCH parameters as in case (i) and the third CH factor has GARCH parameters $\omega_{3,1} = 0.1$ and $\omega_{3,2} = 0.8$. As in the previous design, we set $\kappa = 0.1$, $\omega_{i,0} = 1 - \omega_{i,1} - \omega_{i,2}$ (for $i = 1, 2, 3$) and use Y_t as conditioning variables in constructing the functions $g_l(\cdot)$, $l = 1, \ldots, k$. The empirical power of the $\hat{Z}$ test is presented in the bottom panel of Table 2. In case (i), the rejection rates under the alternative quickly reach levels close to 100%. In case (ii), the power of the test increases more slowly but it is at 100% for $T = 10,000$. Overall, provided that the sample size is reasonably large, the empirical size and power properties of the proposed specification test are quite satisfactory.

Table 2. Empirical rejection rates of specification test $\hat{Z}$ under the null (size) and alternative (power).

n	10%	5%	10%	5%
	$m = 2$		$m = 3$	
	simulation design 2: size			
2000	8.00	3.30	11.30	3.40
5000	8.90	4.00	9.20	3.50
10000	9.80	4.90	12.80	5.40
	simulation design 3: power			
2000	98.30	97.60	39.90	26.90
5000	100.00	100.00	95.50	91.60
10000	100.00	100.00	100.00	100.00

Notes: The nominal level is 5% and 10%. For simulation design 2 (size), $m = 2$ corresponds to a bivariate y_{t+1} with a single common CH factor, and Case $m = 3$ corresponds to a trivariate Y_{t+1} with two common CH factors. For simulation design 3 (power), $m = 2$ corresponds to a bivariate y_{t+1} with two CH factors, and Case $m = 3$ corresponds to a trivariate Y_{t+1} with three CH factors.

5.2 Empirical application

In this section, we investigate the presence of a common CH factor in U.S. bond returns of different maturities. Engle *et al.* (1990) argued that the CH factor model provides a parsimonious approximation of the covariance structure of excess asset returns. After presenting some preliminary evidence on commonality in the GARCH-based volatility dynamics in bond returns, we subject these portfolio returns to the test of common CH features which amounts to testing the validity of a version of the conditional moment restriction $\mathbb{E}\left(u(y_t, \theta_0)|x_t\right) = 0$.

Let $rx_{t+1}^{(j)}$ denote the holding return, between periods t and $t+1$, on a bond with j years to maturity, in excess of the risk-free rate. Let $Y_{t+1} = (rx_{t+1}^{(1)}, ..., rx_{t+1}^{(N)})'$. As in Section 2.2, we posit that the N-vector of excess bond returns Y_{t+1}, adapted to the increasing filtration $\mathcal{F}_t$, admits a common factor representation:

$$Y_{t+1} = \mu_t + \Lambda f_{t+1} + e_{t+1},$$

where the errors e_{t+1} satisfy $\mathbb{E}(e_{t+1}|\mathcal{F}_t) = 0$ and $\text{Var}(e_{t+1}|\mathcal{F}_t) = \Omega$, and f_{t+1} is a vector of r ($r < N$) common factors that satisfy $\text{Cov}(e_{t+1}, f_{t+1}|\mathcal{F}_t) = 0$, $\mathbb{E}(f_{t+1}|\mathcal{F}_t) = 0$, and $\text{Var}(f_{t+1}|\mathcal{F}_t) = D_t = \text{Diag}(\sigma^2_{1,t+1}, \ldots, \sigma^2_{r,t+1})$.[8] Then, Y_{t+1} is characterized by $(N - r)$ time invariant CH common features if its conditional covariance matrix is given by

$$\text{Var}(Y_{t+1}|\mathcal{F}_t) = \Lambda D_t \Lambda' + \Omega.$$

This implies that there exists a vector $\beta \neq 0_N$ in $\mathbb{R}^N$ such that $\mathbb{E}\left((\beta' Y_{t+1})^2|\mathcal{F}_t\right)$ is constant.

In the empirical analysis, we use the Fama bond portfolio returns from CRSP with the following maturities: 1 to 2 years, 2 to 3 years, 3 to 4 years, 4 to 5 years, and 5 to 10 years. The data is at monthly frequency covering the period January 1952 – December 2020. We construct excess bond returns by subtracting the one-month risk-free rate, obtained from CRSP.

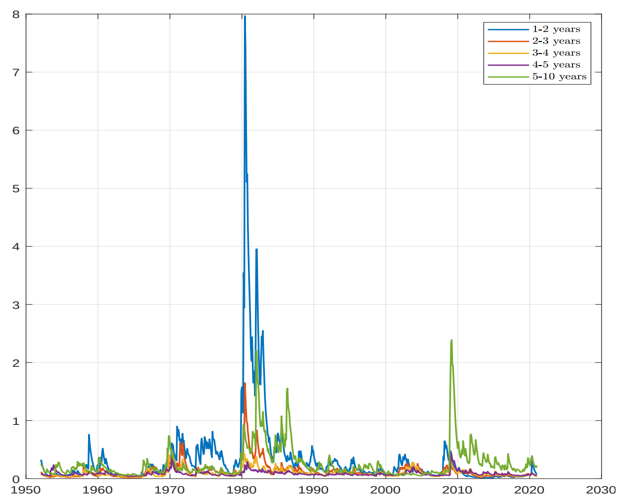

Figure 2: Estimated GARCH(1,1) volatilities for portfolio bond excess returns of different maturities.

We start by fitting a GARCH(1,1) to each of these excess bond returns. The filtered GARCH volatilities are plotted in Figure 2. As the graph reveals, there appears to be a strong co-movement in these GARCH volatilities. This is probably not too surprising since the first principal component in these five bond returns explains in excess of 95% of their volatility.

Crump and Gospodinov (2022) argue that the spread or cross-sectionally differenced returns, $dr^{(j)}_{t+1} = rx^{(j)}_{t+1} - rx^{(j-1)}_{t+1}$, reveal better the underlying factor structure since the term-structure identi-

[8] The conventional term structure models impose no-arbitrage restrictions on the factor loading matrix Λ. Recent research (Duffee, 2011; Joslin *et al.*, 2011; among others) casts doubt on the role of no-arbitrage restrictions in modeling and forecasting bond yields. The forecasting properties are further deteriorated by incorporating stochastic volatility. As Joslin and Le (2021) demonstrate, this is largely attributed to the fact that these models impose a tight link between risk compensation and interest rate volatility, and recommend the use of unrestricted factor models. This is the approach that we follow here.

ties induce elevated local correlations in $rx_{t+1}^{(j)}$ across maturities that obscure the true signal.[9] For this reason, we use the vector of differenced returns $Y_{t+1} = (rx_{t+1}^{(1)}, dr_{t+1}^{(2)} ..., dr_{t+1}^{(N)})'$ which brings down the explained variation by the first principal component to 67%: more muted than that for excess bond returns but still large enough to suggest commonality in bond return dynamics.

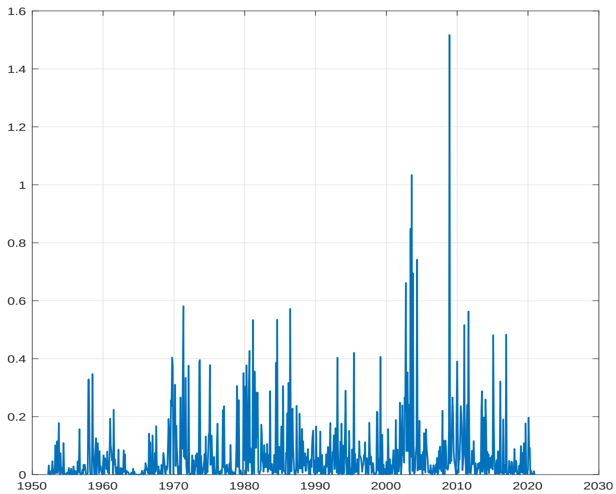

Figure 3: Plot of $u_{t+1}(\hat{\beta}, \hat{c}) := (\hat{\beta}' Y_{t+1})^2 - \hat{c}$, where the estimates are based on $\mathbb{E}[u_{t+1}(\beta, c)|\mathcal{F}_t)] = 0$.

We estimate the parameters by the two-step efficient GMM, based on the conditional restriction $\mathbb{E}(u_{t+1}(\theta_0)|\mathcal{F}_t) = 0$, where $u_{t+1}(\beta, c) := (\beta_1 Y_{1,t+1} + ... + \beta_4 Y_{4,t+1} + (1 - \beta_1 - \cdots - \beta_4) Y_{5,t+1})^2 - c$. We use Y_t as a vector of conditioning variables and the choices of $g_l(\cdot)$ and k are as specified in the previous section. The obtained estimates for $\hat{\beta}$ are $(-0.2087, 0.1913, 0.6420, 0.4376, -0.0622)'$ which, interestingly, produce a tent-shaped pattern as in Cochrane and Piazzesi (2005). Figure 3 plots $(\hat{\beta}_1 Y_{1,t+1} + ... + \hat{\beta}_4 Y_{4,t+1} + (1 - \hat{\beta}_1 - ... - \hat{\beta}_4) Y_{5,t+1})^2$. While this graph reveals that the strong CH structure, observed in individual series, is largely destroyed, we subject this hypothesis to a formal test using our $\hat{Z}$ test statistic. The value of the test is -0.0839 with a p-value of 0.9331 suggesting that the null $H_0 : \mathbb{E}\left((\beta' Y_{t+1})^2 - c|\mathcal{F}_t\right) = 0$ cannot be rejected. This result could have important implications for portfolio allocation and hedging.

6 Conclusions

Economic models are often defined by a set of conditional moment restrictions that can be used to assess the degree of misspecification or the validity of particular economic theory. It is possible, however, that while the model remains globally identified, it suffers from first-order local identification failure.

[9] When the maturity matches the frequency of the data, the differenced returns have the interpretation of returns on a forward trade (see Crump and Gospodinov, 2022).

This setup is the focus of our theoretical analysis. First, we derive the rate of convergence of the GMM estimator with an expanding number of moment conditions under the lack of first-order local identification. This rate of convergence is shown to be faster than the case with a fixed number of restrictions. Unlike the standard case, the contribution of the increasing number of moment restrictions translates into efficiency gain that drives the faster rate of convergence. We also characterize the asymptotic distribution of the estimator which is non-standard and has a point mass at zero. Finally, we establish the asymptotic normality of the conditional moment restriction test in our setup. Importantly, this result is obtained for time series data resorting to central limit theorems for degenerate U-statistics of weakly dependent processes. The finite-sample properties of the test are illustrated using simulated data and the proposed testing framework is applied to study the presence of common (conditionally heteroskedastic) features in bond portfolio returns.

The validity of the standard normal limit for the specification test is established for large samples. For empirical problems with limited sample sizes, finite-sample improvements based on subsampling or resampling methods are often desirable. However, such finite-sample refinements may be difficult to develop and implement due to the highly challenging nature of our setup: a conditional moment restrictions model with local first-order local identification failure, second-order local identification and dependent data. While the analysis in this paper provides some guidance on how one could design asymptotically valid methods with improved finite-sample properties, such an extension proves to be highly nontrivial. This is a fruitful direction for future research.

Appendix A: Consistency of the GMM estimator

This Appendix establishes the consistency of the GMM estimator with an increasing number of moment conditions. The general result in Theorem A.1 is then specialized to the two-step efficient estimator by Corollary A.2. Assumption A.1 provides sufficient conditions for Assumption 3(ii) to hold as established by Lemma B.2.

Theorem A.1 *Under Assumptions 1, 2(i) and 3, if $k/n \to 0$ as $n \to \infty$, then the GMM estimator $\hat{\theta}$ – defined by* (14) *– converges in probability to θ_0.*

The following assumption pertains to the consistency of the two-step efficient GMM estimator.

Assumption A.1 (Consistency of two-step GMM)

(i) $\tilde{\theta} - \theta_0 = O_P(r_n)$, with $\tilde{\theta}$ the first-step GMM estimator of θ_0 and $r_n \to 0$ as $n \to \infty$.

(ii) $\mathbb{E}\left[u(y,\theta_0)^{2\delta}|g_h(x)|^\delta|g_l(x)|^\delta\right] \le \Delta < \infty$*, for some* $\delta > 2$ *and an absolute constant* $\Delta > 0$*.*

(iii) For each y*, the function* $\theta \mapsto u(y,\theta)$ *is continuously differentiable in a neighborhood* $\mathcal{N}$ *of* θ_0 *and* $\mathbb{E}\left(g_l(x)^2 \sup_{\theta\in\mathcal{N}} u(y,\theta)^2\right) \le \Delta < \infty$ *and* $\mathbb{E}\left(g_l(x)^2 \sup_{\theta\in\mathcal{N}} \|\nabla_\theta u(y,\theta)\|^2\right) \le \Delta < \infty$*.*

(iv) Let $\bar\lambda_k := \lambda_{\max}(V_k)$ *and* $\lambda_k := \lambda_{\min}(V_k)$*. There exists* $\underline{\lambda} > 0$*:* $\lambda_k \ge \underline{\lambda}$ *and* $\bar\lambda_k/\lambda_k \le \Delta < \infty$*.*

(v) $k[r_n \vee n^{-1/2}] \to 0$ *as* $n \to \infty$*.*

Corollary A.2 *If Assumptions 1, 2(i), 3(i), and A.1 hold, then the two-step GMM estimator, defined with* $\hat W_k = \hat V_k^{-1}$*, is consistent.*

Proof of Theorem A.1: Let $u_t := u(y_t,\theta_0)$. Note that

$$\mathbb{E}(\bar f_k(\theta_0)'\bar f_k(\theta_0)) = \frac{1}{n}\sum_{t,s=1}^n \mathbb{E}\left[g^{(k)}(x_t)'g^{(k)}(x_s)u_tu_s\right] = \frac{1}{n}\sum_{t=1}^n \mathbb{E}\left[g^{(k)}(x_t)'g^{(k)}(x_t)u_t^2\right]$$
$$= \mathbb{E}\left[g^{(k)}(x_t)'g^{(k)}(x_t)u_t^2\right] \le k\cdot \max_{1\le l\le k}\mathbb{E}[g_l(x_t)^2u_t^2].$$

Since $\max_{1\le l\le k}\mathbb{E}[g_l(x_t)^2u_t^2]$ is bounded, we can claim that $\bar f_k(\theta_0)'\bar f_k(\theta_0) = O_P(k)$. Also, by definition,

$$\lambda_{\min}(\hat W_k)\frac{1}{n}\bar f_{k_0}(\hat\theta)'\bar f_{k_0}(\hat\theta) \le \lambda_{\min}(\hat W_k)\frac{1}{n}\bar f_k(\hat\theta)'\bar f_k(\hat\theta) \le \frac{1}{n}\bar f_k(\hat\theta)'\hat W_k\bar f_k(\hat\theta)$$
$$\le \frac{1}{n}\bar f_k(\theta_0)'\hat W_k\bar f_k(\theta_0) \le \lambda_{\max}(\hat W_k)\frac{1}{n}\bar f_k(\theta_0)'\bar f_k(\theta_0).$$

In particular,

$$\frac{1}{n}\bar f_{k_0}(\hat\theta)'\bar f_{k_0}(\hat\theta) \le \frac{\lambda_{\max}(\hat W_k)}{\lambda_{\min}(\hat W_k)}\frac{1}{n}\bar f_k(\theta_0)'\bar f_k(\theta_0) = \frac{\lambda_{\max}(\hat W_k)}{\lambda_{\max}(W_k)}\frac{\lambda_k}{\lambda_{\min}(\hat W_k)}\frac{\lambda_{\max}(W_k)}{\lambda_k}\frac{1}{n}\bar f_k(\theta_0)'\bar f_k(\theta_0).$$

By Assumption 3(ii), we can therefore claim that:

$$\frac{1}{n}\bar f_{k_0}(\hat\theta)'\bar f_{k_0}(\hat\theta) := \left\|\frac{1}{n}\sum_{t=1}^n g^{(k_0)}(x_t)u_t(\hat\theta)\right\|_2^2 = O_P(k/n) = o_P(1).$$

Thus, by Assumption 3(i), we can claim that $\mathbb{E}(g^{(k_0)}(x_t)u(y_t,\hat\theta)) \xrightarrow{P} 0$. We shall deduce that $\hat\theta$ converges in probability to θ_0 by the following standard argument (see, e.g., Newey and McFadden, 1994). Let $\mathcal{N}$ be an open neighborhood of θ_0. By continuity of $\theta \mapsto \mathbb{E}(g^{(k_0)}(x_t)u(y_t,\theta))$ and compactness of $\Theta\backslash\mathcal{N}$,

$$\inf_{\theta\in\Theta\backslash\mathcal{N}}\|\mathbb{E}(g^{(k_0)}x_t)(u(y_t,\theta))\|_2 = \|\mathbb{E}(g^{(k_0)}(x_t)u(y_t,\theta_*))\|_2 = \epsilon$$

with $\epsilon \ne 0$ because $\Theta\setminus\mathcal{N} \ni \theta_* \ne \theta_0$. Since $\mathbb{E}(g^{(k_0)}(x_t)u(y_t,\hat\theta)) = o_P(1)$,

$$\Pr\left(\|\mathbb{E}(g^{(k_0)}(x_t)u(y_t,\hat\theta))\|_2 < \epsilon/2\right) \to 1.$$

That is, $\Pr(\hat{\theta} \in \mathcal{N}) \to 1$ and this concludes the proof. $\square$

Proof of Corollary A.2: Assumptions 1, and A.1(ii, iii) ensure that the orders of magnitude derived by Lemma B.2 apply and the conditions in Assumption 3(ii) follow from Assumption A.1(iv, v). $\square$

Appendix B: Preliminary lemmas and proofs of main results

B.1 Useful lemmas

Let $u_t(\theta) = u(y_t, \theta)$, $\nabla_\theta \bar{f}_k(\theta) = \partial \bar{f}_k(\theta)/\partial\theta'$, $\nabla_\theta u_t(\theta) = \partial u_t(\theta)/\partial\theta'$, $\nabla_{\theta\theta} u_t(\theta) = \left[\text{vec}\left(\partial^2 u_t(\theta)/\partial\theta\partial\theta'\right)\right]'$,

$$\bar{D}_1(\theta) := \frac{1}{n}\sum_{t=1}^{n} g^{(k)}(x_t)\nabla_\theta u_t(\theta) R_1, \quad D_1 := \mathbb{E}[g^{(k)}(x_t)\nabla_\theta u_t(\theta_0) R_1],$$

$$\bar{H}^{(k)}(\theta) = \frac{1}{n}\sum_{t=1}^{n} g^{(k)}(x_t)\nabla_{\theta\theta} u_t(\theta), \text{ and } H := H^{(k)}(\theta_0) = \mathbb{E}[g^{(k)}(x_t)\nabla_{\theta\theta} u_t(\theta_0)].$$

Lemma B.1 *Let $\bar{\theta}$ be a random sequence converging in probability to θ_0 as n grows. If $\theta \mapsto U_t(\theta)$ is a random $\mathbb{R}^m$-valued function satisfying Condition C in the main text. Then,*

$$\left\|\frac{1}{n}\sum_{t=1}^{n} g^{(k)}(x_t) U_t(\bar{\theta})' - \mathbb{E}\left(g^{(k)}(x_t) U_t(\theta_0)'\right)\right\|_2 = O_P\left(\sqrt{k}\cdot(n^{-1/2} \vee \|\bar{\theta} - \theta_0\|_2)\right).$$

Lemma B.2 *Let $\bar{\theta}$ be a sequence of estimators converging in probability to θ_0 and $v_n = k(n^{-1/2} \vee \|\bar{\theta} - \theta_0\|_2)$. Also, let $\lambda_k := \lambda_{\min}(V_k)$ and $\bar{V}_k$ be defined by*

$$\bar{V}_k = \frac{1}{n}\sum_{t=1}^{n} f_k(x_t, y_t, \bar{\theta}) f_k(x_t, y_t, \bar{\theta})'.$$

If Assumptions 1 and A.1(ii, iii) hold, then

(i) $\|\bar{V}_k - V_k\|_2 = O_P(v_n)$, *(ii)* $|\lambda_{\min}(\bar{V}_k) - \lambda_k| = O_P(v_n)$, *(iii)* $|\lambda_{\min}(\bar{V}_k)\lambda_k^{-1} - 1| = O_P(\lambda_k^{-1} v_n)$,

(iv) $\|\bar{V}_k^{-1} - V_k^{-1}\|_2 = \frac{O_P(\lambda_k^{-2} v_n)}{1 + O_P(\lambda_k^{-1} v_n)}$.

