Quantitative Economics 15 (2024), 849–891 1759-7331/20240849

Specification testing for conditional moment restrictions under local identification failure

Prosper Dovonon
Department of Economics, Concordia University, CIREQ, and CIRANO

Nikolay Gospodinov
Research Department, Federal Reserve Bank of Atlanta

In this paper, we study the asymptotic behavior of specification tests in conditional moment restriction models under first-order local identification failure with dependent data. More specifically, we obtain conditions under which the conventional specification test for conditional moment restrictions retains its standard normal limit when first-order local identification fails but global identification is still attainable. In the process, we derive some novel intermediate results that include extending the first- and second-order local identification framework to models defined by conditional moment restrictions, establishing the rate of convergence of the GMM estimator and characterizing the asymptotic representation for degenerate U-statistics under strong mixing dependence. Importantly, the specification test is robust to first-order local identification failure regardless of the number of directions in which the Jacobian of the conditional moment restrictions is degenerate and remains valid even if the model is first-order identified.

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, robustness.

JEL classification. C01, C1, C14, G12.

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

Prosper Dovonon: prosper.dovonon@concordia.ca
Nikolay Gospodinov: nikolay.gospodinov@atl.frb.org
We thank four referees and seminar and conference participants at LSE, Princeton University, UCL, the 2022 meeting of the Société Canadienne de Science économique (Montreal, Canada), and the 2023 Africa Meeting of the Econometric Society (Nairobi, Kenya) for insightful comments and suggestions. The first author acknowledges financial support from the Social Sciences and Humanities Research Council of Canada. 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.

Available at http://qeconomics.org. https://doi.org/10.3982/QE2242

the 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 standard 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, e.g., 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 underidentified*. 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, Kan, and Robotti (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.

This paper builds on these two strands of literature to obtain conditions under which the conventional specification tests of conditional moment conditions remain valid under first-order local identification failure. It should be noted that this is not the case in unconditional moment restriction models, where the second-order local identification leads to overrejection of the standard specification test (Dovonon and Renault (2013)). By contrast, our proposed test is characterized by robust properties as it preserves its standard asymptotic limit irrespective of whether the model is first-order or second-order identified. More specifically, the proposed conditional specification test retains its validity in the presence of possible uncertainty about the parameter values or moment conditions that determine first-order local identification. These robustness properties constitute the main advantage of our approach and warrant some additional remarks. Importantly, the proposed test is agnostic to the precise form of the local identification failure, that is, the degree of rank deficiency of the expected Jacobian matrix of the moment conditions, the values of the model parameters that give rise to this rank deficiency or if the Jacobian is exactly zero. In the latter case, and if prior knowledge of the zero Jacobian is available to the researcher, Lee and Liao (2018) propose to augment the set of moment conditions with this additional restriction that restores the first-order local identification and standard inference. However, if the source of the rank deficiency is different or first-order local identification holds, imposing these Jacobian restrictions would be invalid and lead to erroneous inference. By contrast, the implementation and validity of our proposed test obviates the need to take a stand on the form of the rank deficiency of the Jacobian matrix or, more generally, whether the model is first-order locally identified or not. When the model is indeed first-order locally identified, the conditional specification test continues to be characterized by the standard normal limit.

To this end, we formalize the concepts of point identification and first-order local identification failure in conditional moment restriction models. Similar 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, Imbens, and Newey (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. While the test proposed by de Jong and Bierens (1994) is obtained in the context of nonlinear regression models, the test by Donald, Imbens, and Newey (2003) is also developed within the GMM framework for a specific choice of basis functions and cross-sectional data. Our results therefore show that the GMM-based test of Donald, Imbens, and Newey (2003) retains its size control with dependent data and under first-order local identification failure so long as second-order local identification holds. 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 nontrivial. 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 by conditional moment restrictions. The limiting behavior of the GMM estimator and the specification test are studied in the setup where point identification holds, 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, estimates the directions of the parameters that are locally first-order identified at the 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 the conditional setting, 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 nonstandard limiting behavior of the estimator. Despite this nonstandard 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 conditional moment restriction setup and the testing framework. It also presents the notions of first- and second-order local identification along with alternative characterizations in the context of conditional moment restrictions. In order to enhance the intuition behind the identification framework, Section 3 discusses the model with common conditionally heteroskedastic features, which is later explored further in simulations and in the empirical application. This example also allows us to highlight the robustness of our testing approach to knowledge about the precise structure of the model. The asymptotic properties of the GMM estimator are analyzed in Section 4. Section 5 proposes a CLT for U-statistics under strong mixing dependence and establishes the asymptotic normality of the specification test statistic under the null hypothesis. In addition, this section shows that this test is consistent against all alternatives. Section 6 reports simulation results for the proposed specification test and provides an empirical application for the presence of common conditionally heteroskedastic features in portfolio bond returns. Section 7 concludes. Proofs and additional results are provided in Appendices A and B, and the Supplemental Material (Dovonon and Gospodinov (2024)).

Throughout the paper, we use the following notation. For any matrix C, $\|C\|_2 = \sqrt{\lambda_{\max}(CC')}$ denotes the spectral norm, where $\lambda_{\max}(\cdot)$ is 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, $\mathrm{Card}(S)$ denotes the cardinality of a finite set S, defined to be the number of elements in the set S, $\mathrm{vec}(C)$ signifies column vectorization of a matrix C, $a \vee b$ denotes the maximum of a and b, $\mathrm{Rank}(C)$ is the rank of a matrix C, and $\mathrm{Diag}(c_{11}, c_{22}, \dots, c_{mm})$ denotes an $m \times m$ diagonal matrix with $(c_{11}, c_{22}, \dots, c_{mm})'$ 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

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

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 and identification

2.1 *Conditional moment restriction setup*

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

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

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 restriction in (1) is

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

against the alternative

$$H_1 : \Pr\big\{\mathbb{E}\big(u(y_t, \theta)|x_t\big) = 0\big\} < 1, \quad \text{for any } \theta \in \Theta. \tag{3}$$