(v) If, in addition, $\lambda_{\max}(V_k)/\lambda_{\min}(V_k) = O(1)$, then $|\lambda_{\max}(\bar{V}_k)\lambda_{\max}(V_k)^{-1} - 1| = O_P(\lambda_k^{-1} v_n)$.

Lemma B.3 *Let $\bar{\theta}$ and $\hat{\theta}$ be two sequences of estimators converging to θ_0 in probability. Let $\hat{W}_k$ be a sequence of (k,k)-positive-definite weighting matrices and W_k be a (k,k)-symmetric positive definite matrix such that the eigenvalues of $\hat{W}_k$ and W_k satisfy Assumption 3(ii). Assume that the functions*

$\theta \mapsto \nabla_\theta u_i(\theta) \cdot R_1$ and $\theta \mapsto \nabla_{\theta\theta} u_i(\theta)$ satisfy Condition C in the main text. Furthermore, let $\bar{D}_1 := \bar{D}_1(\bar{\theta})$, $\hat{D}_1 := \bar{D}_1(\hat{\theta})$, $\bar{D}_2 := \bar{D}_2(\theta_0)$, $\bar{H} := \bar{H}^{(k)}(\bar{\theta})$, $H := H^{(k)}(\theta_0)$, $\bar{M}^{(k)} := I_k - \bar{P}^{(k)}$with $\bar{P}^{(k)} := \hat{W}_k^{1/2}\bar{D}_1\left(\bar{D}_1'\hat{W}_k\bar{D}_1\right)^{-1}\bar{D}_1'\hat{W}_k^{1/2}$, $M^{(k)} := I_k - P^{(k)}$, with $P^{(k)} := W_k^{1/2}D_1\left(D_1'W_kD_1\right)^{-1}D_1'W_k^{1/2}$, $\bar{f}_k := \bar{f}_k(\theta_0)$, and $\hat{M}^{(k)}$ defined as $\bar{M}^{(k)}$ but using $\hat{D}_1$. Finally, let

$$\Delta_{1n} = \bar{H}'\hat{W}_k^{1/2}\bar{M}^{(k)}\hat{W}_k^{1/2}\bar{H} - H'W_k^{1/2}M^{(k)}W_k^{1/2}H \text{ and } \Delta_{2n} = \bar{H}'\hat{W}_k^{1/2}\bar{M}^{(k)}\hat{W}_k^{1/2}\bar{f}_k - H'W_k^{1/2}M^{(k)}W_k^{1/2}\bar{f}_k.$$

If there exist $\alpha_1, \alpha_2 > 0$ such that

$$\alpha_1\sqrt{k} \leq \|D_1\|_2 \leq \alpha_2\sqrt{k}, \quad \alpha_1\sqrt{k} \leq \|H\|_2 \leq \alpha_2\sqrt{k}, \quad \lambda_{\max}(\bar{D}_2'\bar{D}_2) = O_P(k)$$

and

$$\lambda_{\max}(D_1'D_1)/\lambda_{\min}(D_1'D_1) \leq \alpha_2, \text{ and } \|\bar{f}_k\|_2 = O_P(\sqrt{k}),$$

then:

(i) $\|\bar{D}_2'\hat{W}_k\bar{D}_2\|_2 = O_P(\bar{\lambda}_k k)$, *(ii)* $\|\hat{W}_k^{1/2}\bar{f}_k\|_2 = O_P(\sqrt{\bar{\lambda}_k k})$, *(iii)* $\|\hat{W}_k^{1/2}\bar{H}\|_2 = O_P(\sqrt{\bar{\lambda}_k k})$,

(iv) $\|(\bar{D}_1'\hat{W}_k\bar{D}_1)^{-1}\|_2 = O_P(\bar{\lambda}_k^{-1}k^{-1})$, *(v)* $\|\bar{M}^{(k)} - M^{(k)}\|_2 = O_P(\bar{\lambda}_k^{-1}\|\hat{W}_k - W_k\|_2) + O_P(n^{-1/2} \vee \|\bar{\theta} - \theta_0\|_2)$,

(vi) $\|\Delta_{1n}\|_2 = O_P(k\|\hat{W}_k - W_k\|_2) + O_P(\bar{\lambda}_k k[n^{-1/2} \vee \|\bar{\theta} - \theta_0\|_2])$,

(vii) $\|\Delta_{2n}\|_2 = O_P(k\bar{\lambda}_k[n^{-1/2} \vee \|\bar{\theta} - \theta_0\|_2]) + O_P(k\|\hat{W}_k - W_k\|_2)$,

(viii) If $\bar{\theta} \in (\theta_0, \hat{\theta})$, we have $\|\hat{M}^{(k)} - \bar{M}^{(k)}\|_2 = O_P(n^{-1/2} \vee \|\hat{\theta} - \theta_0\|_2)$.

The proofs of Lemmas B.1, B.2 and B.3 are provided in the Online Appendix.

B.2 Proofs of main results

Proof of Proposition 2.2: (i) It suffices to show that (8) holds for k_0 to claim that it holds for all $k \geq k_0$. Since $\alpha_l \equiv 0$ for all $l \geq k_0$, we have

$$\mathbb{E}(u(y,\theta)|x) := \rho(x,\theta) = \sum_{l=1}^{k_0-1}\alpha_l(\theta)g_l(x) \text{ and } [\rho(x,\theta) \equiv 0] \Leftrightarrow [\alpha_l = 0, \forall l = 1,\ldots,k_0-1].$$

By the law of iterated expectations, $\alpha_l(\theta) = \mathbb{E}(g_l(x)u(y,\theta)) = 0, \forall l = 1,\ldots,k_0-1$ and this establishes the claim since $[\rho(x,\theta) \equiv 0 \Leftrightarrow \theta = \theta_0]$ holds by assumption.

To establish the second claim, recall that $\rho(x,\theta) = \sum_{l=1}^{\infty} \alpha_l(\theta) g_l(x)$. Also, by the law of iterated expectations, $\mathbb{E}(g^{(k)}(x)u(y,\theta)) = \mathbb{E}(g^{(k)}(x)\rho(x,\theta))$ so that

$$\mathbb{E}(g^{(k)}(x)u(y,\theta)) = (\alpha_1(\theta), \ldots, \alpha_k(\theta))'.$$

Hence, by the definition of θ_k,

$$\mathbb{E}(\rho(x,\theta_k))^2 = \mathbb{E}\left(\sum_{l \geq k+1} \alpha_l(\theta_k) g_l(x)\right)^2 = \sum_{l \geq k+1} \alpha_l(\theta_k)^2 \to 0, \text{ as } k \to \infty, \tag{B.1}$$

where the convergence follows from (d).

Consider an arbitrary small and open neighborhood of $\mathcal{N}$ of θ_0 and let $\epsilon = \min_{\Theta \setminus \mathcal{N}} \mathbb{E}(\rho(x,\theta))^2$. By the continuity assumption (c), the compactness of $\Theta \setminus \mathcal{N}$, and the identification property in (3), we can claim that $\epsilon > 0$. Also, from (B.1), it is clear that there exists $k_0 \in \mathbb{N}$ such that $\mathbb{E}[\rho(x,\theta_k)]^2 < \epsilon$ for all $k \geq k_0$. It then follows that for $k \geq k_0$, we have $\theta_k \in \mathcal{N}$ which proves the claim.

(ii) First, we establish the necessary condition. If the first-order local identification condition fails, then

$$\text{Rank}\left[\mathbb{E}\left((\mathbb{E}(\nabla_\theta u(y,\theta_0)|x))'(\mathbb{E}(\nabla_\theta u(y,\theta_0)|x))\right)\right] < p,$$

implying that there exists $\delta \neq 0 \in \mathbb{R}^p$ such that $\mathbb{E}(\nabla_\theta u(y,\theta_0)|x) \cdot \delta = 0$ almost surely. Therefore, for any $k \in \mathbb{N}$,

$$\mathbb{E}\left(g^{(k)}(x) \cdot \nabla_\theta u(y,\theta_0)\right) \cdot \delta = \mathbb{E}\left(g^{(k)}(x) \cdot \mathbb{E}(\nabla_\theta u(y,\theta_0)|x)\right) \cdot \delta = 0.$$

As a result,

$$\text{Rank}\left(\mathbb{E}\left(g^{(k)}(x) \cdot \nabla_\theta u(y,\theta_0)\right)\right) \leq p-1, \forall k.$$

Since $k \mapsto \text{Rank}\left(\mathbb{E}\left(g^{(k)}(x) \cdot \nabla_\theta u(y,\theta_0)\right)\right)$ takes integer values, it is nondecreasing and bounded from above, it reaches its maximum, say $r \leq p-1$, as k increases. This shows the necessary condition.

Next, we establish the sufficient condition. Under the stated condition, there exists $\delta \neq 0$ such that

$$\mathbb{E}\left(g_l(x) \cdot \nabla_\theta u(y,\theta_0)\right) \cdot \delta = \mathbb{E}\left(g_l(x) \cdot \mathbb{E}(\nabla_\theta u(y,\theta_0)|x)\right) \cdot \delta = 0, \text{ for all } l \geq 1. \tag{B.2}$$

Since $\mathbb{E}(\nabla_\theta u(y,\theta_0)|x) \in (L^2(P))^p$, its i-th component can be written as $\sum_{l \geq 1} \alpha_{l,i} g_l(x)$, with $\alpha_{l,i}$'s being scalars. Taking the relevant linear combinations (over l) of the equalities in (B.2), we have

$$\mathbb{E}\left((\mathbb{E}(\nabla_\theta u(y,\theta_0)|x))'(\mathbb{E}(\nabla_\theta u(y,\theta_0)|x))\right) \cdot \delta = 0$$

and this completes the proof. $\square$

Proof of Theorem 3.1: Let $R = (R_1 \,|\, R_2)$ and consider the transformation $\theta = R\eta := R_1\eta_1 + R_2\eta_2$, with $\theta, \eta \in R^p$, $\eta_1 \in R^r$ and $\eta_2 \in R^{p-r}$, and set $\hat{\theta} = R\hat{\eta}$, and $\theta_0 = R\eta_0$. Hence, $\bar{f}_k(\hat{\theta}) = \bar{f}_k(R\hat{\eta}) = \bar{f}_k(R_1\hat{\eta}_1 + R_2\hat{\eta}_2)$. By a first-order Taylor expansion of $\eta_1 \mapsto \bar{f}_k(R_1\eta_1 + R_2\hat{\eta}_2)$ around η_{01} and a second-order Taylor expansion of $\eta_2 \mapsto \bar{f}_k(R_1\eta_{01} + R_2\eta_2)$ around η_{02}, we have

$$\bar{f}_k(\hat{\theta}) = \bar{f}_k(\theta_0) + \frac{1}{\sqrt{n}}\nabla_\theta \bar{f}_k(R_1\bar{\eta}_1 + R_2\hat{\eta}_2)R_1\sqrt{n}(\hat{\eta}_1 - \eta_{01}) + \nabla_\theta \bar{f}_k(\theta_0)R_2(\hat{\eta}_2 - \eta_{02}) + \frac{1}{2}\bar{H}^{(k)}(\bar{\theta}) \cdot \sqrt{n} \cdot \text{vec}(R_2(\hat{\eta}_2 - \eta_{02})(\hat{\eta}_2 - \eta_{02})'R_2'),$$

where $\bar{\eta}_1 \in (\eta_{01}, \hat{\eta}_1)$ and $\bar{\theta} \in (\theta_0, \hat{\theta})$ and both may differ from row to row.
Let $\tilde{\theta} = R_1\bar{\eta}_1 + R_2\hat{\eta}_2$, $\bar{D}_1 = \frac{1}{\sqrt{n}}\nabla_\theta \bar{f}_k(\tilde{\theta}) \cdot R_1$, $\bar{D}_2 = \nabla_\theta \bar{f}_k(\theta_0)R_2$, $z_{0n} = \sqrt{n} \cdot vec(R_2(\tilde{\eta}_2 - \eta_{02})(\tilde{\eta}_2 - \eta_{02})'R_2')$ and we write

$$\bar{f}_k(\hat{\theta}) = \bar{f}_k(\theta_0) + \bar{D}_1\sqrt{n}(\hat{\eta}_1 - \eta_{01}) + \bar{D}_2(\hat{\eta}_2 - \eta_{02}) + \frac{1}{2}\bar{H}^{(k)}(\bar{\theta})z_{0n}. \tag{B.3}$$

By pre-multiplying this equation by $\bar{D}_1'\hat{W}_k$ and solving for $\sqrt{n}(\hat{\eta}_1 - \eta_{01})$, we obtain

$$\sqrt{n}(\hat{\eta}_1 - \eta_{01}) = -\left(\bar{D}_1'\hat{W}_k\bar{D}_1\right)^{-1}\bar{D}_1'\hat{W}_k\left(\bar{f}_k(\theta_0) - \bar{f}_k(\hat{\theta}) + \bar{D}_2(\hat{\eta}_2 - \eta_{02}) + \frac{1}{2}\bar{H}^{(k)}(\bar{\theta})z_{0n}\right). \tag{B.4}$$

Plugging this back into (B.3), we have

$$\bar{M}^{(k)}\hat{W}_k^{1/2}\bar{f}_k(\hat{\theta}) = \bar{M}^{(k)}\hat{W}_k^{1/2}\bar{f}_k(\theta_0) + \bar{M}^{(k)}\hat{W}_k^{1/2}\bar{D}_2(\hat{\eta}_2 - \eta_{02}) + \frac{1}{2}\bar{M}^{(k)}\hat{W}_k^{1/2}\bar{H}^{(k)}(\bar{\theta})z_{0n}, \tag{B.5}$$

with $\bar{M}^{(k)} = I_k - \bar{P}^{(k)}$ and $\bar{P}^{(k)} = \hat{W}_k^{1/2}\bar{D}_1\left(\bar{D}_1'\hat{W}_k\bar{D}_1\right)^{-1}\bar{D}_1'\hat{W}_k^{1/2}$. Then, multiplying each side of (B.5) by its own transpose and rearranging yields

$$\begin{aligned}\frac{1}{4}z_{0n}'\bar{H}^{(k)}(\bar{\theta})'\hat{W}_k^{1/2}\bar{M}^{(k)}\hat{W}_k^{1/2}\bar{H}^{(k)}(\bar{\theta})z_{0n} &= (\bar{f}_k(\hat{\theta})'\hat{W}_k^{1/2}\bar{M}^{(k)}\hat{W}_k^{1/2}\bar{f}_k(\hat{\theta}) - \bar{f}_k(\theta_0)'\hat{W}_k^{1/2}\bar{M}^{(k)}\hat{W}_k^{1/2}\bar{f}_k(\theta_0)) \\ &- (\hat{\eta}_2 - \eta_{02})'\bar{D}_2'\hat{W}_k^{1/2}\bar{M}^{(k)}\hat{W}_k^{1/2}\bar{D}_2(\hat{\eta}_2 - \eta_{02}) - 2\bar{f}_k(\theta_0)'\hat{W}_k^{1/2}\bar{M}^{(k)}\hat{W}_k^{1/2}\bar{D}_2(\hat{\eta}_2 - \eta_{02}) \\ &- \bar{f}_k(\theta_0)'\hat{W}_k^{1/2}\bar{M}^{(k)}\hat{W}_k^{1/2}\bar{H}^{(k)}(\bar{\theta})z_{0n} - (\hat{\eta}_2 - \eta_{02})'\bar{D}_2'\hat{W}_k^{1/2}\bar{M}^{(k)}\hat{W}_k^{1/2}\bar{H}^{(k)}(\bar{\theta})z_{0n}.\end{aligned}$$

By definition, $\bar{f}_k(\hat{\theta})'\hat{W}_k\bar{f}_k(\hat{\theta}) \leq \bar{f}_k(\theta_0)'\hat{W}_k\bar{f}_k(\theta_0)$. Hence,

$$\begin{aligned}\frac{1}{4}z_{0n}'\bar{H}^{(k)}(\bar{\theta})'\hat{W}_k^{1/2}\bar{M}^{(k)}\hat{W}_k^{1/2}\bar{H}^{(k)}(\bar{\theta})z_{0n} &= (\bar{f}_k(\hat{\theta})'\hat{W}_k^{1/2}\bar{P}^{(k)}\hat{W}_k^{1/2}\bar{f}_k(\hat{\theta}) - \bar{f}_k(\theta_0)'\hat{W}_k^{1/2}\bar{P}^{(k)}\hat{W}_k^{1/2}\bar{f}_k(\theta_0)) \\ &- (\hat{\eta}_2 - \eta_{02})'\bar{D}_2'\hat{W}_k^{1/2}\bar{M}^{(k)}\hat{W}_k^{1/2}\bar{D}_2(\hat{\eta}_2 - \eta_{02}) - 2\bar{f}_k(\theta_0)'\hat{W}_k^{1/2}\bar{M}^{(k)}\hat{W}_k^{1/2}\bar{D}_2(\hat{\eta}_2 - \eta_{02}) \\ &- \bar{f}_k(\theta_0)'\hat{W}_k^{1/2}\bar{M}^{(k)}\hat{W}_k^{1/2}\bar{H}^{(k)}(\bar{\theta})z_{0n} - (\hat{\eta}_2 - \eta_{02})'\bar{D}_2'\hat{W}_k^{1/2}\bar{M}^{(k)}\hat{W}_k^{1/2}\bar{H}^{(k)}(\bar{\theta})z_{0n}.\end{aligned}$$

We show in the Online Appendix that

$$|\bar{f}_k(\hat{\theta})'\hat{W}_k^{1/2}\bar{P}^{(k)}\hat{W}_k^{1/2}\bar{f}_k(\hat{\theta}) - \bar{f}_k(\theta_0)'\hat{W}_k^{1/2}\bar{P}^{(k)}\hat{W}_k^{1/2}\bar{f}_k(\theta_0)| = O_P(\bar{\lambda}_k k/\sqrt{n}) + O_P(\bar{\lambda}_k k\|\hat{\theta}-\theta_0\|_2) \quad \text{(B.6)}$$

and, since $\bar{\lambda}_k$ is bounded and $k/\sqrt{n} \to 0$, only the second term matters.

Using this fact and letting $H := H^{(k)}(\theta_0)$ and

$$\begin{aligned}\Delta_{1n} &:= \bar{H}^{(k)}(\bar{\theta})'\hat{W}_k^{1/2}\bar{M}^{(k)}\hat{W}_k^{1/2}\bar{H}^{(k)}(\bar{\theta}) - H'W_k^{1/2}M^{(k)}W_k^{1/2}H,\\ \Delta_{2n} &:= \bar{H}^{(k)}(\bar{\theta})'\hat{W}_k^{1/2}\bar{M}^{(k)}\hat{W}_k^{1/2}\bar{f}_k(\theta_0) - H'W_k^{1/2}M^{(k)}W_k^{1/2}\bar{f}_k(\theta_0),\end{aligned}$$

we can write

$$\begin{aligned}\frac{1}{4}z_{0n}'H'W_k^{1/2}M^{(k)}W_k^{1/2}Hz_{0n} \le O_P(\bar{\lambda}_k k)\|\hat{\eta}-\eta_0\|_2 - (\hat{\eta}_2-\eta_{02})'\bar{D}_2'\hat{W}_k^{1/2}\bar{M}^{(k)}\hat{W}_k^{1/2}\bar{D}_2(\hat{\eta}_2-\eta_{02})\\ - 2\bar{f}_k(\theta_0)'\hat{W}_k^{1/2}\bar{M}^{(k)}\hat{W}_k^{1/2}\bar{D}_2(\hat{\eta}_2-\eta_{02}) - \bar{f}_k(\theta_0)'W_k^{1/2}M^{(k)}W_k^{1/2}Hz_{0n} - \Delta_{2n}'z_{0n}\\ - (\hat{\eta}_2-\eta_{02})'\bar{D}_2'\hat{W}_k^{1/2}\bar{M}^{(k)}\hat{W}_k^{1/2}\bar{H}^{(k)}(\bar{\theta})z_{0n} - \frac{1}{4}z_{0n}'\Delta_{1n}z_{0n}. \quad \text{(B.7)}\end{aligned}$$

From (B.4), we can show that $\|\hat{\eta}_1-\eta_{01}\|_2 = O_P(\|\hat{\eta}_2-\eta_{02}\|_2^2)$ so that $\|\hat{\eta}-\eta_0\|_2 = O_P(\|\hat{\eta}_2-\eta_{02}\|_2)$. Also, from the second-order local identification property, we have

$$\frac{1}{4}z_{0n}'H'W_k^{1/2}M^{(k)}W_k^{1/2}Hz_{0n} \ge \frac{1}{4}\gamma_k\|z_{0,n}\|_2^2 = \frac{1}{4}\gamma_k n\|\hat{\eta}_2-\eta_{02}\|_2^4.$$

Let $z_{1n} = \gamma_k^{1/4}n^{1/4}(\hat{\eta}_2-\eta_{02})$. By the Cauchy-Schwarz inequality, (B.7) yields

$$\begin{aligned}\frac{1}{4}\|z_{1n}\|_2^4 \le \frac{1}{\gamma_k^{1/4}n^{1/4}}O_P(\bar{\lambda}_k k)\|z_{1n}\|_2 + \frac{1}{\sqrt{n\gamma_k}}\|\bar{D}_2'\hat{W}_k\bar{D}_2\|_2\cdot\|z_{1n}\|_2^2 + \frac{2}{(n\gamma_k)^{1/4}}\|\hat{W}_k^{1/2}\bar{f}_k(\theta_0)\|_2\|\hat{W}_k^{1/2}\bar{D}_2\|_2\cdot\|z_{1n}\|_2\\ + \|\gamma_k^{-1/2}H'W_k^{1/2}M^{(k)}W_k^{1/2}\bar{f}_k(\theta_0)\|_2\cdot\|z_{1n}\|_2^2 + \frac{1}{\sqrt{\gamma_k}}\|\Delta_{2n}\|_2\cdot\|z_{1n}\|_2^2\\ + \frac{1}{\gamma_k^{3/4}n^{1/4}}\|\hat{W}_k^{1/2}\bar{D}_2\|_2\|\hat{W}_k^{1/2}\bar{H}^{(k)}(\bar{\theta})\|_2\cdot\|z_{1n}\|_2^3 + \frac{1}{4\gamma_k}\|\Delta_{1n}\|_2\cdot\|z_{1n}\|_2^4.\end{aligned}$$

Since $\gamma_k/k = O(1)$, by Lemma B.3, we have

$$\frac{1}{\sqrt{n\gamma_k}}\|\bar{D}_2'\hat{W}_k\bar{D}_2\|_2 = O_P(\bar{\lambda}_k\sqrt{k/n}), \quad \frac{1}{(n\gamma_k)^{1/4}}\|\hat{W}_k^{1/2}\bar{f}_k(\theta_0)\|_2\|\hat{W}_k^{1/2}\bar{D}_2\|_2 = O_P(\bar{\lambda}_k k^{3/4}/n^{1/4}),$$

$$\frac{1}{\gamma_k^{3/4}n^{1/4}}\|\hat{W}_k^{1/2}\bar{D}_2\|_2\|\hat{W}_k^{1/2}\bar{H}^{(k)}(\bar{\theta})\|_2 = O_P(\bar{\lambda}_k k^{1/4}/n^{1/4}),$$

$$\frac{1}{\gamma_k}\|\Delta_{1n}\|_2 = O_P(\|\hat{W}_k-W_k\|_2) + O_P(\bar{\lambda}_k/\sqrt{n}) + O_P(\bar{\lambda}_k\|\hat{\theta}-\theta_0\|_2),$$

and

$$\frac{1}{\sqrt{\gamma_k}}\|\Delta_{2n}\|_2 = O_P(\bar{\lambda}_k\sqrt{k/n}) + O_P(\bar{\lambda}_k\sqrt{k}\|\hat{\theta}-\theta_0\|_2) + O_P(\sqrt{k}\|\hat{W}_k-W_k\|_2).$$

Since $\bar{\lambda}_k$ is bounded, $k^3/n \to 0$ and $\sqrt{k}\|\hat{W}_k - W_k\|_2 = o_P(1)$, it follows that

$$\frac{1}{\gamma_k}\|\Delta_{1n}\|_2 = o_P(1), \text{ and } \frac{1}{\sqrt{\gamma_k}}\|\Delta_{2n}\|_2 = O_P(k^{1/4}/n^{1/4})\|z_{1n}\|_2 = o_P(1)\|z_{1n}\|_2.$$

Hence,

$$\|z_{1n}\|_2^4 \leq \|\gamma_k^{-1/2} H' W_k^{1/2} M^{(k)} W_k^{1/2} \bar{f}_k(\theta_0)\|_2 \cdot \|z_{1n}\|_2^2 + o_P(1) \cdot \|z_{1n}\|_2 + o_P(1) \cdot \|z_{1n}\|_2^2 + o_P(1) \cdot \|z_{1n}\|_2^3 + o_P(1) \cdot \|z_{1n}\|_2^4. \quad \text{(B.8)}$$

Since $\gamma_k^{-1/2} H' W_k^{1/2} M^{(k)} W_k^{1/2} \bar{f}_k(\theta_0) = O_P(1)$, we can readily claim that $\|z_{1n}\| = O_P(1)$. Indeed, (B.8) amounts to

$$(1 + o_P(1))\|z_{1n}\|_2 \leq \frac{O_P(1)}{\|z_{1n}\|_2} + \frac{o_P(1)}{\|z_{1n}\|_2^2} + \frac{o_P(1)}{\|z_{1n}\|_2} + o_P(1).$$

Hence, if $\|z_{1n}\|_2 > 1$, this inequality implies $(1 + o_P(1))\|z_{1n}\|_2 \leq O_P(1) + o_P(1)$. Thus, we either have $(\|z_{1n}\|_2 < 1)$ or $(1 + o_P(1))\|z_{1n}\|_2 \leq O_P(1)$, which ensures that $\|z_{1n}\|_2 = O_P(1)$; that is,

$$\gamma_k^{1/4} n^{1/4}(\hat{\eta}_2 - \eta_{02}) = O_P(1).$$

Using (B.4), we obtain that $\sqrt{n}(\hat{\eta}_1 - \eta_{01}) = O_P(1)$. Recalling that $\hat{\theta} - \theta_0 = R_1(\hat{\eta}_1 - \eta_{01}) + R_2(\hat{\eta}_2 - \eta_{02})$, we have

$$\|\hat{\theta} - \theta_0\|_2 = O_P(n^{-1/2}) + O_P(\gamma_k^{-1/4} n^{-1/4}) = O_P(\gamma_k^{-1/4} n^{-1/4}).$$

Also, by the definition of R_1 and R_2 as spanning, respectively, the range of the transpose a matrix and the null space of that same matrix, we have $R_1' R_2 = 0$. Hence,

$$R_1'(\hat{\theta} - \theta_0) = R_1' R_1(\hat{\eta}_1 - \eta_{01}) = O_P(n^{-1/2})$$

and

$$R_2'(\hat{\theta} - \theta_0) = R_2' R_2(\hat{\eta}_2 - \eta_{02}) = O_P(\gamma_k^{-1/4} n^{-1/4}).$$

To complete the proof, it only remains to establish (B.6), which is done below. □

Proof of Theorem 3.1: Since θ_0 is in the interior of Θ and $\hat{\theta}$ converges in probability to θ_0, $\hat{\theta}$ is also an interior optimum with probability approaching one. Therefore, this estimator solves:

$$(\nabla_\theta \bar{f}_k(\hat{\theta}))' \hat{W}_k \bar{f}_k(\hat{\theta}) = 0. \quad \text{(B.9)}$$

By a mean-value expansion of $\nabla \bar{f}_k(\hat{\theta})$ and a second-order Taylor expansion $\bar{f}_k(\hat{\theta})$ around θ_0, we have

$$\nabla_\theta \bar{f}_k(\hat{\theta}) = \nabla_\theta \bar{f}_k(\theta_0) + \bar{H}^{(k)}(\dot{\theta})\sqrt{n}(\hat{\theta} - \theta_0)$$

and

$$\bar{f}_k(\hat{\theta}) = \bar{f}_k(\theta_0) + \nabla_\theta \bar{f}_k(\theta_0)(\hat{\theta} - \theta_0) + \frac{1}{2}\bar{H}^{(k)}(\ddot{\theta})\sqrt{n}(\hat{\theta} - \theta_0)^2,$$

with $\dot{\theta}, \ddot{\theta} \in (\theta_0, \hat{\theta})$. Let $\dot{h} := \bar{H}^{(k)}(\dot{\theta})$, $\ddot{h} := \bar{H}^{(k)}(\ddot{\theta})$, and $\bar{D}_2 := \nabla_\theta \bar{f}_k(\theta_0)$. The first-order condition (B.9) yields

$$n^{-1/4}(\nabla \bar{f}_k(\hat{\theta}))'\hat{W}_k \bar{f}_k(\hat{\theta}) = n^{-1/4}\bar{D}_2'\hat{W}_k \bar{f}_k(\theta_0) + n^{-1/4}\bar{D}_2'\hat{W}_k \bar{D}_2(\hat{\theta} - \theta_0) + \frac{1}{2}\bar{D}_2'\hat{W}_k \ddot{h} n^{1/4}(\hat{\theta} - \theta_0)^2$$
$$+ \dot{h}'\hat{W}_k \bar{f}_k(\theta_0) n^{1/4}(\hat{\theta} - \theta_0) + \dot{h}'\hat{W}_k \bar{D}_2 n^{1/4}(\hat{\theta} - \theta_0)^2 + \frac{1}{2}\dot{h}'\hat{W}_k \ddot{h} n^{3/4}(\hat{\theta} - \theta_0)^3 = 0.$$

In this framework, $\gamma_k = H'W_kH$. With $H := H^{(k)}(\theta_0)$, we can write

$$\frac{1}{2}(\gamma_k n)^{3/4}(\hat{\theta} - \theta_0)^3 + \frac{1}{2\gamma_k}(\dot{h}'\hat{W}_k\ddot{h} - H'W_kH)(\gamma_k n)^{3/4}(\hat{\theta} - \theta_0)^3 + \frac{3}{2\gamma_k^{3/4} n^{1/4}} H'\hat{W}_k \bar{D}_2 \sqrt{\gamma_k n}(\hat{\theta} - \theta_0)^2$$
$$+ \frac{1}{\gamma_k^{3/4} n^{1/4}}(\dot{h} - H)'\hat{W}_k \bar{D}_2 \sqrt{\gamma_k n}(\hat{\theta} - \theta_0)^2 + \frac{1}{\sqrt{\gamma_k}} H'W_k \bar{f}_k(\theta_0)(\gamma_k n)^{1/4}(\hat{\theta} - \theta_0)$$
$$+ \frac{1}{\sqrt{\gamma_k}}(\dot{h}'\hat{W}_k - H'W_k)\bar{f}_k(\theta_0)(\gamma_k n)^{1/4}(\hat{\theta} - \theta_0) + \frac{1}{2\gamma_k^{3/4} n^{1/4}} \bar{D}_2'\hat{W}_k(\ddot{h} - H)\sqrt{\gamma_k n}(\hat{\theta} - \theta_0)^2$$
$$+ \frac{1}{\sqrt{\gamma_k n}} \bar{D}_2'\hat{W}_k \bar{D}_2 (\gamma_k n)^{1/4}(\hat{\theta} - \theta_0) + n^{-1/4}\bar{D}_2'\hat{W}_k \bar{f}_k(\theta_0) = 0.$$

Similar to the lines of the proof of Lemma B.3, it is not hard to see that

$$\frac{1}{\gamma_k}(\dot{h}'\hat{W}_k\ddot{h} - H'W_kH) = o_P(1), \quad H'\hat{W}_k\bar{D}_2 = O_P(k), \quad \|\dot{h} - H\|_2 = O_P(\sqrt{k}[n^{-1/2} \vee \|\hat{\theta} - \theta_0\|_2]) = O_P(k^{1/4}/n^{1/4}),$$

$$(\dot{h} - H)'\hat{W}_k\bar{D}_2 = O_P(k^{3/4}/n^{1/4}), \text{ and } \frac{1}{\sqrt{\gamma_k}}(\dot{h}'\hat{W}_k - HW_k)\bar{f}_k(\theta_0) = o_P(1).$$

Then, letting $z_{1n} := (\gamma_k n)^{1/4}(\hat{\theta} - \theta_0)$ and $Z_n := \frac{1}{\sqrt{\gamma_k}} H'W_k \bar{f}_k(\theta_0)$, we have

$$z_{1n}(z_{1n}^2 + 2Z_n) = o_P(1). \tag{B.10}$$

Since $(z_{1n}, Z_n) = O_P(1)$, by the Prokhorov's theorem, each subsequence of has a further subsequence that converges in distribution to, say, (V, Z). Thus, along this converging subsequence, (B.10) implies that

$$V(V^2 + 2Z) = 0.$$

Therefore, it is not difficult to see that $|V| = 1_{Z \leq 0}\sqrt{-2Z}$ and, since as a Gaussian random variable Z has a symmetric distribution, $|V| = 1_{Z \geq 0}\sqrt{2Z}$. The fact that this limit distribution is not specific to the subsequence implies that the whole sequence converges to (V, Z). By the continuous mapping theorem, it follows that $z_{1n}^2 \xrightarrow{d} V^2 = 1_{Z \geq 0}(2Z)$ and this completes the proof. $\square$

Proof of Theorem 4.3: We proceed in two steps by showing first that $\hat{Z}$ is bounded by two statistics $\hat{Z}_1$ and $\hat{Z}_2$; that is,

$$\hat{Z}_1 + o_P(1) \leq \hat{Z} \leq \hat{Z}_2 + o_P(1). \tag{B.11}$$

We then show in the second step that $\hat{Z}_1$ and $\hat{Z}_2$ converge in distribution to $N(0,1)$ which establishes the stated result.