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 $\theta \in \Theta$,

$$\rho(x_t, \theta) := \mathbb{E}\big(u(y_t, \theta)|x_t\big) = 0, \quad \text{a.s.} \quad \Leftrightarrow \quad \theta = \theta_0. \tag{4}$$

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 estimator of θ_0. The standard approach to inference relies on a local identification condition, which states that

$$\mathbb{E}\big(\rho_\theta(x, \theta_0)'\rho_\theta(x, \theta_0)\big) \quad \text{is nonsingular}, \tag{5}$$

where $\rho_\theta(x, \theta) := \mathbb{E}(\nabla_\theta u(y, \theta)|x)$ with $\nabla_\theta u(y, \theta_0) = \partial u(y_t, \theta)/\partial\theta'|_{\theta=\theta_0}$. 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.

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

The rest of our main analytical and testing framework can be summarized as follows. 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 $\mathbb{R}^{k_x}$. We equip $L^2(P)$ with the inner product: $\langle f(\cdot), h(\cdot)\rangle = \mathbb{E}(f(x_t)g(x_t))$. Let $(g_l)_{l\in\mathbb{N}}$ be a countable basis (not necessarily orthonormal) of $L^2(P)$ and $g^{(k)} := (g_1, \dots, g_k)'$.[2] We investigate the null hypothesis in (2) by proposing a test for the sequence of unconditional moment restrictions

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

or written more compactly as

$$\mathbb{E}\big(g^{(k)}(x_t)u(y_t, \theta_0)\big) = 0, \tag{6}$$

where $k = k(n)$ with $k(n) \to \infty$ as $n \to \infty$. The GMM estimator, based on these restrictions, is defined as

$$\hat{\theta} = \arg\min_{\theta\in\Theta} \bar{f}_k(\theta)'\hat{W}_k\bar{f}_k(\theta), \tag{7}$$

where

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

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 two-step GMM estimator, which uses the weighting matrix

$$\hat{W}_k = \hat{V}_k^{-1}, \qquad \hat{V}_k = \frac{1}{n}\sum_{t=1}^{n} f_k(x_t, y_t, \tilde{\theta})f_k(x_t, y_t, \tilde{\theta})'. \tag{8}$$

The preliminary (first-step) GMM estimator $\tilde{\theta}$ used in (8) is obtained by commonly setting $\hat{W}_k = I_k$ or, more generally, to a nonrandom matrix sequence $W_{k,0}$. Furthermore, we will derive our results under the condition that the sequence $(f_k(x_t, y_t, \theta_0))_t$ is serially uncorrelated. This is ensured by our maintained assumption that

$$\mathbb{E}\big(u(y_t, \theta_0)|\mathcal{F}_t\big) = 0, \quad \text{where } \mathcal{F}_t = \sigma\big(x_t, u(y_{t-1}, \theta_0), x_{t-1}, u(y_{t-2}, \theta_0), \dots\big). \tag{9}$$

[2] 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. Furthermore, as discussed by de Jong and Bierens (1994), the conditioning random variable x can be considered to be bounded since

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

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 *componentwise* `arc` tangent function, that is, $x \mapsto \arctan(x) = (\arctan(x_1), \dots, \arctan(x_{k_x}))'$, while choices of enumeration of weight functions $(g_l)_l$ include polynomial, trigonometric, and flexible Fourier form families (de Jong and Bierens (1994); see also Andrews (1991), and Gallant (1981), for more details on these families).

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 weighting matrix for the two-step GMM estimator is thus constructed as 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 (8).

Our goal is to derive the asymptotic distribution of the test statistic of the null hypothesis in (2)

$$\hat{Z} = \frac{1}{\sqrt{2k}}\big(\bar{f}_k(\hat{\theta})'\hat{V}_k^{-1}\bar{f}_k(\hat{\theta}) - k\big) \tag{10}$$

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. We show that even in this highly nonstandard identification setup, the test $\hat{Z}$ retains its $N(0, 1)$ limit under H_0 in (2)—which is the asymptotic distribution of the test in the standard identification setting (de Jong and Bierens (1994))—and is consistent under H_1 in (3).

2.2 *Identification*

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

$$\mathbb{E}\big(z_t \cdot u(y_t, \theta_0)\big) = 0. \tag{11}$$

Following Sargan (1983) and Dovonon and Renault (2013), among others, the unconditional moment restriction (11) locally identifies θ_0 at first order if

$$\text{Rank}\big(\mathbb{E}\big(z_t \cdot \nabla_\theta u(y_t, \theta_0)\big)\big) = p, \tag{12}$$

whereas lack of first-order local identification occurs when

$$\text{Rank}\big(\mathbb{E}\big(z_t \cdot \nabla_\theta u(y_t, \theta_0)\big)\big) < p. \tag{13}$$

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 (12) holds. Relatedly, first-order local identification fails if and only if (13) 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*, $\text{Rank}(\mathbb{E}(z_t \cdot \nabla_\theta u(y_t, \theta_0))) < p$, *where $z_t = g(x_t)$ and assuming that the moment exists.*

(ii) *There exists at least one linear combination of the elements of* $\mathbb{E}(\nabla_\theta u(y_t, \theta_0)|x_t)$ *that is almost surely nil.*[3]

[3]Note that, when u is a q-vector, "$\cdot$" in part (i) should be replaced by the Kronecker product "$\otimes$" and "*elements*" in part (ii) should be replaced by "*columns*," with the understanding that $\nabla_\theta u(y_t, \theta_0)$ is a (q, p)-matrix.

A proof of an alternative formulation of this proposition (Proposition A.1(ii)) is provided in Appendix B. The characterization in (ii) validates (5) as the first-order local identification condition in model (1). Furthermore, this highlights some similarities with the first-order local identification failure in parametric models as studied by Lee and Chesher (1986) and Rotnitzky, Cox, Bottai, and Robins (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.

With the basis functions $g^{(k)}(x)$ as defined in Section 2.1, Proposition A.1 in Appendix A establishes the connection between point identification (resp., first-order local identification failure) in conditional moment models such as (1) with point identification (resp., first-order local identification failure) in their corresponding sequences of unconditional moment models given by (6). While point identification by (1) amounts to point identification by (6) for some k_0, the fact that $g^{(k)}(x)$ is an increasingly embedded sequence of vectors allows us to claim that (1) fails first-order local identification if and only if there exists r such that, for any k large enough (say $k \geq k_0$), $\text{Rank}(G^{(k)}) = r < p$, where $G^{(k)}$ is the expected Jacobian matrix:

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

By construction, the null space of $G^{(k)}$ and the range of its transpose are fixed for any k large enough. 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 only 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 $a \in \mathbb{R}^r$ and $b \in \mathbb{R}^{p-r}$, we have[4]

$$\Big(G^{(k)} R_1 a + \big(b' R_2' \mathbb{E}\big(g_l^{(k)}(x) \nabla_{\theta\theta} u(y, \theta_0)\big) R_2 b\big)_{1 \leq l \leq k}\Big) = 0 \quad \Leftrightarrow \quad \big((a, b) = (0, 0)\big), \tag{14}$$

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

$$\begin{aligned} &\big\| M^{(k)} \big(b' R_2' \mathbb{E}\big(g_l^{(k)}(x) \nabla_{\theta\theta} u(y, \theta_0)\big) R_2 b\big)_{1 \leq l \leq k} \big\|_2 \geq \sqrt{\gamma_k} \|b\|_2^2, \\ &\text{with } \gamma_k = \inf_{\|b\|_2 = 1} \big\| M^{(k)} \big(b' R_2' \mathbb{E}\big(g_l^{(k)}(x) \nabla_{\theta\theta} u(y, \theta_0)\big) R_2 b\big)_{1 \leq l \leq k} \big\|_2^2. \end{aligned} \tag{15}$$

[4]From our discussion above, if (14) holds for a given k, it holds for all $k' \geq k$.

This inequality is instrumental in deriving the rate of convergence of estimators of θ_0. We show in Appendix B 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. We maintain (see Assumption 4(ii)) that $\gamma_k = O(k)$.

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

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

By definition, it follows that

$$\big(b'R_2'\mathbb{E}\big(g_l^{(k)}(x)\nabla_{\theta\theta}u(y, \theta_0)\big)R_2b\big)_{1\le l\le k} = H^{(k)}(\theta)\cdot\text{vec}\big(R_2bb'R_2\big).$$

First, note that 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. Second, 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

$$\text{Rank}\big[G^{(k)}R_1 \quad H^{(k)}(\theta_0)\cdot\text{vec}\big(R_2R_2'\big)\big] = p,$$

for R_1 and R_2 defined as in (14). More generally, suppose that $\text{Rank}(G^{(k)}) = p - 1$ for all $k \ge k_0$, and in some parameter direction, say θ_h, we have $\mathbb{E}(g^{(k)}(x)[\partial u(y, \theta_0)/\partial\theta_h]) = 0$. Then second-order local identification amounts to

$$\text{Rank}\big[G^{(k)}d_h \quad H^{(k)}(\theta_0)\cdot\text{vec}\big(e_he_h'\big)\big] = p,$$

where $G^{(k)}d_h$ is the $(k, p-1)$-matrix of columns of $G^{(k)}$, except the hth column, and e_h is the p-vector of zeros with 1 in its hth entry so that $H^{(k)}(\theta_0)\cdot\text{vec}(e_he_h') = \mathbb{E}(g^{(k)}(x)[\partial^2u(y, \theta_0)/\partial\theta_h^2])$.

3. Example: Common conditionally heteroskedastic features

In this section, we use a model with common conditionally heteroskedastic features to illustrate the identification structure of interest in which the conditional moment restriction fails first-order local identification but is point identified and satisfies the second-order local identification condition. This example, which captures some important common drivers in financial asset returns (Engle and Kozicki (1993); see also Dovonon and Renault (2013)), is further explored in our simulations and empirical application.

Let $\mathcal{F}_t$ denote an increasing filtration to which all information available at date t is adapted and $f_{t+1} := (f_{1,t+1}, f_{2,t+1})'$ be a vector of unobserved conditionally heteroskedastic (CH) factors with conditional mean and variance given by

$$\mathbb{E}(f_{t+1}|\mathcal{F}_t) = 0, \qquad D_t := \text{Var}(f_{t+1}|\mathcal{F}_t) = \text{Diag}\big(\sigma_{1,t}^2, \sigma_{2,t}^2\big),$$

where $\sigma_{i,t}^2$ $(i = 1, 2)$ are time-varying volatility processes. Furthermore, let Y_t be a bivariate process of asset (stock, bond, or other financial asset) excess returns that admits the following CH factor representation (Engle, Ng, and Rothschild (1990); King, Sentana, and Wadhwani (1994); Sentana and Fiorentini (2001)):

$$Y_{t+1} = \Lambda D_t \tau + \Lambda f_{t+1} + e_{t+1}, \tag{16}$$

where Λ is a (2×2) matrix of factor loadings, $\tau \in \mathbb{R}^2$ represents the vector of market prices of risk associated with each of the volatility factors, and e_{t+1} is the vector of idiosyncratic shocks that is assumed to satisfy the following restrictions:

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

with $\Omega = \left(\begin{smallmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{smallmatrix}\right)$ being time invariant. In this model, $\Lambda D_t \tau$ is the vector of risk premia and, by definition, is also the conditional mean of Y_{t+1}.