Step 1: By definition,

$$\bar{f}_k(\hat{\theta})'\hat{V}_k^{-1}\bar{f}_k(\hat{\theta}) \leq \bar{f}_k(\theta_0)'\hat{V}_k^{-1}\bar{f}_k(\theta_0) = \bar{f}_k(\theta_0)'V_k^{-1}\bar{f}_k(\theta_0) + \bar{f}_k(\theta_0)'(\hat{V}_k^{-1} - V_k^{-1})\bar{f}_k(\theta_0).$$

Note that

$$|\bar{f}_k(\theta_0)'(\hat{V}_k^{-1} - V_k^{-1})\bar{f}_k(\theta_0)| \leq \|\hat{V}_k^{-1} - V_k^{-1}\|_2 \|\bar{f}_k(\theta_0)\|_2^2.$$

From Theorem 3.1, the first-step GMM estimator $\tilde{\theta}$ is such that $\tilde{\theta} - \theta_0 = O_P(k^{-1/4}n^{-1/4})$. Hence, Lemma B.2(iv) implies that $\|\hat{V}_k^{-1} - V_k^{-1}\|_2 = O_P(k^{3/4}n^{-1/4})$. Thus,

$$\|\hat{V}_k^{-1} - V_k^{-1}\|_2 \|\bar{f}_k(\theta_0)\|_2^2 = O_P(k^{7/4}n^{-1/4}) = \sqrt{k}O_P(k^{5/4}n^{-1/4}) = o_P(\sqrt{k}).$$

As a result, we have

$$\hat{Z} = \frac{\bar{f}_k(\hat{\theta})'\hat{V}_k^{-1}\bar{f}_k(\hat{\theta}) - k}{\sqrt{2k}} \leq \frac{\bar{f}_k(\theta_0)'V_k^{-1}\bar{f}_k(\theta_0) - k}{\sqrt{2k}} + o_P(1) := \hat{Z}_2 + o_P(1). \tag{B.12}$$

On the other hand, using (B.5), we can write

$$\begin{aligned}
&\bar{f}_k(\hat{\theta})'\hat{V}_k^{-1}\bar{f}_k(\hat{\theta}) \\
&\quad = \left[\bar{f}_k(\hat{\theta})'\hat{V}_k^{-1/2}\bar{P}^{(k)}\hat{V}_k^{-1/2}\bar{f}_k(\hat{\theta}) - \bar{f}_k(\theta_0)'\hat{V}_k^{-1/2}\bar{P}^{(k)}\hat{V}_k^{-1/2}\bar{f}_k(\theta_0)\right] + \bar{f}_k(\theta_0)'\hat{V}_k^{-1}\bar{f}_k(\theta_0) \\
&\quad + (\hat{\eta}_2 - \eta_{02})'\bar{D}_2'\hat{V}_k^{-1/2}\bar{M}^{(k)}\hat{V}_k^{-1/2}\bar{D}_2(\hat{\eta}_2 - \eta_{02}) + \frac{1}{4}z_{0n}'\bar{H}^{(k)}(\bar{\theta})'\hat{V}_k^{-1/2}\bar{M}^{(k)}\hat{V}_k^{-1/2}\bar{H}^{(k)}(\bar{\theta})z_{0n} \\
&\quad + 2\bar{f}_k(\theta_0)'\hat{V}_k^{-1/2}\bar{M}^{(k)}\hat{V}_k^{-1/2}\bar{D}_2(\hat{\eta}_2 - \eta_{02}) + \bar{f}_k(\theta_0)'\hat{V}_k^{-1/2}\bar{M}^{(k)}\hat{V}_k^{-1/2}\bar{H}^{(k)}(\bar{\theta})z_{0n} \\
&\quad + (\hat{\eta}_2 - \eta_{02})'\bar{D}_2'\hat{V}_k^{-1/2}\bar{M}^{(k)}\hat{V}_k^{-1/2}\bar{H}^{(k)}(\bar{\theta})z_{0n} := (1) + (2) + (3) + (4) + (5) + (6) + (7).
\end{aligned}$$

From (B.6), and Lemma B.2, we have

$$\begin{aligned}
&(1) = O_P(k^{3/4}n^{-1/4}) = o_P(1), \quad (3) = O_P(k^{1/2}n^{-1/2}) = o_P(1), \\
&\quad (4) = \frac{1}{4}z_{0n}'H'V_k^{-1/2}M^{(k)}V_k^{-1/2}Hz_{0n} + o_P(1), \quad (5) = O_P(k^{3/4}n^{-1/4}) = o_P(1), \\
&\qquad (6) = \bar{f}_k(\theta_0)'V_k^{-1/2}M^{(k)}V_k^{-1/2}Hz_{0n} + o_P(1), \quad (7) = O_P(k^{1/4}n^{-1/4}) = o_P(1)
\end{aligned}$$

and from the lines above, (2) $= \bar{f}_k(\theta_0)'V_k^{-1}\bar{f}_k(\theta_0) + o_P(\sqrt{k})$. As a result, we write

$$\bar{f}_k(\hat{\theta})'\hat{V}_k^{-1}\bar{f}_k(\hat{\theta}) = \bar{f}_k(\theta_0)'V_k^{-1}\bar{f}_k(\theta_0) + \bar{f}_k(\theta_0)'V_k^{-1/2}M^{(k)}V_k^{-1/2}Hz_{0n} + \frac{1}{4}z_{0n}'H'V_k^{-1/2}M^{(k)}V_k^{-1/2}Hz_{0n} + o_P(\sqrt{k}). \quad \text{(B.13)}$$

Let the rank factorization of $M^{(k)}V_k^{-1/2}H$ be $M^{(k)}V_k^{-1/2}H = H_1H_2$, where H_1 and H_2 are a (k, r_h)-matrix and a (r_h, p^2)-matrix, respectively, with the same rank $r_h = \text{Rank}(M^{(k)}V_k^{-1/2}H) \leq p^2$. By second-order local identification, $r_h \neq 0$ so that $M^{(k)}V_k^{-1/2}H \neq 0$.

Thus, (B.13) can be written as

$$\bar{f}_k(\hat{\theta})'\hat{V}_k^{-1}\bar{f}_k(\hat{\theta}) = \bar{f}_k(\theta_0)'V_k^{-1}\bar{f}_k(\theta_0) + \bar{f}_k(\theta_0)'V_k^{-1/2}H_1H_2z_{0n} + \frac{1}{4}z_{0n}'H_2'H_1'H_1H_2z_{0n} + o_P(\sqrt{k}).$$

Letting $m(u) := \bar{f}_k(\theta_0)'V_k^{-1}\bar{f}_k(\theta_0) + \bar{f}_k(\theta_0)'V_k^{-1/2}H_1u + \frac{1}{4}u'H_1'H_1u$, and $M_1 := I_k - H_1(H_1'H_1)^{-1}H_1'$, we can claim that

$$\min_{u\in\mathbb{R}^{r_h}} m(u) + o_P(\sqrt{k}) = \bar{f}_k(\theta_0)'V_k^{-1/2}M_1V_k^{-1/2}\bar{f}_k(\theta_0) + o_P(\sqrt{k}) \leq \bar{f}_k(\hat{\theta})'\hat{V}_k^{-1}\bar{f}_k(\hat{\theta}).$$

Letting $\hat{Z}_1 := \frac{\bar{f}_k(\theta_0)'V_k^{-1/2}M_1V_k^{-1/2}\bar{f}_k(\theta_0) - k}{\sqrt{2k}}$, we obtain (B.11).

Step 2: We now show that both $\hat{Z}_1$ and $\hat{Z}_2$ are asymptotically standard normal. We first consider $\hat{Z}_2$. We have

$$\hat{Z}_2 = \frac{1}{n\sqrt{2k}}\sum_{t\neq s:t,s=1}^{n} f_k(x_s, y_s, \theta_0)'V_k^{-1}f_k(x_t, y_t, \theta_0) + \frac{\frac{1}{n}\sum_{t=1}^{n} f_k(x_t, y_t, \theta_0)'V_k^{-1}f_k(x_t, y_t, \theta_0) - k}{\sqrt{2k}} := U_{1n} + U_{2n}.$$

The asymptotic normality of U_{1n} follows readily from the central limit theorem stated by Theorem 4.2. In addition, it is not hard to see that $E(U_{2n}) = 0$. Using similar arguments as in the proof of Proposition 4.1, we can show that $E(U_{2n}^2) = o(1)$. This establishes that $U_{2n} = o_P(1)$. We can then conclude that $\hat{Z}_2$ is asymptotically standard normal.

We now consider $\hat{Z}_1$. Note first that, since M_1 is an orthogonal projection matrix on a space of dimension $k - r_h$, there exists a $(k, k - r_h)$-matrix S_1 such that $S_1'S_1 = I_{k-r_h}$ and $M_1 = S_1S_1'$. In that respect, $\bar{f}_k(\theta_0)'V_k^{-1/2}M_1V_k^{-1/2}\bar{f}_k(\theta_0) = \bar{f}_k(\theta_0)'V_k^{-1/2}S_1S_1'V_k^{-1/2}\bar{f}_k(\theta_0)$. Also, $Var(S_1'V_k^{-1/2}\bar{f}_k(\theta_0)) = I_{k-r_h}$. Using Theorem 4.2 and similarly to the lines above for $\hat{Z}_2$, we can claim that

$$\hat{Z}_3 := \frac{\bar{f}_k(\theta_0)'V_k^{-1/2}S_1S_1'V_k^{-1/2}\bar{f}_k(\theta_0) - (k - r_h)}{\sqrt{2(k - r_h)}} \xrightarrow{d} N(0, 1).$$

Since $0 \leq r_h \leq p^2$ with p fixed, we can see that $\hat{Z}_1 = \hat{Z}_3 + o_P(1)$. Therefore, $\hat{Z}_1 \xrightarrow{d} N(0, 1)$. □

Proof of Lemma 4.4: Note that conditions of the theorem ensure that the maps $\theta \mapsto \mathbb{E}(g_l(x)u(\theta))$ are continuous on Θ for each l. By Lemma 1 of de Jong and Bierens (1994), we can claim that, for each $\theta_i \in \Theta$, there exists l_i such that $|\mathbb{E}(g_{l_i}(x)u(\theta_i))| = \delta_i > 0$. By continuity of the map $\theta \mapsto \mathbb{E}(g_{l_i}(x)u(\theta))$, there exists an open neighborhood V_i of θ_i such that $|\mathbb{E}(g_{l_i}(x)u(\theta))| > \delta_i/2$ for all $\theta \in V_i$. Clearly, $\Theta \subset \cup_{\theta_i \in \Theta} V_i$ and, by compactness of Θ, we can extract a finite number of elements from the sequence of V_i to cover Θ. That is, $\Theta \subset \cup_{j=1}^{m_0} V_{i_j}$, with m_0 finite. Let $k_0 = \max_{1 \le j \le m_0} l_{i_j}$. k_0 is a finite integer and, by construction, for any $\theta \in \Theta$, $\mathbb{E}(g^{(k_0)}(x)u(\theta)) \neq 0$. Also, taking $\delta_0 = \min_{1 \le j \le m_0}(\delta_{i_j}/2)^2$, we obviously have $\|\mathbb{E}(g^{(k_0)}(x)u(\theta))\|_2^2 \ge \delta_0 > 0$ and this concludes the proof. $\square$

Proof of Theorem 4.5: From Lemma 4.4, there exist k_0 and $\delta_0 > 0$ such that $\mathbb{E}(f_{k_0}(x_t, y_t, \hat{\theta}))'\mathbb{E}(f_{k_0}(x_t, y_t, \hat{\theta})) \ge \delta_0$. We have

$$\begin{aligned} k^{3/2}n^{-1}|\hat{Z}| &\ge 2^{-1/2}kn^{-1}(\bar{f}_k(\hat{\theta})'\hat{V}_k^{-1}\bar{f}_k(\hat{\theta}) - k) \ge 2^{-1/2}kn^{-1}\bar{f}_k(\hat{\theta})'\bar{f}_k(\hat{\theta})/\lambda_{\max}(\hat{V}_k) + 2^{1/2}k^2n^{-1} \\ &\ge (2^{-1/2}/\bar{\lambda})n^{-1}\bar{f}_{k_0}(\hat{\theta})'\bar{f}_{k_0}(\hat{\theta}) + o(1), \text{ with probability approaching one.} \end{aligned}$$

Also,

$$\begin{aligned} n^{-1}\bar{f}_{k_0}(\hat{\theta})'\bar{f}_{k_0}(\hat{\theta}) &\ge \mathbb{E}(f_{k_0}(x_t, y_t, \hat{\theta}))'\mathbb{E}(f_{k_0}(x_t, y_t, \hat{\theta})) + 2\left(\frac{1}{n}\sum_{t=1}^{n} f_{k_0}(x_t, y_t, \hat{\theta}) - \mathbb{E}[f_{k_0}(x_t, y_t, \hat{\theta})]\right)' \mathbb{E}(f_{k_0}(x_t, y_t, \hat{\theta})) \\ &\ge \delta_0 - 2\left\|\frac{1}{n}\sum_{t=1}^{n} f_{k_0}(x_t, y_t, \hat{\theta}) - \mathbb{E}[f_{k_0}(x_t, y_t, \hat{\theta})]\right\|_2 \|\mathbb{E}(f_{k_0}(x_t, y_t, \hat{\theta}))\|_2 = \delta_0 + o_P(1)O_P(1). \end{aligned}$$

It follows that, with probability approaching one, $k^{3/2}n^{-1}|\hat{Z}| \ge (2^{-1/2}/\bar{\lambda})\delta_0 + o_P(1)$ and this concludes the proof by setting $\delta := (2^{-1/2}/\bar{\lambda})\delta_0$. $\square$

References

[1] Ando, T., and J.L. van Hemmen (1980), An inequality for trace ideals, Communications in Mathematical Physics 76, 143–148.

[2] Andrews, D. W. K. (1984), Non-strong mixing autoregressive processes, Journal of Applied Probability 21, 930–934.

[3] Antoine, B., K. Proulx, and E. Renault (2020), Pseudo-true SDFs in conditional asset pricing models, Journal of Financial Econometrics 18, 656–714.

[4] Berman, S.M. (1964), Limit theorems for the maximum term in stationary sequences, Annals of Mathematical Statistics 35, 502–516.

[5] Bierens, H.J. (1982), Consistent model specification tests, Journal of Econometrics 20, 105–134.

[6] Bierens, H.J., and W. Ploberger (1987), Asymptotic theory of integrated conditional moment tests, Econometrica 65, 1129–1151.

[7] Bougerol, P., and N. Picard (1992), Stationarity of GARCH processes and some nonnegative time series, Journal of Econometrics 52, 115–127.

[8] Carrasco, M., and X. Chen (2002), Mixing and moment properties of various GARCH and stochastic volatility models, Econometric Theory 18, 17–39.

[9] Carrasco, M., and J.P. Florens (2000), Generalization of GMM to a continuum of moment conditions, Econometric Theory 16, 797–834.

[10] Chao, J.C., and N.R. Swanson (2005), Consistent estimation with a large number of weak instruments, Econometrica 73, 1673–1692.

[11] Cochrane, J., and M. Piazzesi (2005), Bond risk premia, American Economic Review 95, 138–160.

[12] Crump, R.K., and N. Gospodinov (2022), On the factor structure of bond returns, Econometrica 90, 295–314.

[13] Dedecker, J., and C. Prieur (2005), New dependence coefficients. Examples and applications to statistics, Probability Theory and Related Fields 132, 203–236.

[14] de Jong, R., and H.J. Bierens (1994), On the limit behavior of a chi-square type test if the number of conditional moments tested approaches infinity, Econometric Theory 10, 70–90.

[15] Dominguez, M., and I. Lobato (2004), Consistent estimation of models defined by conditional moment restrictions, Econometrica 72, 1601–1615.

[16] Domowitz, I., and H. White (1982), Misspecified models with dependent observations, Journal of Econometrics 20, 35–58.

[17] Donald, S.G., G.W. Imbens and W.K. Newey (2003), Empirical likelihood estimation and consistent tests with conditional moment restrictions, Journal of Econometrics 117, 55–93.

[18] Dovonon, P., and Y.F. Atchadé, (2020), Efficiency bounds for semiparametric models with singular score functions, Econometric Reviews 39, 612–648.

[19] Dovonon, P., and A. Hall (2018), The asymptotic properties of GMM and indirect inference under second-order identification, Journal of Econometrics 205, 76–111.

[20] Dovonon, P., A. Hall, and F. Kleibergen (2020), Inference in second-order identified models, Journal of Econometrics 218, 346–372.

[21] Dovonon, P., and E. Renault (2013), Testing for common conditionally heteroskedastic factors, Econometrica 81, 2561–2586.

[22] Dovonon, P., and E. Renault (2020), GMM overidentification test with first order underidentification, Working paper.

[23] Duffee, G. (2011), Forecasting with the term structure: the role of no-arbitrage restrictions, Working paper.

[24] Engle, R.F., and S. Kozicki (1993), Testing for common features, Journal of Business & Economic Statistics 11, 369–380.

[25] Engle, R.F., V.K. Ng, and M. Rothschild (1990), Asset pricing with a factor-ARCH covariance structure: Empirical estimates for treasury bills, Journal of Econometrics 45, 213–237.

[26] Fan, Y., and Q. Li (1999), Central limit theorem for degenerate U-statistics of absolutely regular processes with applications to model specification testing, Journal of Nonparametric Statistics 10, 245–271.

[27] Francq, C., and J.-M. Zakoian (2006), Mixing properties of a general class of GARCH(1,1) models without moment assumptions on the observed process, Econometric Theory 22, 815–834.