Each process is characterized by CH dynamics so that each return 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 the two columns of the matrix of factor loadings Λ and is tested within (16) 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:

$$\mathbb{E}\big(u_{1,t+1}(\theta)|\mathcal{F}_t\big) = 0, \quad \text{with } u_{1,t+1}(\theta) := (Y_{1,t+1} + \beta Y_{2,t+1})^2 - c, \tag{17}$$

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 θ that offsets the CH factor is $\beta_0 = -\Lambda_{1,1}/\Lambda_{2,1}$ 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, with $\omega = 2(\Omega_{21} + \beta_0 \Omega_{22})$, simple calculations yield

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

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

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

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

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

Hence, for $a, b \in \mathbb{R}$, we have

$$G^{(2)} R_1 a + b^2 \begin{pmatrix} R_2' \mathbb{E}\big(\nabla_{\theta\theta} u_{1,t+1}(\theta_0)\big) R_2 \\ R_2' \mathbb{E}\big(z_t \cdot \nabla_{\theta\theta} u_{1,t+1}(\theta_0)\big) R_2 \end{pmatrix} = 0 \quad \Leftrightarrow$$

$$\begin{pmatrix} 1 \\ \mathbb{E}(z_t) \end{pmatrix} (\omega^2 + 1) a + \begin{pmatrix} \mathbb{E}\big(Y_{2,t+1}^2\big) \\ \mathbb{E}\big(z_t Y_{2,t+1}^2\big) \end{pmatrix} b^2 = 0,$$

which yields $a = b = 0$.

Instead of constructing the test only based on (17), one may exploit information about the conditional mean by adding the following moment restriction:

$$\mathbb{E}\big(u_{2,t+1}(\beta) | \mathcal{F}_t\big) = 0, \quad \text{with } u_{2,t+1}(\beta) := Y_{1,t+1} + \beta Y_{2,t+1}. \tag{18}$$

The joint model (17)–(18) is obviously point identified. Furthermore, the possibility for this extended model to be first-order locally identified by the restriction (18) depends on the value of the market price of risk τ. In fact, the extended model is first-order locally identified if and only if $\tau \neq 0$. Indeed, with $u_{t+1}(\theta) = (u_{1,t+1}(\beta, c), u_{2,t+1}(\beta))'$ and $\Lambda_{2\bullet}$ denoting the second row of Λ,

$$\mathbb{E}\big(\nabla_\theta u_{t+1}(\theta_0) | \mathcal{F}_t\big) = \begin{pmatrix} \omega & -1 \\ \Lambda_{2\bullet} D_t \tau & 0 \end{pmatrix}, \tag{19}$$

which is nondegenerate if and only if $\tau \neq 0$. Note that the second row in (19) is effectively $\mathbb{E}(Y_{2,t+1} | \mathcal{F}_t)$, which characterizes the predictability of $Y_{2,t+1}$.

Nevertheless, when $\tau = 0$, the augmented model (17)–(18) continues to be second-order locally identified by the conditional moment restriction (17) while the additional restriction (18) is completely uninformative about θ_0 due to the unpredictability of $Y_{2,t+1}$. When the researcher is agnostic about the true value of τ, the robustness properties of the test proposed in this paper prove to be particularly appealing. More specifically, irrespective of the value of τ, this test maintains the same asymptotic distribution under the null regardless of whether the simpler moment restriction (17) or the restrictions implied by the extended model (17)–(18) are used.

4. 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.

4.1 *Assumptions*

This section 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* (9), *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 $f_k(x_t, y_t, \theta_0)$ 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* $k \geq k_0$,

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

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

$$\operatorname{Rank}\big(G^{(k)}\big) = r < p.$$

(iii) (Second-order local identification) *For any* $k \geq k_0$ *and for* R_1 *and* R_2 *defined as in* (14), *we have*

$$\begin{aligned} &\Big(G^{(k)} R_1 a + \big(b' R_2' \mathbb{E}\big(g_l^{(k)}(x)(\nabla_{\theta\theta} u)(\theta_0)\big) R_2 b\big)_{1\leq l\leq k} = 0\Big) \\ &\quad\Rightarrow \quad \big((a, b) = (0, 0)\big) \end{aligned} \tag{20}$$

for all $a \in \mathbb{R}^r$, *all* $b \in \mathbb{R}^{p-r}$, *and all* $k \geq k_0$.

Assumption 2 characterizes the identification framework for our analysis. Assumption 2(i) is necessary for establishing the consistency of GMM estimators. As we argue in Proposition A.1 in Appendix A, this condition is equivalent, under mild regularity conditions, to point identification of the conditional moment restriction model. More specifically, it is worth mentioning that this condition holds under the mild assumptions of Proposition A.1(i) if we assume second-order local identification—which rules out solutions of $\mathbb{E}[g^{(k)}(x_t) u(y_t, \theta)] = 0$ lying in $\Theta \setminus \{\theta_0\}$ and converging to θ_0 as k grows—and point identification of the conditional moment restriction 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. Assumption 2(iii) claims second-order local identification.

Theorem A.2 in Appendix A shows that the GMM estimator, defined by (7), is consistent under Assumptions 1, 2(i) and the following Assumption 3.

ASSUMPTION 3 (Consistency of GMM estimators). *Assume that*:

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

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

(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*

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

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) serves as a sufficient condition and is not restrictive as it merely rules out the possibility that W_k is ill-conditioned. Note that similar conditions on the boundedness of the eigenvalues are routinely assumed in models with an expanding number of instruments or covariates (e.g., Han and Phillips (2006); Cattaneo, Jansson, and Newey (2018)), and for estimation of large covariance matrices (e.g., Fan, Liao, and Mincheva (2013)). In standard problems where k is fixed, $\hat{W}_k$ is assumed to converge in probability to W_k. With an 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, as stated by the second and third conditions of Assumption 3(ii), is sufficient to establish consistency of GMM. The conditions in Assumption 3(ii) are trivially fulfilled by estimation procedures using a nonrandom weighting matrix with eigenvalues bounded away from zero and from above. This is the case for the first-step GMM estimator introduced above, which uses $W_{k,0}$ as a weighting matrix and, therefore, is consistent. For the optimal weighting matrix, Assumption 3(ii)—the boundedness of the smallest eigenvalue from zero, in particular—may appear restrictive. While the provision of more primitive and less restrictive regularity conditions on the optimal weighting matrix may still be possible,[5] a more formal treatment is beyond the scope of this paper.