[28] Fryzlewicz, P., and S. Subba Rao (2011), Mixing properties of ARCH and time-varying ARCH processes, Bernoulli 17, 320–346.

[29] Gao, J. (2007), Nonlinear Time Series: Semiparametric and Nonparametric Methods, CRC Press, Florida USA.

[30] Gao, J., and H. Hong (2008), Central limit theorems for generalized U-statistics with applications in nonparametric specification, Journal of Nonparametric Statistics 20, 61–76.

[31] Gospodinov, N., R. Kan, and C. Robotti (2017), Spurious inference in reduced-rank asset-pricing models, Econometrica 85, 1613–1628.

[32] Gospodinov, N., and S. Ng (2015), Minimum distance estimation of possibly noninvertible moving average models, Journal of Business & Economic Statistics 33, 403–417.

[33] Hall, P. (1984), Central limit theorem for integrated square error of multivariate nonparametric density estimators, Journal of Multivariate Analysis 14, 1–16.

[34] Han, C., and P.C.B. Phillips (2006), GMM with many moment conditions, Econometrica 74, 147–192.

[35] Hansen, L. P. (2014), Uncertainty outside and inside economic models, Journal of Political Economy 122, 945–987.

[36] Joslin, S., and A. Le (2021), Interest rate volatility and no-arbitrage affine term structure models, Management Science 67, 7391–7416.

[37] Joslin, S., K.J. Singleton, and H. Zhu (2011), A new perspective on Gaussian dynamic term structure models, Review of Financial Studies 24, 926–970.

[38] Isaev, M., I.V. Rodionov, R.-R. Zhang, and M.E. Zhukovskii (2020), Extreme value theory for triangular arrays of dependent random variables, Russian Mathematical Surveys 75, 968–970.

[39] Kim, T.Y, Z.-M. Luo and C. Kim (2011), The central limit theorem for degenerate variable U-statistics under dependence, Journal of Nonparametric Statistics 23, 683–699.

[40] Kitamura, Y., G. Tripathi, and H. Ahn (2004), Empirical likelihood-based inference in conditional moment restriction models, Econometrica 72, 1667–1714.

[41] Lee, J.H., and Z. Liao (2018), On standard inference for GMM with local identification failure of known forms, Econometric Theory 34, 790–814.

[42] Leucht, A. (2012), Degenerate U- and V-statistics under weak dependence: Asymptotic theory and bootstrap consistency, Bernoulli 18, 552–585.

[43] Newey, W.K., and D. McFadden (1994), Large sample estimation and hypothesis testing, chapter 36 in Handbook of Econometrics Vol. 4, 2111–2245.

[44] Pereira, H.I. (1983), Rate of convergence towards a Fréchet type limit distribution, Annales scientifiques de l'Université de Clermont-Ferrand II 76, 67–80.

[45] Pötscher, B.M., and I.R. Prucha (1989), A uniform law of large numbers for dependent and heterogeneous data processes, Econometrica 57, 675–683.

[46] Rotnitzky, A., D.R. Cox, M. Bottai, and J. Robins (2000), Likelihood based inference with singular information matrix, Bernoulli 6, 243–284.

[47] Roussas, G.G., and D. Ioannides (1987), Moment inequalities for mixing sequences of random variables, Stochastic Analysis and Applications 5, 61–120.

[48] Sargan, J.D. (1983), Identification and lack of identification, Econometrica 51, 1605–1633.

[49] Smith, R. (2007), Efficient information theoretic inference for conditional moment restrictions, Journal of Econometrics 138, 430–460.

[50] Tripathi, G., and Y. Kitamura (2003), Testing conditional moment restrictions, Annals of Statistics 31, 2059–2095.

Online Appendix for

"Specification Testing for Conditional Moment Restrictions under Local Identification Failure"

Prosper Dovonon and Nikolay Gospodinov

Proofs of Lemmas B.1, B.2 and B.3 in Appendix B

The proofs of Lemma B.1(i), Proposition 4.1 and Theorem 4.2 rely on the following result.

Lemma OA.1 (Theorem 7.4 of Roussas and Ioannides, 1987) *Suppose that F_n^m are the σ-algebras generated by a stationary α-mixing process x_i with mixing coefficient $\alpha(i)$. For some positive integer ℓ, let $\eta_i \in F_{s_i}^{t_i}$, where $s_1 < t_1 < s_2 < t_2 < \cdots < t_\ell$ and $s_{i+1} - t_i \geq \tau$ for all i. In addition, let $\|\eta\|_p = [E|\eta|^p]^{1/p}$ for $1 < p < \infty$ and $\|\eta\|_\infty = esssup|\eta|$, respectively, and assume further that*

$$\|\eta\|_{p_i} < \infty \text{ for some } 1 < p_i \leq \infty \text{ with } q = \sum_{i=1}^{\ell} \frac{1}{p_i} < 1.$$

Then,

$$\left| \mathbb{E}\left(\prod_{i=1}^{\ell} \eta_i \right) - \prod_{i=1}^{\ell} \mathbb{E}(\eta_i) \right| \leq 10(\ell - 1)\alpha(\tau)^{1-q} \prod_{i=1}^{\ell} \|\eta_i\|_{p_i}.$$

Proof of Lemma B.1:

$$\begin{aligned}
\left\| \frac{1}{n} \sum_{t=1}^{n} g^{(k)}(x_t) U_t(\bar{\theta})' - \mathbb{E}\left(g^{(k)}(x_t) U_t(\theta_0)' \right) \right\|_2 &\leq O_P\left(\sqrt{\frac{k}{n}} \right) + \left\| \mathbb{E}\left(g^{(k)}(x_t) U_t(\bar{\theta})' \right) - \mathbb{E}\left(g^{(k)}(x_t) U_t(\theta_0)' \right) \right\|_2 \\
&\leq O_P\left(\sqrt{\frac{k}{n}} \right) + \left(\sum_{l=1}^{k} \sum_{r=1}^{m} \left[\mathbb{E}\left(g_l(x_t) U_{t,r}(\bar{\theta}) \right) - \mathbb{E}\left(g_l(x_t) U_{t,r}(\theta_0) \right) \right]^2 \right)^{1/2} \\
&\leq O_P\left(\sqrt{\frac{k}{n}} \right) + \sqrt{km} \cdot c \cdot \|\bar{\theta} - \theta_0\|_2 = O_P\left(\sqrt{k} \cdot (n^{-1/2} \vee \|\bar{\theta} - \theta_0\|_2) \right). \quad \square
\end{aligned}$$

Proof of Lemma B.2: (i) Recall that

$$\|\bar{V}_k - V_k\|_2 \leq k \max_{1 \leq h,l \leq k} |[\bar{V}_k - V_k]_{hl}|.$$

Letting $s_{i,hl} = v_i(\theta_0) - E(v_i(\theta_0))$, with $v_i(\theta) = g_h(x_i) g_l(x_i) u_i(\theta)^2$, we have

$$\begin{aligned}
[\bar{V}_k - V_k]_{hl} &= \frac{1}{n} \sum_{i=1}^{n} v_i(\bar{\theta}) - \mathbb{E}[v_i(\theta_0)] \\
&= \left(\frac{1}{n} \sum_{i=1}^{n} s_{i,hl} \right) + \frac{2}{n} \sum_{i=1}^{n} g_h(x_i) g_l(x_i) u_i(\bar{\theta}) \nabla_\theta u_i(\tilde{\theta})(\bar{\theta} - \theta_0) := (a') + (b'). \quad \text{(OA.1)}
\end{aligned}$$

The first equality follows by definition and the second one follows from a mean-value expansion of $(1/n) \sum_{i=1}^{n} v_i(\bar{\theta})$ around θ_0 with $\tilde{\theta} \in (\bar{\theta}, \theta_0)$. We need to determine the order of magnitude of (a') and (b'). Pick $0 < \xi < 1/2$ and let $\alpha = (1 - 2\xi)/2$. We have

$$\mathbb{E}\left(n^{\xi} a' \right)^2 = n^{2\xi - 2} \sum_{i,j=1}^{n} \mathbb{E}(s_{i,hl} s_{j,hl}).$$

From Assumption A.1-(ii), we have $\max_{h,l} E|s_{i,hl}|^{\delta} < \infty$ with $\delta > 2$. Also, using Assumption 1, we can apply Lemma OA.1, and claim (with $q = 2/\delta < 1$) that: for all i, j,

$$|\mathbb{E}(s_{i,hl}s_{j,hl}) - \mathbb{E}(s_{i,hl})\mathbb{E}(s_{j,hl})| \leq 10\rho^{(1-q)|i-j|}(\mathbb{E}|s_{i,hl}|^{\delta})^{2/\delta} \leq 10\rho^{(1-q)|i-j|}(\max_{h,l}\mathbb{E}|s_{i,hl}|^{\delta})^{2/\delta} := c\rho^{(1-q)|i-j|}.$$

Note that for $|i-j| > n^{\alpha}$,

$$|\mathbb{E}(s_{i,hl}s_{j,hl})| = |\mathbb{E}(s_{i,hl}s_{j,hl}) - \mathbb{E}(s_{i,hl})\mathbb{E}(s_{j,hl})| \leq c\rho^{(1-q)n^{\alpha}}$$

and for $|i-j| \leq n^{\alpha}$,

$$|\mathbb{E}((s_{i,hl}s_{j,hl}))| \leq \left(\mathbb{E}(s_{i,hl}^2)\mathbb{E}(s_{j,hl}^2)\right)^{1/2} \leq c.$$

It follows that

$$\sum_{i,j=1}^{n}\mathbb{E}(s_{i,hl}s_{j,hl}) = \sum_{|i-j|\leq n^{\alpha}}\mathbb{E}(s_{i,hl}s_{j,hl}) + \sum_{|i-j|>n^{\alpha}}\mathbb{E}(s_{i,hl}s_{j,hl}) \leq c\cdot n\cdot n^{\alpha} + 2c\cdot n\cdot\sum_{j=n^{\alpha}+1}^{n}\rho^{(1-q)j}$$
$$\leq c\cdot n\cdot n^{\alpha} + 2c\cdot n\rho^{(1-q)(n^{\alpha}+1)}\frac{1-\rho^{(1-q)(n-n^{\alpha})}}{1-\rho^{1-q}} \leq c\cdot n\cdot n^{\alpha} + \frac{4c}{1-\rho^{1-q}}\cdot n.$$

As a result, for any $n \geq 1$,

$$n^{2\xi-2}\sum_{i,j}\mathbb{E}(s_{i,hl}s_{j,hl}) \leq c\cdot n^{-\alpha} + c'\cdot n^{2\xi-1} = c\cdot n^{-\alpha} + c'\cdot n^{-2\alpha}$$

uniformly in h, l. Letting $\xi \to 1/2$, $\alpha \to 0$ and, hence, $n^{-1}\sum_{i,j}E(s_{i,hl}s_{j,hl}) \leq c + c'$. We can therefore claim that

$$\max_{h,l}\left|\frac{1}{n}\sum_{i=1}^{n}s_{i,hl}\right| = O_P(n^{-1/2}).$$

Besides,

$$|(b')| \leq 2\left(\frac{1}{n}\sum_{i=1}^{n}|g_h(x_i)||g_l(x_i)|\sup_{\theta\in\mathcal{N}}|u_i(\theta)|\|\nabla_{\theta}u_i(\theta)\|\right)\|\bar{\theta}-\theta_0\|$$
$$\leq 2\left(\frac{1}{n}\sum_{i=1}^{n}g_l(x_i)^2\sup_{\theta\in\mathcal{N}}|u_i(\theta)|^2\right)^{1/2}\left(\frac{1}{n}\sum_{i=1}^{n}g_h(x_i)^2\sup_{\theta\in\mathcal{N}}\|\nabla_{\theta}u_i(\theta)\|^2\right)^{1/2}\|\bar{\theta}-\theta_0\| = O_P(\|\bar{\theta}-\theta_0\|).$$

We deduce from (OA.1) that:

$$\max_{1\leq h,l\leq k}|[\hat{V}_k - V_k]_{hl}| = O_P(n^{-1/2}) + O_P(\|\bar{\theta}-\theta_0\|) = O_P(n^{-1/2}\vee\|\bar{\theta}-\theta_0\|)$$

and the result follows.

(ii) and (iii): Let c be the unit vector of $\mathbb{R}^k$ such that $c'V_kc = \lambda_k$ if $\lambda_k \leq \lambda_{\min}(\bar{V}_k)$. (Choose c such that $c'\bar{V}_kc = \lambda_{\min}(\bar{V}_k)$ otherwise.) We then have

$$|\lambda_k - \lambda_{\min}(\bar{V}_k)| \leq \left|c'(V_k - \bar{V}_k)c\right| \leq \sum_{h,l=1}^{k} |c_h||c_l| \left|[V_k - \bar{V}_k]_{hl}\right| \leq \max_{1\leq h,l\leq k} \left|[V_k - \bar{V}_k]_{hl}\right| \left(\sum_{l=1}^{k} |c_l|\right)^2$$
$$\leq k \cdot \max_{1\leq h,l\leq k} \left|[V_k - \bar{V}_k]_{hl}\right|$$

and (ii) follows from the proof of part (i). Part (iii) is obtained by dividing each side of (ii) by λ_k.

We establish (iv) by recalling that $\bar{V}_k^{-1} - V_k^{-1} = -\bar{V}_k^{-1}(\bar{V}_k - V_k)V_k^{-1}$. Thus, by the Cauchy-Schwarz inequality, we have:

$$\|\bar{V}_k^{-1} - V_k^{-1}\|_2 \leq \lambda_{\max}(\bar{V}_k^{-1})\|\hat{V}_k - V_k\|_2\lambda_{\max}(V_k^{-1}) = \frac{1}{\lambda_{\min}(\bar{V}_k)\lambda_k}\|\bar{V}_k - V_k\|_2 = \frac{\lambda_k}{\lambda_{\min}(\bar{V}_k)}\lambda_k^{-2}\|\bar{V}_k - V_k\|_2.$$

This yields the stated order by using (i) and (iii). The proof of (v) follows the same arguments as those in the proof of (iii). □

Proof of Lemma B.3: (i) We have $\|\bar{D}_2'\hat{W}_k\bar{D}_2\|_2 = \lambda_{\max}(\bar{D}_2'\hat{W}_k\bar{D}_2) \leq [\lambda_{\max}(\hat{W}_k)/\bar{\lambda}_k]\bar{\lambda}_k\lambda_{\max}(\bar{D}_2'\bar{D}_2) = O_P(\bar{\lambda}_k k)$.

(ii) Same as (i) by noting that $\|\hat{W}_k^{1/2}\bar{f}_k\|_2 = \sqrt{\lambda_{\max}(\bar{f}_k'\hat{W}_k\bar{f}_k)} = \sqrt{\bar{f}_k'\hat{W}_k\bar{f}_k}$.

(iii) We have $\|\hat{W}_k^{1/2}\bar{H}\|_2 \leq \|\hat{W}_k^{1/2}(\bar{H} - H)\|_2 + \|\hat{W}_k^{1/2}H\|_2 \leq \|\hat{W}_k^{1/2}\|_2\|\bar{H} - H\|_2 + O_P(\sqrt{\bar{\lambda}_k k})$. Note that
$\|\hat{W}_k^{1/2}\|_2\|\bar{H} - H\|_2 \leq \sqrt{\lambda_{\max}(\hat{W}_k)}O_P(\sqrt{k}[n^{-1/2} \vee \|\bar{\theta} - \theta_0\|_2]) = o_P(\sqrt{\bar{\lambda}_k k})$ and the result follows.