In Appendix A, we show that these conditions continue to hold for the two-step GMM estimator and this estimator is consistent as well (see Assumption A.1 and Corollary A.3 in Appendix A). To summarize, Assumptions 1, 2(i), and 3 already ensure the consistency of the GMM estimator in point-identified, conditional restriction model under first-order local identification failure (see Theorem A.2 in Appendix A). However, for

[5]In the Supplemental Material, we provide more explicit conditions—on the conditional expectation of $u(y, \theta_0)^2$ and the density of x_t—for the boundedness of the eigenvalues in our context. We show by simulations that these conditions still appear to be only sufficient conditions and can be further relaxed.

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 $\bar{H}^{(k)}(\theta) = n^{-1}\sum_{t=1}^{n} g^{(k)}(x_t)[\text{vec}'(\nabla_{\theta\theta}u(y_t,\theta))]$ and $\bar{G}^{(k)}(\theta) = n^{-1}\sum_{t=1}^{n} g^{(k)}(x_t)\nabla_\theta u(y_t,\theta)$ 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 $U_t(\theta)$ (with m fixed) that proves useful in obtaining the order of magnitude of quantities such as $(1/n)\sum_{t=1}^{n} g^{(k)}(x_t)U_t(\bar{\theta}) - \mathbb{E}(g^{(k)}(x_t)U_t(\theta_0))$, 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*

$$\sup_{\theta\in\mathcal{N}}\left\|\frac{1}{n}\sum_{t=1}^{n} g^{(k)}(x_t)U_t(\theta)' - \mathbb{E}\big(g^{(k)}(x_t)U_t(\theta)'\big)\right\|_2 = O_P\left(\sqrt{\frac{k}{n}}\right),$$

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

$$\forall\theta_1,\theta_2\in\mathcal{N},\quad \big\|\mathbb{E}\big(g_l(x_t)U_{t,r}(\theta_1)\big) - \mathbb{E}\big(g_l(x_t)U_{t,r}(\theta_2)\big)\big\|_2 \le c\|\theta_1-\theta_2\|_2.$$

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 $\sup_{\theta\in\mathcal{N}}\mathbb{E}(|g_l(x)|\|\partial U_{t,r}(\theta)/\partial\theta'\|_2) \le \Delta < \infty$. We now present the final assumption that is useful 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, $\theta \mapsto u(y,\theta)$ is twice continuously differentiable in a neighborhood of θ_0 for each y, and the maps $\theta \mapsto \nabla_\theta u(y,\theta)$ and $\theta \mapsto \text{vec}(\nabla_{\theta\theta}u(y,\theta))$ satisfy Condition C.*

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

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

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

Note that D_1 and $H^{(k)}$ are nonzero matrices and the condition on their spectral norms in Assumption 4(ii) follows if each has at least one column with a number of nonzero 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 prevents 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, for example, 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

$$\begin{aligned} c' \operatorname{Var}(S_k)c &= \gamma_k^{-1}c'H^{(k)\prime}W_k^{1/2}M^{(k)}W_k^{1/2}V_kW_k^{1/2}M^{(k)}W_k^{1/2}H^{(k)}c \\ &\leq \Delta\gamma_k^{-1}c'H^{(k)\prime}W_k^{1/2}M^{(k)}W_k^{1/2}H^{(k)}c \leq \Delta\lambda_{\max}(W_k)\gamma_k^{-1}\|H^{(k)}\|_2^2 = O(1), \end{aligned}$$

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 close to W_k as n grows. Note that these two conditions are fulfilled by the GMM estimator with nonrandom matrix having bounded eigenvalues such as $W_{k,0}$. These conditions are also satisfied for the two-step GMM estimator as we show in Appendix B in the context of the rate-of-convergence results in the next subsection.

4.2 *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 $\operatorname{Rank}(G^{(k)}) = r < p$.

Theorem 4.1. *If Assumptions 1–4 hold and $k \to \infty$ as $n \to \infty$ with $k^3/n \to 0$, then*

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

Theorem 4.1 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 vector 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 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 (in a location model) even with possibly irrelevant instruments. In our framework, point identification is maintained and consistent estimation is therefore possible even if the number of moment restrictions does not grow. But, as Theorem 4.1 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.

Although the rates of convergence that are stated in Theorem 4.1 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 4.2. *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 4.2 first demonstrates that the slow rate of convergence derived in Theorem 4.1 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 4.2 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 nonstandard 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.

5. 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 a 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:

$$U_n = \frac{1}{n}\sum_{t\neq s} \frac{f_k(x_t)' V_k^{-1} f_k(x_s)}{\sqrt{k}}, \tag{21}$$

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. While the asymptotic theory for degenerate U-statistics has been extensively studied in the literature (see the Supplemental Material), the available results are 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 (21). Since this result may be of independent interest, we collect the sufficient conditions 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*

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

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

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

$$\mathbb{E}\Big(\max_{1\leq t\leq n} \big\|V_k^{-1/2} f_k(x_t)\big\|/\sqrt{k}\Big) = O\big(\log^\beta n\big) \quad \textit{and}$$

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

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 S.2 in the Supplemental Material and Theorem 5.1 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 fourth-order moments of $f_k(x_t)$. Replacing these moments by their analogues under independence is a common approach in the literature. The remainder is then controlled by resorting to Lemma S.1 in the Supplemental Material, 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 particularly 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.

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.[6] 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 (21).