(iv) We have $\|(\bar{D}_1'\hat{W}_k\bar{D}_1)^{-1}\|_2 = [\lambda_{\min}(\bar{D}_1'\hat{W}_k\bar{D}_1)]^{-1}$. But, $\lambda_{\min}(\bar{D}_1'\hat{W}_k\bar{D}_1) \geq \lambda_{\min}(\hat{W}_k)\lambda_{\min}(\bar{D}_1'\bar{D}_1)$. Note that, proceeding as in the proof of Lemma B.2-(iii), we obtain

$$|\lambda_{\min}(\bar{D}_1'\bar{D}_1)/\lambda_{\min}(D_1'D_1) - 1| = O_P(\|\bar{D}_1'\bar{D}_1 - D_1'D_1\|_2/\lambda_{\min}(D_1'D_1)).$$

Using Lemma B.1, it is not difficult to see that $\|\bar{D}_1'\bar{D}_1 - D_1'D_1\|_2 = O_P(k[n^{-1/2} \vee \|\bar{\theta} - \theta_0\|_2]) = o_P(k)$. As a result, $\lambda_{\min}(\bar{D}_1'\bar{D}_1)/\lambda_{\min}(D_1'D_1) - 1 = o_P(1)$. We can therefore claim that

$$\|(\bar{D}_1'\hat{W}_k\bar{D}_1)^{-1}\|_2 \leq \frac{1}{\lambda_{\min}(\hat{W}_k)\lambda_{\min}(\bar{D}_1'\bar{D}_1)} = O_P([\lambda_{\min}(W_k)\lambda_{\min}(D_1'D_1)]^{-1}) = O_P(\bar{\lambda}_k^{-1}k^{-1}).$$

(v) We have

$$\bar{M}^{(k)} - M^{(k)} = (\hat{W}_k^{1/2} - W_k^{1/2})\bar{D}_1(\bar{D}_1'\hat{W}_k\bar{D}_1)^{-1}\bar{D}_1'\hat{W}_k^{1/2} + W_k^{1/2}(\bar{D}_1 - D_1)'(\bar{D}_1'\hat{W}_k\bar{D}_1)^{-1}\bar{D}_1'\hat{W}_k^{1/2}$$
$$+ W_k^{1/2}D_1[(\bar{D}_1'\hat{W}_k\bar{D}_1)^{-1} - (D_1'W_kD_1)^{-1}]\bar{D}_1'\hat{W}_k^{1/2} + W_k^{1/2}D_1(D_1'W_kD_1)^{-1}(\bar{D}_1 - D_1)\hat{W}_k^{1/2}$$
$$+ W_k^{1/2}D_1(D_1'W_kD_1)^{-1}D_1(\hat{W}_k^{1/2} - W_k^{1/2}) := (1) + (2) + (3) + (4) + (5).$$

Note that

$$(\bar{D}_1'\hat{W}_k\bar{D}_1)^{-1} - (D_1'W_kD_1)^{-1} = -(\bar{D}_1'\hat{W}_k\bar{D}_1)^{-1}(\bar{D}_1'\hat{W}_k\bar{D}_1 - D_1'W_kD_1)(D_1'W_kD_1)^{-1}.$$

From (iv), $\|(\bar{D}_1'\hat{W}_k\bar{D}_1)^{-1}\|_2 = O_P(\bar{\lambda}_k^{-1}k^{-1})$ and $\|D_1'W_kD_1\|_2 = O_P(\bar{\lambda}_k^{-1}k^{-1})$. Also,

$$\bar{D}_1'\hat{W}_k\bar{D}_1 - D_1'W_kD_1 = (\bar{D}_1 - D_1)'\hat{W}_k\bar{D}_1 + D_1'(\hat{W}_k - W_k)\bar{D}_1 + D_1'W_k(\bar{D}_1 - D_1).$$

Hence,

$$\|\bar{D}_1'\hat{W}_k\bar{D}_1 - D_1'W_kD_1\|_2 = O_P(\bar{\lambda}_k k[n^{-1/2} \vee \|\bar{\theta} - \theta_0\|_2]) + O_P(k\|\hat{W}_k - W_k\|_2).$$

Thus,

$$\|(\bar{D}_1'\hat{W}_k\bar{D}_1)^{-1} - (D_1'W_kD_1)^{-1}\|_2 = O_P(\bar{\lambda}_k^{-1}k^{-1}[n^{-1/2} \vee \|\bar{\theta} - \theta_0\|_2]) + O_P(\bar{\lambda}_k^{-2}k^{-1}\|\hat{W}_k - W_k\|_2).$$

Also, using the Ando-van Hemmen inequality (see Ando and van Hemmen, 1980), we have

$$\left\|\hat{W}_k^{1/2} - W_k^{1/2}\right\|_2 \le \frac{1}{\lambda_{\min}(\hat{W}_k)^{1/2} + \lambda_{\min}(W_k)^{1/2}}\|\hat{W}_k - W_k\|_2 \le \lambda_{\min}(W_k)^{-1/2}\|\hat{W}_k - W_k\|_2.$$

Then, going back to the expression for $\bar{M}^{(k)} - M^{(k)}$, we have

$$\|(1)\|_2 \le \lambda_{\min}(W_k)^{-1/2}\|\hat{W}_k - W_k\|_2 O_P(\sqrt{k})O_P(\bar{\lambda}_k^{-1}k^{-1})O_P(\sqrt{k})O_P(\sqrt{\bar{\lambda}_k}) = O_P(\bar{\lambda}_k^{-1}\|\hat{W}_k - W_k\|_2).$$

Similarly, we can verify that $\|(5)\|_2 = O_P(\bar{\lambda}_k^{-1}\|\hat{W}_k - W_k\|_2)$ and

$$\|(2)\|_2 \le \bar{\lambda}_k^{1/2}O_P(\sqrt{k}[n^{-1/2} \vee \|\bar{\theta} - \theta_0\|_2])O_P(\bar{\lambda}_k^{-1}k^{-1})O_P(\sqrt{k})O_P(\bar{\lambda}_k^{1/2}) = O_P(n^{-1/2} \vee \|\bar{\theta} - \theta_0\|_2).$$

The same holds for $\|(4)\|_2$. Finally,

$$\|(3)\|_2 \le O_P(\bar{\lambda}_k)O_P(k)\left[O_P(\bar{\lambda}_k^{-1}k^{-1}[n^{-1/2} \vee \|\bar{\theta} - \theta_0\|_2]) + O_P(\bar{\lambda}_k^{-2}k^{-1}\|\hat{W}_k - W_k\|_2)\right]$$
$$= O_P([n^{-1/2} \vee \|\bar{\theta} - \theta_0\|_2]) + O_P(\bar{\lambda}_k^{-1}\|\hat{W}_k - W_k\|_2)$$

and the result follows.

(vi) Note that

$$\|\Delta_{1n}\|_2 = \|\bar{H}'\hat{W}_k^{1/2}\bar{M}^{(k)}\hat{W}_k^{1/2}\bar{H} - H'W_k^{1/2}M^{(k)}W_k^{1/2}H\|_2$$
$$\leq \|\bar{H}-H\|_2\|\hat{W}_k\|_2\|\bar{H}\|_2+\|H\|_2\left\|\hat{W}_k^{1/2}-W_k^{1/2}\right\|_2\|\hat{W}_k^{1/2}\|_2\|\bar{H}\|_2+\|H\|_2\|W_k^{1/2}\|_2\|\bar{M}^{(k)}-M^{(k)}\|_2\|\bar{H}\|_2\|\hat{W}_k^{1/2}\|_2$$
$$+ \|H\|_2\|W_k^{1/2}\|_2\left\|\hat{W}_k^{1/2}-W_k^{1/2}\right\|_2\|\bar{H}\|_2 + \|H\|_2\|W_k\|_2\|\bar{H}-H\|_2.$$

The result follows by the same steps as above.

(vii) Similarly,

$$\|\Delta_{2n}\|_2 = \|\bar{H}'\hat{W}_k^{1/2}\bar{M}^{(k)}\hat{W}_k^{1/2}\bar{f}_k - H'W_k^{1/2}M^{(k)}W_k^{1/2}\bar{f}_k\|_2$$
$$\leq \|\bar{H}-H\|_2\|\hat{W}_k\|_2\|\bar{f}_k\|_2+\|H\|_2\left\|\hat{W}_k^{1/2}-W_k^{1/2}\right\|_2\|\hat{W}_k^{1/2}\|_2\|\bar{f}_k\|_2+\|H\|_2\|W_k^{1/2}\|_2\|\bar{M}^{(k)}-M^{(k)}\|_2\|\hat{W}_k^{1/2}\|_2\|\bar{f}_k\|_2$$
$$+ \|H\|_2\|W_k^{1/2}\|_2\left\|\hat{W}_k^{1/2}-W_k^{1/2}\right\|_2\|\bar{f}_k\|_2$$

and the result follows along similar lines as above.

(viii) We have

$$\|\hat{M}^{(k)}-\bar{M}^{(k)}\|_2 \leq \|\hat{W}_k\|_2\Big(\|\hat{D}_1-\bar{D}_1\|_2\|(\hat{D}_1'\hat{W}_k\hat{D}_1)^{-1}\|_2\|\hat{D}_1\|_2 + \|\hat{D}_1\|_2\|\bar{D}_1\|_2\|(\hat{D}_1'\hat{W}_k\hat{D}_1)^{-1} - (\bar{D}_1'\hat{W}_k\bar{D}_1)^{-1}\|_2$$
$$+ \|\bar{D}_1\|_2\|(\bar{D}_1'\hat{W}_k\bar{D}_1)^{-1}\|_2\|\hat{D}-\bar{D}_1\|_2\Big) := (1)+(2)+(3).$$

From (iv), $\|(\bar{D}_1'\hat{W}_k\bar{D})_1^{-1}\|_2 = O_P(\bar{\lambda}_k^{-1}k^{-1})$ and $\|(\hat{D}_1'\hat{W}_k\hat{D}_1)^{-1}\|_2 = O_P(\bar{\lambda}_k^{-1}k^{-1})$. Note that

$$\|\hat{D}_1-\bar{D}_1\|_2 \leq \|\hat{D}_1-D_1\|_2 + \|\bar{D}_1-D_1\|_2 \leq O_P(\sqrt{k}[n^{-1/2}\vee\|\hat{\theta}-\theta_0\|_2]) + O_P(\sqrt{k}[n^{-1/2}\vee\|\bar{\theta}-\theta_0\|_2]).$$

Since $\bar{\theta}\in(\theta_0,\hat{\theta})$, we can write $\bar{\theta} = t\theta_0 + (1-t)\hat{\theta}$ for $t\in(0,1)$. Therefore, we can claim that $\|\hat{D}_1-\bar{D}_1\|_2 = O_P(\sqrt{k}[n^{-1/2}\vee\|\hat{\theta}-\theta_0\|_2])$. Using similar arguments as above, we can also show that

$$\|(\hat{D}_1'\hat{W}_k\hat{D}_1)^{-1} - (\bar{D}_1'\hat{W}_k\bar{D}_1)^{-1}\|_2 = O_P(\bar{\lambda}_k^{-1}k^{-1}[n^{-1/2}\vee\|\hat{\theta}-\theta_0\|_2]).$$

By combining this with the fact that $\|\hat{W}_k\|_2 = O_P(\bar{\lambda}_k)$, $\|\bar{D}_1\|_2 = O_P(\sqrt{k})$ and $\|\hat{D}_1\|_2 = O_P(\sqrt{k})$, the result follows readily. □

Proof of Equation (B.6) in Appendix B

We have

$$\bar{f}_k(\hat{\theta})'\hat{W}_k^{1/2}\bar{P}^{(k)}\hat{W}_k^{1/2}\bar{f}_k(\hat{\theta}) - \bar{f}_k(\theta_0)'\hat{W}_k^{1/2}\bar{P}^{(k)}\hat{W}_k^{1/2}\bar{f}_k(\theta_0) = 2\bar{f}_k(\hat{\theta})'\hat{W}_k^{1/2}\bar{P}^{(k)}\hat{W}_k^{1/2}\bar{f}_k(\hat{\theta})$$
$$- 2\bar{f}_k(\theta_0)'\hat{W}_k^{1/2}\bar{P}^{(k)}\hat{W}_k^{1/2}\bar{f}_k(\hat{\theta}) - (\bar{f}_k(\hat{\theta})-\bar{f}_k(\theta_0))'\hat{W}_k^{1/2}\bar{P}^{(k)}\hat{W}_k^{1/2}(\bar{f}_k(\hat{\theta})-\bar{f}_k(\theta_0)).$$

From (B.3) and (B.4), $\|\bar{f}_k(\hat{\theta}) - \bar{f}_k(\theta_0)\|_2 = O_P(\sqrt{k}\|\hat{\eta} - \eta_0\|_2) = O_P(\sqrt{k}\|\hat{\theta} - \theta_0\|_2)$. Hence,

$$|(\bar{f}_k(\hat{\theta}) - \bar{f}_k(\theta_0))'\hat{W}_k^{1/2}\bar{P}^{(k)}\hat{W}_k^{1/2}(\bar{f}_k(\hat{\theta}) - \bar{f}_k(\theta_0))| \leq \lambda_{\max}(\hat{W}_k)\|\bar{f}_k(\hat{\theta}) - \bar{f}_k(\theta_0)\|_2^2 = O_P(\lambda_k k\|\hat{\theta} - \theta_0\|_2^2)).$$

Let $\hat{D}_1 := \bar{D}_1(\hat{\theta})$. By the first-order necessary optimality condition, $\hat{D}_1'\hat{W}_k\bar{f}_k(\hat{\theta}) = 0$.
Let $\hat{P}^{(k)} := \hat{W}_k^{1/2}\hat{D}_1(\hat{D}_1'\hat{W}_k\hat{D}_1)^{-1}\hat{D}_1'\hat{W}_k^{1/2}$. Obviously, $\hat{P}^{(k)}\hat{W}_k^{1/2}\bar{f}_k(\hat{\theta}) = 0$. Thus,

$$\bar{P}^{(k)}\hat{W}_k^{1/2}\bar{f}_k(\hat{\theta}) = (\bar{P}^{(k)} - \hat{P}^{(k)})\hat{W}_k^{1/2}\bar{f}_k(\hat{\theta}) \quad \text{and} \quad \|\bar{P}^{(k)}\hat{W}_k^{1/2}\bar{f}_k(\hat{\theta})\|_2 \leq \|\bar{P}^{(k)} - \hat{P}^{(k)}\|_2 \cdot O_P(\sqrt{\bar{\lambda}_k k}).$$

From Lemma B.3, $\|\bar{P}^{(k)} - \hat{P}^{(k)}\|_2 = O_P(n^{-1/2} \vee \|\hat{\theta} - \theta_0\|_2)$ and it follows that

$$\|\bar{P}^{(k)}\hat{W}_k^{1/2}\bar{f}_k(\hat{\theta})\|_2 = O_P(\sqrt{k\bar{\lambda}_k}[n^{-1/2} \vee \|\hat{\theta} - \theta_0\|_2]).$$

Thus:

$$\bar{f}_k(\hat{\theta})'\hat{W}_k^{1/2}\bar{P}^{(k)}\hat{W}_k^{1/2}\bar{f}_k(\hat{\theta}) = O_P(\bar{\lambda}_k k[n^{-1} \vee \|\hat{\theta} - \theta_0\|_2^2])$$

and

$$\bar{f}_k(\theta_0)'\hat{W}_k^{1/2}\bar{P}^{(k)}\hat{W}_k^{1/2}\bar{f}_k(\hat{\theta}) = O_P(\bar{\lambda}_k k[n^{-1/2} \vee \|\hat{\theta} - \theta_0\|_2]).$$

As a result,

$$|\bar{f}_k(\hat{\theta})'\hat{W}_k^{1/2}\bar{P}^{(k)}\hat{W}_k^{1/2}\bar{f}_k(\hat{\theta}) - \bar{f}_k(\theta_0)'\hat{W}_k^{1/2}\bar{P}^{(k)}\hat{W}_k^{1/2}\bar{f}_k(\theta_0)| = O_P(\bar{\lambda}_k k n^{-1/2}) + O_P(\bar{\lambda}_k k\|\hat{\theta} - \theta_0\|_2). \quad \square$$

Proofs of Proposition 4.1 and Theorem 4.2

Proof of Proposition 4.1: We shall assume, without loss of generality, that $V_k = I_k$. This amounts to taking the scaled version of $f_k(\theta_0)$ by $V_k^{-1/2}$. In the following expressions, we let $f_i = f_k(x_i)$. We have

$$\begin{aligned} Var(U_n) &= \frac{1}{n^2 k}\mathbb{E}\left(\sum_{i=1}^{n}\sum_{j\neq i} f_i' f_j\right)^2 = \frac{1}{n^2 k}\sum_{i_1\neq j_1}\sum_{i_2\neq j_2}\mathbb{E}(f_{i_1}' f_{j_1} \cdot f_{i_2}' f_{j_2}) \\ &= \frac{2}{n^2 k}\sum_{i\neq j}\mathbb{E}\left[(f_i' f_j)^2\right] + \frac{1}{n^2 k}\sum_{\substack{i_1\neq j_1, i_2\neq j_2 \\ \text{3 diff. indices}}}\mathbb{E}\left[(f_{i_1}' f_{j_1}) \cdot (f_{i_2}' f_{j_2})\right] \\ &\quad + \frac{1}{n^2 k}\sum_{\substack{i_1\neq j_1, i_2\neq j_2 \\ \text{4 diff. indices}}}\mathbb{E}\left[(f_{i_1}' f_{j_1}) \cdot (f_{i_2}' f_{j_2})\right]. \end{aligned} \quad \text{(OA.2)}$$

Because of the martingale difference dynamics of f_i, all the terms in the last summation are 0. We next show that the second expression is $o(1)$ and the first one is equal to $2 + o(1)$.

(a) Consider:

$$A_n = \frac{1}{n^2 k} \sum_{\substack{i \neq j_1, i \neq j_2 \\ j_1 \neq j_2}} \mathbb{E}[f_i' f_{j_1} \cdot f_i' f_{j_2}].$$

Let $\pi_n = n^a$, $a \in (0,1)$. In the following, we will consider two indices i, j to be connected iff $|i - j| \leq \pi_n$ and a set of three indices to be connected iff any one of them is connected to at least another one of them.

In the summation above, we have three configurations that stand out:

(i) The three indices i, j_1, j_2 are connected. Denote S_1 to be the collection of such indices.

(ii) Only two of the indices, say $\{i, j_1\}$, $\{i, j_2\}$ or $\{j_1, j_2\}$ are connected. Denote S_2, S_3 and S_4 the subset of indices (i, j_1, j_2) satisfying these descriptions, respectively.

(iii) All the three indices are isolated from one another. Denote S_5 to be the collection of such indices.

We first deal with (i). We have

$$\begin{aligned}|\mathbb{E}[f_i' f_{j_1} \cdot f_i' f_{j_2}]| &= \left| \sum_{h,h'=1}^{k} \mathbb{E}[f_{ih} f_{j_1 h} f_{ih'} f_{j_2 h'}] \right| \leq \sum_{h,h'=1}^{k} \mathbb{E}|f_{ih} f_{j_1 h} f_{ih'} f_{j_2 h'}| \\ &\leq \sum_{h,h'=1}^{k} \left[\mathbb{E}(f_{ih}^4)\mathbb{E}(f_{j_1 h}^4)\mathbb{E}(f_{ih'}^4)\mathbb{E}(f_{j_2 h'}^4)\right]^{1/4} \leq \sum_{h,h'=1}^{k} \left(\mathbb{E}|f_{ih}|^{4+\epsilon}\right)^{2/4+\epsilon} \left(\mathbb{E}|f_{ih'}|^{4+\epsilon}\right)^{2/4+\epsilon},\end{aligned}$$

where the first inequality is the triangle inequality, the second one follows from the Cauchy-Schwarz inequality, the third one follows from the monotonicity (in p) of L_p-norms and the stationarity assumption. Hence,

$$\begin{aligned}|\mathbb{E}[f_i' f_{j_1} \cdot f_i' f_{j_2}]| &\leq k^2 \left(\frac{1}{k} \sum_{h=1}^{k} \left(\mathbb{E}|f_{ih}|^{4+\epsilon}\right)^{2/4+\epsilon} \right)^2 \leq k^2 \left(\frac{1}{k} \sum_{h=1}^{k} \left(\mathbb{E}|f_{ih}|^{4+\epsilon}\right)^{4/4+\epsilon} \right) \\ &\leq k^2 \left(\frac{1}{k} \sum_{h=1}^{k} \mathbb{E}|f_{ih}|^{4+\epsilon} \right)^{4/4+\epsilon} \leq k^2 \Delta_\epsilon,\end{aligned}$$

where $\Delta_\epsilon = \Delta^{4/4+\epsilon}$ is an absolute constant. The last inequality holds by assumption and the previous two follow from the Jensen's inequality. Thus,

$$A_{1n} \equiv \left| \frac{1}{n^2 k} \sum_{(i,j_1,j_2) \in S_1} \mathbb{E}[f_i' f_{j_1} \cdot f_i' f_{j_2}] \right| \leq N(S_1) \frac{k^2 \Delta_\epsilon}{n^2 k},$$

where $N(S)$ denotes the cardinality of S. It is not hard to see that $N(S_1) \leq n\pi_n^2$. Then,

$$A_{1n} \leq \Delta_\epsilon k \pi_n^2 / n.$$

Choosing $\pi_n = o(\sqrt{n/k})$ is sufficient to claim that $A_{1n} = o(1)$.