THEOREM 5.1. *Under Assumptions*-CLT 1, 2, *and* 3,

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

The proof of Theorem 5.1 follows similar arguments as in Kim, Luo, and Kim (2011) and is provided in the Supplemental Material. 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 $\hat{Z} = \frac{1}{\sqrt{2k}}(\bar{f}_k(x_t, y_t, \hat{\theta})'\hat{V}_k^{-1}\bar{f}_k(x_t, y_t, \hat{\theta}) - k)$ under the null hypothesis that the conditional moment restriction (1) is correctly specified.

THEOREM 5.2. *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 5.2. First, it is important to underscore that the standard normal limit distribution in Theorem 5.2 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 nonstandard 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

[6]For instance, $\beta = 1$ if z_t has a Gamma distribution, and $\beta = 1/2$ if z_t is Gaussian. It is worth noting 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 (Pereira (1983); see also Berman (1964) and Isaev, Rodionov, Zhang, and Zhukovskii (2020) for similar results for time-dependent processes).

limit in Theorem 5.2. While the $\hat{Z}$ test statistic is based on the two-step GMM estimator with $\hat{W}_k = [\frac{1}{n}\sum_{t=1}^n f_k(x_t, y_t, \tilde{\theta}) f_k(x_t, y_t, \tilde{\theta})']^{-1}$, 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 1 in Appendix B.) 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 two-step 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, e.g., Carrasco and Florens (2000); Donald, Imbens, and Newey (2003); Tripathi and Kitamura (2003); among others). Theorem 5.2 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. In our simulations and empirical application, we set $k \propto n^{1/6}$. One may even choose k to grow arbitrarily slowly with n and the results in this paper would continue to hold. However, a k that grows too slowly may compromise power. It is worth stressing that the construction and implementation of the test is agnostic about the precise form of first-order local identification failure. This robustness property is further enhanced by the fact that the test remains valid even if the model happens to be first-order identified.

We complete our theoretical analysis by characterizing the limiting behavior of $\hat{Z}$ under the alternative hypothesis H_1, specified in (3). 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 Sargan–Hansen specification test for this unconditional restriction—albeit infeasible because k_0 is unknown—would be consistent. Theorem 5.3 shows that this result carries over to the feasible statistic $\hat{Z}$, which makes our specification test consistent against all alternatives.[7]

Theorem 5.3. *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 B.4 in Appendix B, and the conditions of that lemma are satisfied with* $u(\theta) := u(y, \theta)$. *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* $\sup_{\theta\in\Theta} |(1/n)\sum_{t=1}^n g_l(x_t)u(y_t, \theta) -$

[7] 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.

$\mathbb{E}(g_l(x_t)u(y_t, \theta))| = o_P(1)$ *for each* l. *Then, under* H_1,

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

The conditions of this theorem are essentially a subset of those of the main Theorem 5.2. 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 primitive conditions. Theorem 5.3 shows that $|\hat{Z}|$ diverges to infinity if k is such that $k^{3/2} = o(n)$. Note that in Theorem 5.2, which 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.

6. Simulations and empirical analysis

In this section, we provide simulation evidence on the empirical size and power 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.

6.1 *Simulations*

The simulation design for assessing the finite-sample properties of the specification test $\hat{Z}$ is tailored to the common CH factor example discussed in Section 3 and used in the subsequent empirical application. More specifically, the data generating process has the form:

$$Y_{t+1} = \Lambda D_t \tau + \Lambda f_{t+1} + e_{t+1}, \tag{22}$$

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 $m \times 1$ vector of unobserved CH factors. The ith component $f_{i,t+1}$ of f_{t+1} follows a GARCH(1,1) process:

$$f_{i,t+1} = \sigma_{i,t}\varepsilon_{i,t+1}, \qquad \sigma^2_{i,t} = \omega_{i,0} + \omega_{i,1} f^2_{i,t} + \omega_{i,2}\sigma^2_{i,t-1}, \tag{23}$$

with $\omega_{i,0}, \omega_{i,1}, \omega_{i,2} > 0$ and $\varepsilon_{i,t+1} \sim iidN(0, 1)$.[8] Finally, Λ is an $m \times m$ matrix of factor loadings, $D_t := \mathrm{Diag}(\sigma^2_{1,t}, \dots, \sigma^2_{m,t})$ and τ is an $m \times 1$ vector of market prices of risk (see, e.g., King, Sentana, and Wadhwani (1994)). In all cases considered below, we set $\kappa = 0.1$ and $\omega_{i,0} = 1 - \omega_{i,1} - \omega_{i,2}$ for $i = 1, 2, 3$, where $(\omega_{1,1}, \omega_{1,2}) = (0.2, 0.6)$, $(\omega_{2,1}, \omega_{2,2}) = (0.4, 0.4)$ and $(\omega_{3,1}, \omega_{3,2}) = (0.1, 0.8)$.

The parameterization of the factor loading matrix determines if the model is under the null (common CH features) or under the alternative. In evaluating the size properties of the specification tests, 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. For case (i), $\Lambda = \begin{pmatrix} 1 & 0 \\ 0.5 & 0 \end{pmatrix}$, and for case (ii), $\Lambda = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0.5 & 0.5 & 0 \end{pmatrix}$. In assessing the power properties of the tests, we set Λ to be the identity matrix.

[8] Bougerol and Picard (1992) derive the conditions for strict stationarity of GARCH processes.