We next deal with (ii). Assume that (i, j_1) are connected and j_2 is isolated from both so that $(i, j_1, j_2) \in S_2$. Take $p_1 = \frac{4+\epsilon}{3}$ and $p_2 = 4 + \epsilon$. Let $q = \frac{1}{p_1} + \frac{1}{p_2} = \frac{4}{4+\epsilon}$. We have $q < 1$ and by Lemma OA.1, it follows that

$$\begin{aligned}\mathbb{E}(f_{ih} f_{ih'} f_{j_1 h} f_{j_2 h'}) &= \mathbb{E}(f_{ih} f_{ih'} f_{j_1 h}) \mathbb{E}(f_{j_2 h'}) + O\left(\rho^{(1-q)\pi_n} \|f_{ih} f_{ih'} f_{j_1 h}\|_{p_1} \|f_{j_2 h'}\|_{p_2}\right) \\ &= O\left(\rho^{(1-q)\pi_n} \|f_{ih} f_{ih'} f_{j_1 h}\|_{p_1} \|f_{j_2 h'}\|_{p_2}\right).\end{aligned}$$

By the Cauchy-Schwarz inequality, we have

$$\|f_{ih} f_{ih'} f_{j_1 h}\|_{p_1} \|f_{j_2 h'}\|_{p_2} \leq \left(\mathbb{E}|f_{ih}|^{4+\epsilon}\right)^{\frac{1}{4+\epsilon}} \left(\mathbb{E}|f_{ih'}|^{4+\epsilon}\right)^{\frac{1}{4+\epsilon}} \left(\mathbb{E}|f_{j_1 h}|^{4+\epsilon}\right)^{\frac{1}{4+\epsilon}} = \left(\mathbb{E}|f_{ih}|^{4+\epsilon}\right)^{\frac{2}{4+\epsilon}} \left(\mathbb{E}|f_{ih'}|^{4+\epsilon}\right)^{\frac{1}{4+\epsilon}},$$

where the last equality holds by stationarity. Hence,

$$\mathbb{E}(f_{ih} f_{ih'} f_{j_1 h} f_{j_2 h'}) = O(\rho^{(1-q)\pi_n} \|f_{ih}\|_{4+\epsilon}^2 \|f_{ih'}\|_{4+\epsilon}^2).$$

This yields

$$\sum_{h,h'=1}^{k} \mathbb{E}(f_{ih} f_{ih'} f_{j_1 h} f_{j_2 h'}) = O\left(\rho^{(1-q)\pi_n} k^2 \left(\frac{1}{k} \sum_{h=1}^{k} \|f_{ih}\|_{4+\epsilon}^2\right)^2\right).$$

But, by the Jensen's inequality,

$$\begin{aligned}\left(\frac{1}{k} \sum_{h=1}^{k} \|f_{ih}\|_{4+\epsilon}^2\right)^2 &\leq \frac{1}{k} \sum_{h=1}^{k} \|f_{ih}\|_{4+\epsilon}^4 = \frac{1}{k} \sum_{h=1}^{k} \left(\mathbb{E}|f_{ih}|^{4+\epsilon}\right)^{\frac{4}{4+\epsilon}} \\ &\leq \left(\frac{1}{k} \sum_{h=1}^{k} \mathbb{E}|f_{ih}|^{4+\epsilon}\right)^{\frac{4}{4+\epsilon}} \leq \left(\sup_k \frac{1}{k} \sum_{h=1}^{k} \mathbb{E}|f_{ih}|^{4+\epsilon}\right)^{\frac{4}{4+\epsilon}} = \Delta_\epsilon < \infty.\end{aligned}$$

Thus,

$$\sum_{h,h'=1}^{k} \mathbb{E}(f_{ih} f_{ih'} f_{j_1 h} f_{j_2 h'}) = O\left(k^2 \rho^{(1-q)\pi_n}\right).$$

Hence,

$$A_{2n} \equiv \left| \frac{1}{n^2 k} \sum_{(i,j_1,j_2) \in S_2} \mathbb{E}[f_i' f_{j_1} \cdot f_i' f_{j_2}] \right| = O\left(\frac{N(S_2)}{n^2 k} k^2\right).$$

Clearly, $N(S_2) \leq n^2 \pi_n$. Thus, choosing $\pi_n = n^a$ for some $a > 0$ ensures that $A_{2n} = o(1)$. Likewise, summation over $(i, j_1, j_2) \in S_3$ leads to a negligible quantity.

Now, consider S_4, where j_1 and j_2 are connected while i is isolated. Applying Lemma OA.1, we have

$$\begin{aligned}\mathbb{E}(f_{ih} f_{ih'} f_{j_1 h} f_{j_2 h'}) &= \mathbb{E}(f_{ih} f_{ih'}) \mathbb{E}(f_{j_1 h} f_{j_2 h'}) + O\left(\rho^{(1-q)\pi_n} \|f_{ih} f_{ih'}\|_{p_1} \|f_{j_1 h} f_{j_2 h'}\|_{p_2}\right) \\ &= O\left(\rho^{(1-q)\pi_n} \|f_{ih} f_{ih'} f_{j_1 h}\|_{p_1} \|f_{j_2 h'}\|_{p_2}\right),\end{aligned}$$

with $p_1 = p_2 = \frac{4+\epsilon}{2}$, and then, $q = \frac{4}{4+\epsilon}$. From the Cauchy-Schwarz inequality and the stationarity assumption, we have

$$\|f_{ih}f_{ih'}f_{j_1h}\|_{p_1}\|f_{j_2h'}\|_{p_2} \leq \left(|f_{ih}|^{4+\epsilon}\right)^{\frac{2}{4+\epsilon}}\left(|f_{ih'}|^{4+\epsilon}\right)^{\frac{2}{4+\epsilon}}.$$

Similar derivations as above lead to

$$A_{3n} \equiv \left|\frac{1}{n^2k}\sum_{(i,j_1,j_2)\in S_4}\mathbb{E}[f_i'f_{j_1}\cdot f_i'f_{j_2}]\right| = O\left(\frac{N(S_4)}{n^2k}k^2\right).$$

Again, $N(S_4) \leq n^2\pi_n$. Thus, choosing $\pi_n = n^a$ for some $a > 0$ ensures that $A_{3n} = o(1)$.

We next consider (iii), where $\{i, j_1, j_2\}$ has no pairs connected. Again, by Lemma OA.1, we have

$$\begin{aligned}\mathbb{E}(f_{ih}f_{ih'}f_{j_1h}f_{j_2h'}) &= \mathbb{E}(f_{ih}f_{ih'})\mathbb{E}(f_{j_1h})\mathbb{E}(f_{j_2h'}) + O\left(\rho^{(1-q)\pi_n}\|f_{ih}f_{ih'}\|_{p_1}\|f_{j_1h}\|_{p_2}\|f_{j_2h'}\|_{p_3}\right)\\ &= O\left(\rho^{(1-q)\pi_n}\|f_{ih}f_{ih'}\|_{p_1}\|f_{j_1h}\|_{p_2}\|f_{j_2h'}\|_{p_3}\right),\end{aligned}$$

with $p_1 = \frac{4+\epsilon}{2}$, and $p_2 = p_3 = 4+\epsilon$. By the Cauchy-Schwarz inequality and stationarity, we have, as before,

$$\|f_{ih}f_{ih'}\|_{p_1}\|f_{j_1h}\|_{p_2}\|f_{j_2h'}\|_{p_3} \leq \left(\mathbb{E}(|f_{ih}|^{4+\epsilon}\right)^{\frac{2}{4+\epsilon}}\left(\mathbb{E}(|f_{ih'}|^{4+\epsilon}\right)^{\frac{2}{4+\epsilon}}.$$

Hence,

$$\sum_{h,h'=1}^{k}\mathbb{E}(f_{ih}f_{ih'}f_{j_1h}f_{j_1h'}) = O(\rho^{(1-q)\pi_n}k^2).$$

Thus,

$$A_{4n} \equiv \left|\frac{1}{n^2k}\sum_{(i,j_1,j_2)\in S_5}\mathbb{E}[f_i'f_{j_1}\cdot f_i'f_{j_2}]\right| = O\left(\frac{1}{n^2k}N(S_5)\rho^{(1-q)\pi_n}k^2\right).$$

Note that $N(S_5) \leq n^3$. Thus $A_{4n} = O(n^2\rho^{(1-q)\pi_n})$. Taking $\pi_n = n^a$ for some $a > 0$ ensures that $A_{4n} = o(1)$.

Therefore, since $k = n^\alpha$, with $\alpha \in (0,1)$, we can always find $\pi_n = n^a$ with $a > 0$ small enough so that all A_{1n}, A_{2n}, A_{3n} and A_{4n} are all $o(1)$. This shows that $A_n = o(1)$. Hence, the second term in the expression of the variance of U_n in (OA.2) is $o(1)$; that is,

$$\frac{1}{n^2k}\sum_{\substack{i_1\neq j_1, i_2\neq j_2\\ \text{3 diff. indices}}}\mathbb{E}\left[(f_{i_1}'f_{j_1})\cdot(f_{i_2}'f_{j_2})\right] = o(1).$$

(b) Consider the first term in the expression of the variance of U_n in (OA.2). We have

$$\frac{1}{n^2k}\sum_{i\neq j}\mathbb{E}\left[(f_i'f_j)^2\right] = \frac{1}{n^2k}\sum_{i\neq j}\sum_{h,h'=1}^{k}\mathbb{E}(f_{ih}f_{jh}f_{ih'}f_{jh'}).$$

As previously, we can show that

$$|\mathbb{E}(f_{ih}f_{jh}f_{ih'}f_{jh'})| \le (\mathbb{E}|f_{ih}|^{4+\epsilon})^{\frac{2}{4+\epsilon}}(\mathbb{E}|f_{jh}|^{4+\epsilon})^{\frac{2}{4+\epsilon}}$$

and

$$\sum_{h,h'=1}^{k} \mathbb{E}(f_{ih}f_{jh}f_{ih'}f_{jh'}) \le k^2 \cdot \Delta_\epsilon.$$

If i, j are connected,

$$A_{5n} \equiv \left| \frac{1}{n^2 k} \sum_{\substack{i\neq j \\ |i-j|\le \pi_n}} \sum_{h,h'=1}^{k} \mathbb{E}(f_{ih}f_{jh}f_{ih'}f_{jh'}) \right| \le \frac{Nk^2\Delta_\epsilon}{n^2 k} = O\left(\frac{k\pi_n}{n}\right),$$

where $N \le 2n\pi_n$ is the number of connected pairs (i, j). Choosing $\pi_n = o(n/k)$ is enough to claim that $A_{5n} = o(1)$.

If i, j are not connected, $|i - j| \ge \pi_n$, we have

$$\mathbb{E}(f_{ih}f_{jh}f_{ih'}f_{jh'}) = \mathbb{E}(f_{ih}f_{ih'})\mathbb{E}(f_{jh}f_{jh'}) + O\left(\rho^{(1-q)\pi_n}\|f_{ih}f_{ih'}\|_{p_1}\|f_{jh}f_{jh'}\|_{p_2}\right),$$

with $p_1 = p_2 = \frac{4+\epsilon}{2}$ and $q = \frac{4}{4+\epsilon}$. As in the previous lines, we can claim that

$$\frac{1}{n^2 k} \sum_{\substack{i\neq j \\ |i-j|\ge \pi_n}} \sum_{h,h'=1}^{k} \mathbb{E}(f_{ih}f_{jh}f_{ih'}f_{jh'}) = \frac{1}{n^2 k} \sum_{\substack{i\neq j \\ |i-j|\ge \pi_n}} \mathbb{E}^*((f_i' f_j)^2) + O\left(k\rho^{(1-q)\pi_n}\right),$$

where E^* denote expectation under independence of f_i and f_j. By taking $\pi_n = n^a$ for some $a > 0$ small enough, the second term in the right hand side is $o(1)$. Note that the number N of connected pairs (i, j) satisfies:

$$n(n - 2\pi_n) \le N \le n^2.$$

Also,

$$\mathbb{E}^*(f_i' f_j)^2 = \mathbb{E}^*(f_i' f_j f_j' f_i) = \text{trace}(\mathbb{E}^*[f_j f_j' f_i f_i']) = k.$$

It follows that $\frac{1}{n^2 k}\sum_{i\neq j, |i-j|\ge \pi_n} E^*((f_i' f_j)^2) = 1 + o(1)$ and as a result,

$$\frac{1}{n^2 k}\sum_{i\neq j} \mathbb{E}\left[(f_i' f_j)^2\right] = 1 + o(1).$$

This completes the proof. $\square$

Proof of Theorem 4.2: As previously, we assume, without loss of generality, that $V_k = I_k$. Again, we let $f_i := f_k(x_i)$ and set $z_i := \max_{1\leq i\leq n} \|f_i\|/\sqrt{k}$ and $z_{ij} := \max_{1\leq i\neq j\leq n} |f_i' f_j|/\sqrt{k}$. Let $M > 0$. Then, we have

$$U_n = U_n 1_{\{(z_i\leq M\log^\beta n)\cap(z_{ij}\leq M\log^\beta n)\}} + U_n 1_{\{(z_i> M\log^\beta n)\cup(z_{ij}> M\log^\beta n)\}}. \tag{OA.3}$$

For the second term in this expression, we have

$$\mathbb{E}[|U_n|1_{\{z_i>M\log^\beta n\cup z_{ij}>M\log^\beta n\}}] \leq \mathbb{E}[|U_n|1_{\{z_i>M\log^\beta n\}}] + \mathbb{E}[|U_n|1_{\{z_{ij}>M\log^\beta n\}}].$$

By the Cauchy-Schwarz and the Markov inequalities,

$$\mathbb{E}[|U_n|1_{\{z_i>M\log^\beta n\}}] \leq \Big(\mathbb{E}(U_n^2)\Pr(z_i > M\log^\beta n)\Big)^{1/2} \leq \frac{1}{\sqrt{M\log^\beta n}}\left(\mathbb{E}(U_n^2)\mathbb{E}(z_i)\right)^{1/2}.$$

Thus, by Proposition 4.1 and Assumption-CLT 3, there exists a constant $c > 0$ such that

$$\sup_n \mathbb{E}[|U_n|1_{\{z_i>M\log^\beta n\}}] \leq \frac{c}{\sqrt{M}}.$$

The same claim can be made about $\sup_n E[|U_n|1_{\{z_{ij}>M\log^\beta n\}}]$ and it then follows that

$$\lim_{M\to\infty}\sup_n \mathbb{E}[|U_n|1_{\{(z_i>M\log^\beta n)\cup(z_{ij}>M\log^\beta n)\}}] = 0.$$

Hence, $U_n 1_{\{(z_i>M\log^\beta n)\cup(z_{ij}>M\log^\beta n)\}}$ can be made stochastically small by taking M large and we can only focus on the first term in (OA.3). We shall therefore consider throughout that

$$\max_{1\leq i\leq n}\frac{\|f_i\|}{\sqrt{k}\log^\beta n} \leq M \text{ and } \max_{1\leq i\neq j\leq n}\frac{|f_i' f_j|}{\sqrt{k}\log^\beta n} \leq M \tag{OA.4}$$

for a fixed constant $M > 0$.

Next, we show that the moments of $U_n/\sqrt{2}$ converge to those of the standard normal distribution. Let $r \in N$. We have

$$\begin{aligned} U_n^{2r+1} &= \frac{1}{(n\sqrt{k})^{2r+1}} \sum_{\substack{i_1,j_1,\ldots,i_{2r+1},j_{2r+1}=1\\ i_a\neq j_a, a=1,\ldots,2r+1}}^{n} \prod_{s=1}^{2r+1} f_{i_s}' f_{j_s} \\ &= \frac{1}{(n\sqrt{k})^{2r+1}} \sum_{\substack{i_1,j_1,\ldots,i_{2r+1},j_{2r+1}=1\\ i_a\neq j_a, a=1,\ldots,2r+1}}^{n} \sum_{h_1,\ldots,h_{2r+1}=1}^{k} \prod_{s=1}^{2r+1} f_{i_s,h_s} f_{j_s,h_s}. \end{aligned} \tag{OA.5}$$

To derive this moment, we shall rely, following Kim et al. (2011), on some notions from graph theory. To make the discussion self-contained, we introduce some of these notions below. More details may be found in Kim et al. (2011). Each term of in the right-hand side of (OA.5) can be associated with an undirected graph with vertices $i_1, j_1, i_2, j_2, \ldots, i_{2r+1}, j_{2r+1}$. Let $\pi_n(\leq n)$ be an increasing sequence of n. We say that a_1, a_2 in the graph $\{i_1, j_1, i_2, j_2, \ldots, i_{2r+1}, j_{2r+1}\}$ are connected if: $|a_1 - a_2| \leq \pi_n$ or there exist $\{b_1, \ldots b_l\} \subset \{i_1, j_1, i_2, j_2, \ldots, i_{2r+1}, j_{2r+1}\}$ such that

$$|a_1 - b_1| \leq \pi_n, |b_1 - b_2| \leq \pi_n, \ldots, |b_{l-1} - b_l| \leq \pi_n, |b_l - a_2| \leq \pi_n.$$

Note that in a graph, many i_s and/or j_s may take the same value. A component of the graph is a subset I of $\{i_1, j_1, i_2, j_2, \ldots, i_{2r+1}, j_{2r+1}\}$ such that every vertex in I is connected to at least another one. A graph can be partitioned into m components: $I_1, \ldots I_m$. We may assume, without loss of generality, that the components are arranged in the increasing order of vertices so that, for $u < v$,

$$i < j \text{ for all } i \in \mathcal{I}_u \text{ and } j \in \mathcal{I}_v.$$

The distance between two successive components I_u and I_{u+1} is defined as

$$d_u := d(\mathcal{I}_u, \mathcal{I}_{u+1}) = \inf_{i \in \mathcal{I}_u, j \in \mathcal{I}_{u+1}} (j - i).$$

Note that $d_u \geq \pi_n$ for any u. For given k components $I_1, \ldots, I_m$, let $d_{(u)}$ denote the u-th smallest distance among $d_1, \ldots, d_{m-1}$. The size of a component is the number of vertices (accounting for multiplicity) it contains.