As demonstrated in Section 3, the presence of common CH features amounts to testing the conditional moment restriction $\mathbb{E}(u_{1,t+1}(\theta)|\mathcal{F}_t)) = 0$. In case (i), $u_{1,t+1}(\theta)$ is parameterized as $u_{1,t+1}(\theta) := (\beta_1 Y_{1,t+1} + (1-\beta_1)Y_{2,t+1})^2 - c$ with $\theta = (\beta_1, c)'$ and for case (ii), $u_{1,t+1}(\theta) := (\beta_1 Y_{1,t+1} + \beta_2 Y_{2,t+1} + (1-\beta_1-\beta_2)Y_{3,t+1})^2 - c$ with $\theta = (\beta_1, \beta_2, c)'$.[9] For some choice of instruments $z_t \in \mathcal{F}_t$, the parameter vector θ is estimated by the two-step GMM based on the moment conditions $\mathbb{E}[z_t u_{1,t+1}(\theta)] = 0$. For the unconditional GMM approach and the corresponding J (Sargan–Hansen) test, we use $z_t = (Y_t', Y_t^{2\prime})'$, with $Y_t^2 := (Y_{1,t}^2, \dots, Y_{m,t}^2)'$, which gives rise to $k_1 - p$ overidentifying restrictions: with $k_1 = 4$ and $p = 2$ for case (i), and $k_1 = 6$ and $p = 3$ for case (ii). We report two versions of the J test: one based on critical values from $\chi^2(k_1 - p)$ and one based on critical values from $\chi^2(k_1)$. As argued above, the former test is not valid when the model is not first-order locally identified while the latter remains valid but is conservative.

The $\hat{Z}$ test uses $x := Y_t$ as conditioning variables in constructing the functions $g_l(\cdot)$, $l = 1, \dots, k$, so that the GMM estimation is based on $\mathbb{E}(z_t u_{1,t+1}(\theta)) = \mathbb{E}(g_l(Y_t) \times u_{1,t+1}(\theta)) = 0$. In the case where $k_x := \text{size}(x) = 1$, the tuning parameters for the test 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 $l = 1, \dots k$, and $k = n^{1/6}$.[10] In the case where $k_x > 1$, $\Psi(\cdot)$ and $g_l(\cdot)$ are applied componentwise to x leading to $k = k_x \cdot n^{1/6}$ moment restrictions.[11] We report results for the one-sided test $\hat{Z}$ at nominal level α, $\hat{Z} > q_{1-\alpha}$, where $q_{1-\alpha}$ denotes the $(1-\alpha)$ quantile of the $N(0, 1)$ distribution.

The fourth specification that we consider is also a J test but it is based on the augmented set of moment conditions $\mathbb{E}(z_t \binom{u_{1,t+1}(\theta)}{u_{2,t+1}(\beta)}) = 0$, where $u_{2,t+1}(\beta) := \beta_1 Y_{1,t+1} + (1-\beta_1)Y_{2,t+1}$ for case (i), and $u_{2,t+1}(\beta) := \beta_1 Y_{1,t+1} + \beta_2 Y_{2,t+1} + (1-\beta_1-\beta_2)Y_{3,t+1}$ for case (ii). The value of the parameter τ in model (23) determines if the additional restriction $\mathbb{E}(z_t u_{2,t+1}(\beta)) = 0$ restores the first-order local identification ($\tau \neq 0$) or not ($\tau = 0$). For case (i), we set $\tau = (0, 0)'$ or $\tau = (0.1, 0.1)'$, and for case (ii), $\tau = (0, 0, 0)'$ or $\tau = (0.1, 0.1, 0.1)'$. The J statistic in this augmented model is compared to critical values from $\chi^2(k_2 - p)$, where $k_2 = 8$ and $p = 2$ for case (i) and $k_2 = 12$ and $p = 3$ for case (ii).

The empirical rejection probabilities of the four specification tests are based on $n = (2000, 5000, 10{,}000)$ and 10,000 Monte Carlo replications. Tables 1 and 2 present the results for case (i), $m = 2$, and case (ii), $m = 3$, respectively, with the top panel in

[9] We consider in this section a linear combination of assets with coefficients adding to one to mimic portfolio formation. This yields the same identifying properties for the resulting moment restrictions as those obtained using the weight $(1, \beta)'$, considered in Section 3, so long as the matrix of factor loadings does not contain a relevant column with equal elements.

[10] Even for large sample sizes, the choice $k = n^{1/6}$ may result in a relatively small number of instruments. It appears that further improvements can be obtained if we set $k = \text{const} \cdot n^{1/6}$, where the constant "const" is calibrated to the particular setup (to the values of n and m). Ideally, it seems desirable to target a more data-driven choice of k via subsampling or resampling methods. But, as argued in the concluding section, such methods may be difficult to 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. Fortunately, our simulation and empirical results suggest that the properties of the test are not particularly sensitive to the values of "const" in $k = \text{const} \cdot n^{1/6}$ for const ≥ 1.

[11] We experimented with the construction of the basis functions as in Bierens (1990) but the results are broadly similar.

Table 1. Empirical rejection rates of specification tests under the null (size) and alternative (power): case $m = 2$.

	Panel A: $\tau = (0, 0)'$								Panel B: $\tau = (0.1, 0.1)'$							
	J Test		J Test		J Augment		$\hat{Z}$ Test		J Test		J Test		J Augment		$\hat{Z}$ Test	
n	$\chi^2(k_1\text{-}p)$		$\chi^2(k_1)$		$\chi^2(k_2\text{-}p)$		$N(0,1)$		$\chi^2(k_1\text{-}p)$		$\chi^2(k_1)$		$\chi^2(k_2\text{-}p)$		$N(0,1)$	
size	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%
2000	8.1	4.0	1.6	0.7	5.3	2.2	6.6	4.0	8.2	4.3	1.7	0.7	5.2	2.3	6.7	4.2
5000	10.7	5.7	2.4	1.1	6.6	3.1	9.0	5.6	11.3	6.0	2.6	1.2	5.4	2.3	8.8	5.7
10,000	13.3	7.7	3.6	1.7	8.9	4.5	9.4	5.9	13.6	7.6	3.7	1.8	6.2	2.7	9.7	6.2
power	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%
2000	10.8	5.5	2.3	0.9	10.7	5.6	99.1	98.4	14.9	8.0	3.3	1.4	89.4	82.6	99.0	98.5
5000	10.8	5.4	2.2	0.9	10.7	5.4	100	100	24.8	14.3	6.9	3.3	99.8	99.5	100	100
10,000	11.2	5.9	2.4	1.1	11.5	5.9	100	100	40.2	27.4	15.8	8.6	100	100	100	100