Suppose that the graph $G = \{i_1, j_1, i_2, j_2, \ldots, i_{2r+1}, j_{2r+1}\}$ comprises m components.

Since $\|f_i\|/\sqrt{k} \log^\beta n \leq M$ for all i, we also have $|f_{i,h}|/\sqrt{k} \log^\beta n \leq M$ for all i and $h = 1, \ldots, k$. We can then apply Lemma OA.1 with $p_1 = \cdots = p_k = \infty$ and claim that

$$\mathbb{E}\left(\prod_{s=1}^{2r+1} f_{i_s,h_s} f_{j_s,h_s}\right) = \prod_{\ell=1}^{m} \mathbb{E}\left(\prod_{s: i_s \in \mathcal{I}_\ell} f_{i_s,h_s} \prod_{s: j_s \in \mathcal{I}_\ell} f_{j_s,h_s}\right) + O\left(\rho^{d_{(1)}} k^{2r+1} \log^{\beta(4r+1)} n\right).$$

We shall use the type of decomposition above routinely throughout our subsequent derivations.

Let G_m be the set of graphs having exactly m components. We have

$$\begin{aligned}
\mathbb{E}(U_n^{2r+1}) &= \frac{1}{(n\sqrt{k})^{2r+1}} \sum_{m=1}^{4r+2} \sum_{\substack{i_1, j_1, \ldots, i_{2r+1}, j_{2r+1} \in \mathcal{G}_m \\ i_a \neq j_a, a=1,\ldots,2r+1}} \mathbb{E}\left(\prod_{s=1}^{2r+1} f'_{i_s} f_{j_s}\right) \\
&= \frac{1}{(n\sqrt{k})^{2r+1}} \sum_{m=1}^{4r+2} \sum_{\substack{i_1, j_1, \ldots, i_{2r+1}, j_{2r+1} \in \mathcal{G}_m \\ i_a \neq j_a, a=1,\ldots,2r+1}} \sum_{h_1, \ldots, h_{2r+1}=1}^{n} \mathbb{E}\left(\prod_{s=1}^{2r+1} f_{i_s,h_s} f_{j_s,h_s}\right). \quad \text{(OA.6)}
\end{aligned}$$

(a) Consider a graph with $m \geq 2r+2$ components. Then there exists at least one component, $I_{\ell_0} = \{i_0\}$, with size 1 and

$$\mathbb{E}\left(\prod_{s:i_s \in \mathcal{I}_{\ell_0}} f_{i_s,h_s} \prod_{s:j_s \in \mathcal{I}_{\ell_0}} f_{j_s,h_s}\right) = \mathbb{E}(f_{i_0,h_s}) = 0.$$

Thus, for such graphs,

$$\mathbb{E}\left(\prod_{s=1}^{2r+1} f_{i_s,h_s} f_{j_s,h_s}\right) = O\left(\rho^{d_{(1)}} k^{2r+1} \log^{\beta(4r+2)} n\right) = O\left(\rho^{\pi_n} k^{2r+1} \log^{\beta(4r+2)} n\right).$$

Clearly, the total number of graphs with at least $2r+2$ components is less than or equal to $(n(n-1))^{2r+1}$, the total number of graphs in the expansion in (OA.6). As a result,

$$\frac{1}{(n\sqrt{k})^{2r+1}} \sum_{m=2r+2}^{4r+2} \sum_{\substack{i_1,j_1,\ldots,i_{2r+1},j_{2r+1} \in \mathcal{G}_m \\ i_a \neq j_a, a=1,\ldots,2r+1}} \sum_{h_1,\ldots,h_{2r+1}=1}^{n} \mathbb{E}\left(\prod_{s=1}^{2r+1} f_{i_s,h_s} f_{j_s,h_s}\right)$$
$$= O\left(n^{2r+1} k^{3(r+1/2)} \rho^{\pi_n} \log^{\beta(4r+2)} n\right) = o(1),$$

where the final order of magnitude is obtained by setting $\pi_n = n^a$ for some $a \in (0,1)$.

(b) Consider graphs with $m \leq 2r$ components. Recalling (OA.4), we have

$$A_n := \left| \frac{1}{(n\sqrt{k})^{2r+1}} \sum_{m=1}^{2r} \sum_{\substack{i_1,j_1,\ldots,i_{2r+1},j_{2r+1} \in \mathcal{G}_m \\ i_a \neq j_a, a=1,\ldots,2r+1}} \mathbb{E}\left(\prod_{s=1}^{2r+1} f'_{i_s} f_{j_s}\right) \right| \leq \frac{c_{M,r}(\sqrt{k}\log^\beta n)^{2r+1}}{(n\sqrt{k})^{2r+1}} \sum_{m=1}^{2r} N_m,$$

where N_m is the number of graphs with exactly m components and $c_{M,r}$ is a constant depending only on M and r. We next find an upper bound for N_m. Define the degree of a component as the number of different vertices it contains and let δ_ℓ denote the degree of the component I_ℓ for $\ell = 1, \ldots, m$. Let $N_m(\delta_1, \ldots, \delta_m)$ be the number of graphs with m components with δ_ℓ's being the respective components. Recall that $m \leq \sum_{u=1}^{m} \delta_u \leq 4r+2$. Hence,

$$N_m(\delta_1, \ldots, \delta_m) = n\pi_n^{\delta_1 - 1} n\pi_n^{\delta_2 - 1} \cdots n\pi_n^{\delta_m - 1} = n^m \pi_n^{\delta_1 + \cdots + \delta_m - m} \leq n^m \pi_n^{4r+2-m}.$$

Thus,

$$N_m \leq (4r+2)^m n^m \pi_n^{4r+2-m}.$$

For $1 \leq m \leq 2r$, we have:

$$N_m \leq (4r+2)^{2r} n^{2r} \pi_n^{4r+1} = O\left(n^{2r}\pi_n^{4r+1}\right) \text{ and } \sum_{m=1}^{2r} N_m \leq 2r(4r+2)^{2r} n^{2r} \pi_n^{4r+1} = O\left(n^{2r}\pi_n^{4r+1}\right).$$

Taking $\pi_n = n^{\frac{1}{4r+1}-\epsilon}$ for some $\epsilon \in (0, \frac{1}{4r+1})$ ensures that $A_n = o(1)$.

(c) We are left with the graphs with $m = 2r+1$ components. Note that if there is one component of such graphs that has more than two vertices (accounting for multiplicities), then, there is at least one component with exactly one vertex. Similarly to part (a) of the proof, the contribution of such graphs to the moment in (OA.6) is negligible. We are left with the case where each component has exactly 2 vertices. Again, if any of these components, say I_{ℓ_0}, has size 2, then, by the martingale difference assumption, $E(f_{i,h} f_{j,h'}) = 0$ for $i \neq j \in I_{\ell_0}$ and $h, h' = 1, \ldots, k$. Graphs containing such components also have negligible contribution. There only remain graphs with components each having two equal vertices. Let G'_{2r+1} be the set of all graphs with $2r+1$ components, each of size 1. Note that each graph in G'_{2r+1} contains $2r+1$ different vertices, each featuring exactly twice. We can check along the same lines as in (a) that:

$$\frac{1}{(n\sqrt{k})^{2r+1}} \sum_{\substack{i_1,j_1,\ldots,i_{2r+1},j_{2r+1} \in \mathcal{G}'_{2r+1} \\ i_a \neq j_a, a=1,\ldots,2r+1}} \mathbb{E}\left(\prod_{s=1}^{2r+1} f'_{i_s} f_{j_s}\right) = \frac{1}{(n\sqrt{k})^{2r+1}} \sum_{\substack{i_1,j_1,\ldots,i_{2r+1},j_{2r+1} \in \mathcal{G}'_{2r+1} \\ i_a \neq j_a, a=1,\ldots,2r+1}} \mathbb{E}^*\left(\prod_{s=1}^{2r+1} f'_{i_s} f_{j_s}\right) + o(1),$$

where E^* is the expectation under pairwise independence of $(2r+1)$ distinct $f_\bullet$'s appearing in the expectation.

Take $E := \prod_{s=1}^{2r+1} f'_{i_s} f_{j_s}$ with $i_1, j_1, \ldots, i_{2r+1}, j_{2r+1} \in G'_{2r+1}$. Since each f_{i_s} (or f_{j_s}) appears exactly twice in the product, it either appears in a square inner product: $(f'_{i_s} f_{j_s})^2$ and nowhere else, or it appears in two different inner products: $(f'_{i_s} f_{j_s})(f'_{i_{s'}} f_{j_{s'}})$ (with $i_s = i_{s'}$) and nowhere else. Because we are in the presence of odd number $(2r+1)$ of inner product terms in E, there is at least one i_{s_1} such that $f_{i_{s_1}}$ appears in two different inner products; that is, factors such as $(f'_{i_{s_1}} f_{j_{s_1}})$ and $(f'_{i_{s_2}} f_{j_{s_2}})$ with $i_{s_1} = i_{s_2}$ and $j_{s_1} \neq j_{s_2}$ appear in E.

Let

$$(f'_{i_{s_1}} f_{j_{s_1}}) \cdot (f'_{i_{s_2}} f_{j_{s_2}}) \cdots (f'_{i_{s_u}} f_{j_{s_u}})$$

be a product of factors from E featuring pairwise different inner products but with each vertex appearing exactly twice. Clearly, $u \geq 3$. From the discussion above, such a product can be found for any graph in G'_{2r+1}.

Now, consider $E^*\left(\prod_{s=1}^{2r+1} f'_{i_s} f_{j_s}\right)$. By (OA.4), we have

$$\begin{aligned}
\mathbb{E}^*\left(\prod_{s=1}^{2r+1} f'_{i_s} f_{j_s}\right) &\leq c_{M,r}\left(\sqrt{k}\log^\beta n\right)^{2r+1-u} \mathbb{E}^*\left((f'_{i_{s_1}} f_{j_{s_1}}) \cdot (f'_{i_{s_2}} f_{j_{s_2}}) \cdots (f'_{i_{s_u}} f_{j_{s_u}})\right) \\
&\leq c_{M,r}\left(\sqrt{k}\log^\beta n\right)^{2r-2} \mathbb{E}^*\left((f'_{i_{s_1}} f_{j_{s_1}}) \cdot (f'_{i_{s_2}} f_{j_{s_2}}) \cdots (f'_{i_{s_u}} f_{j_{s_u}})\right) = c_{M,r}\left(\sqrt{k}\log^\beta n\right)^{2r-2} k,
\end{aligned}$$

where $c_{M,r}$ is a constant depending only on M and r. Since the number of graphs in G'_{2r+1} is less than n^{2r+1} (a graph in G'_{2r+1} corresponds in particular to a subset of $2r+1$ elements from a set of n elements), we have

$$\frac{1}{(n\sqrt{k})^{2r+1}} \sum_{\substack{i_1,j_1,\ldots,i_{2r+1},j_{2r+1}\in\mathcal{G}'_{2r+1}\\ i_a\neq j_a, a=1,\ldots,2r+1}} \mathbb{E}^*\left(\prod_{s=1}^{2r+1} f'_{i_s} f_{j_s}\right) \leq \frac{c_{M,r}\left(\sqrt{k}\log^{\beta} n\right)^{2r-2} kn^{2r+1}}{(n\sqrt{k})^{2r+1}} = \frac{c_{M,r}\log^{\beta(2r-2)} n}{\sqrt{k}} = o(1).$$

Parts (a), (b) and (c) allow us to conclude that $E(U_n^{2r+1}) = o(1)$.

(d) Let us now obtain the limit of $E(U_n/\sqrt{2})^{2r}$. Similarly to (OA.6), we can claim that

$$\mathbb{E}((U_n/\sqrt{2})^{2r}) = \frac{1}{(n\sqrt{2k})^{2r}} \sum_{m=1}^{4r} \sum_{\substack{i_1,j_1,\ldots,i_{2r},j_{2r}\in\mathcal{G}_m\\ i_a\neq j_a, a=1,\ldots,2r}} \mathbb{E}\left(\prod_{s=1}^{2r} f'_{i_s} f_{j_s}\right). \tag{OA.7}$$

Here, we shall distinguish the cases: $m \geq 2r+1$, $m \leq 2r-1$ and $m = 2r$. Similarly to the arguments in (a) and (b) above, we can claim that those graphs have negligible contribution to $E((U_n/\sqrt{2})^{2r})$. The discussion in (c) is also valid here and only graphs in G'_{2r} can contribute to this expectation. Also graphs with vertices of equal value appearing in different inner products do not contribute significantly. We are left with graphs featuring only square of inner products. Such graphs exist since the number of inner product in each term of the summation in (OA.7) is even ($2r$). We can write

$$\mathbb{E}((U_n/\sqrt{2})^{2r}) = \frac{1}{(n\sqrt{2k})^{2r}} \sum_{\substack{i_1,j_1,\ldots,i_r,j_r\in\mathcal{G}'_{2r}\\ i_a\neq j_a, a=1,\ldots,r}} \mathbb{E}^*\left(\prod_{s=1}^{r} (f'_{i_s} f_{j_s})^2\right) + o(1).$$

(In this expression, we represent graphs in G'_{2r} that determine $\prod_{s=1}^{r}(f'_{i_s}f_{j_s})^2$ by $i_1, j_1, \ldots, i_r, j_r$ without the need to repeat each vertex.) Recall that $U_n = (2/n)\sum_{i<j}(f'_i f_j/\sqrt{k})$ so that

$$\mathbb{E}\left(\frac{U_n}{\sqrt{2}}\right)^{2r} = 2^{2r}\frac{1}{(n\sqrt{2k})^{2r}} \sum_{\substack{i_1,j_1,\ldots,i_r,j_r\in\mathcal{G}'_{2r}\\ i_a< j_a, a=1,\ldots,r}} \mathbb{E}^*\left(\prod_{s=1}^{r} (f'_{i_s} f_{j_s})^2\right) + o(1). \tag{OA.8}$$

We have

$$\mathbb{E}^*\left(\prod_{s=1}^{r}(f'_{i_s}f_{j_s})^2\right) = \prod_{s=1}^{r}\mathbb{E}^*\left((f'_{i_s}f_{j_s})^2\right) = \prod_{s=1}^{r}\mathbb{E}^*\left(f'_{i_s}f_{j_s}f'_{j_s}f_{i_s}\right) = \prod_{s=1}^{r}\text{trace}\mathbb{E}^*(f_{i_s}f'_{i_s}) = k^r.$$

The number of times that a specific term $\prod_{s=1}^{r}(f'_{i_s}f_{j_s})^2$ appears in the expansion of $E(U_n/\sqrt{2})^{2r}$ is given by the multinomial formula as $(2r)!/2^r$. We are left to determine the cardinality number of graphs in G'_{2r} such that $i_s < j_s$ for $s = 1, \ldots, r$. That is the cardinality of the set

$$S = \left\{\{(i_1,j_1),\ldots,(i_r,j_r)\} \subset \Delta_n : \forall s, s' = 1,\ldots,r, |a_s - a_{s'}| \geq \pi_n \text{ with } a_\bullet = i_\bullet \text{ or } j_\bullet\right\},$$

with $\Delta_n := \{(i,j) : i < j, \text{ and } i,j = 1,\ldots,n\}$.

Define

$$S_r = \{\{(i_1,j_1),\ldots(i_r,j_r)\} \subset \Delta_n\},$$

and

$$S_2 = \left\{\{(i_1,j_1),\ldots(i_r,j_r)\} \subset \Delta_n : \exists s,s' \in \{1,\ldots,r\} : |a_s - a_{s'}| \leq \pi_n\right\}.$$

We have $S = S_r \setminus S_2$ so that $\text{Card}(S) = \text{Card}(S_r) - \text{Card}(S_2)$. Note that:

$$\text{Card}(S_1) = \begin{pmatrix} \frac{n(n-1)}{2} \\ r \end{pmatrix} = \frac{1}{r!}\frac{n(n-1)}{2}\left(\frac{n(n-1)}{2} - 1\right)\cdots\left(\frac{n(n-1)}{2} - r + 1\right)$$

and

$$\text{Card}(S_2) \leq \sum_{j=1}^{2r-1} \begin{pmatrix} 2r-1 \\ j \end{pmatrix} \pi_n^j n^{2r-j} = o(n^{2r}),$$

where in this summation j represents the number of increments smaller than π_n as we sort the $2r$ vertices i_s's and j_s's in increasing order. Hence,

$$\text{Card}(S) = \frac{n^{2r}}{r!2^r}(1 + o(1)).$$

The result follows by noting that (OA.8) can be rewritten as

$$\mathbb{E}\left(\frac{U_n}{\sqrt{2}}\right)^{2r} = 2^{2r}\frac{1}{(n\sqrt{2k})^{2r}}\frac{n^{2r}}{r!2^r}(1+o(1))\frac{(2r)!}{2^r}k^r = \frac{(2r)!}{2^r r!} + o(1). \quad \square$$