Note: In this simulation design ($m = 2$), "size" corresponds to a bivariate Y_{t+1} with a single common CH factor, and "power" corresponds to a bivariate Y_{t+1} with two CH factors. The table presents the empirical size and power at 5% and 10% nominal level of three J tests for overidentifying restrictions ($p = 2$, $k_1 = 4$, and $k_2 = 8$) and the $\hat{Z}$ test, proposed in this paper. "J augment" stands for the J test, augmented with an additional conditional moment restriction. The value of τ ($\tau \neq 0$ or $\tau = 0$) determines if the augmented model is first-order locally identified or not. The results are based on 10,000 Monte Carlo replications.

each table reporting the empirical size of the tests and the bottom panel reporting their empirical power.

In the setup where the model is not first-order locally identified but globally identified, the standard J test for overidentifying restrictions is known to overreject under the null (Dovonon and Renault (2013)). These overrejections are confirmed in Table 1

Table 2. Empirical rejection rates of specification tests under the null (size) and alternative (power): case $m = 3$.

	Panel A: $\tau = (0, 0, 0)'$								Panel B: $\tau = (0.1, 0.1, 0.1)'$							
	J Test		J Test		J Augment		$\hat{Z}$ Test		J Test		J Test		J Augment		$\hat{Z}$ Test	
n	$\chi^2(k_1\text{-}p)$		$\chi^2(k_1)$		$\chi^2(k_2\text{-}p)$		$N(0,1)$		$\chi^2(k_1\text{-}p)$		$\chi^2(k_1)$		$\chi^2(k_2\text{-}p)$		$N(0,1)$	
size	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%
2000	4.3	1.8	0.5	0.1	2.4	1.0	3.6	2.0	4.8	2.1	0.7	0.2	3.9	1.6	3.7	2.2
5000	6.5	2.8	0.7	0.3	3.3	1.5	5.2	2.9	7.0	3.3	0.9	0.4	5.3	2.5	5.4	3.2
10,000	8.1	3.7	1.1	0.5	4.3	1.8	6.1	3.9	9.4	4.7	1.3	0.4	6.7	2.9	6.1	3.5
power	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%
2000	2.5	1.2	0.4	0.1	4.7	2.1	50.8	41.9	3.3	1.6	0.5	0.1	55.6	46.2	52.9	43.4
5000	1.8	0.6	0.1	0.0	2.4	1.0	97.3	95.7	3.0	0.9	0.3	0.1	95.2	91.0	97.8	96.4
10,000	5.2	1.9	0.3	0.1	4.6	2.3	100	100	9.7	3.9	0.8	0.3	100	99.9	100	100

Note: In this simulation design ($m = 3$), "size" corresponds to a trivariate Y_{t+1} with two common CH factors, and "power" corresponds to a trivariate Y_{t+1} with three CH factors. The table presents the empirical size and power at 5% and 10% nominal level of three J tests for overidentifying restrictions ($p = 3$, $k_1 = 6$, and $k_2 = 12$) and the $\hat{Z}$ test, proposed in this paper. "J augment" stands for the J test, augmented with an additional conditional moment restriction. The value of τ ($\tau \neq 0$ or $\tau = 0$) determines if the augmented model is first-order locally identified or not. The results are based on 10,000 Monte Carlo replications.

and they continue to get larger as the sample size increases.[12] While the J test based on $\chi^2(k_1)$ critical values is valid, it is very conservative, which results in loss of power. For $\tau = 0$, the standard J test also exhibits lack of power due to the fact that only the instruments Y_t^2 carry information about the model parameters while the restrictions based on the instruments Y_t are uninformative as these restrictions are correct under both the null and the alternative hypotheses and render the model just-identified. When $\tau \neq 0$, the power of the standard J test is somewhat improved (in the case $m = 2$) but it remains low because this test does not exploit explicitly the moment restriction $\mathbb{E}(z_t u_{2,t+1}(\beta)) = 0$. This restriction is used by the J test based on the augmented model ("J augment" in Tables 1 and 2) whose power approaches 100% when $\tau \neq 0$. Because the augmented model regains its first-order local identifiability for $\tau \neq 0$, the standard inference (based on the $\chi^2(k_2 - p)$ distribution) restores its validity under both the null and the alternative hypotheses.

But when $\tau = 0$, the additional moment restriction $\mathbb{E}(z_t u_{2,t+1}(\beta)) = 0$ is redundant and the model remains first-order locally unidentified. This should manifest itself in size distortions under the null (for $n = 50{,}000$, the rejection rates of the augmented J test under the null are 14.9% and 8.6% at the 10% and 5% nominal level, respectively) and lack of power under the alternative. This highlights the need for a robust test that does not require prior knowledge of the true structure of the model and the value of τ.

Indeed, the $\hat{Z}$ test proposed in this paper offers precisely this type of robustness as it remains agnostic about the first-order identifiability of the model. Tables 1 and 2 demonstrate that the $\hat{Z}$ test controls size for both $\tau = 0$ and $\tau \neq 0$. The minor underrejections of the test for $m = 3$ arise from the fact that even for $n = 10{,}000$, the finite-sample distribution of the $\hat{Z}$ test is slightly asymmetric and it requires even larger sample sizes for the $N(0, 1)$ asymptotics to fully assert itself. Further simulation evidence and discussion regarding the absolute and relative finite-sample performance of the $\hat{Z}$ test is provided in the Supplemental Material.

6.2 *Empirical application*

In this subsection, we investigate the presence of a common CH factor in U.S. bond returns of different maturities. 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 $\mathbb{E}(u(y_t, \theta_0)|x_t) = 0$.

Let $r_{t+1}^{(j)}$ 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 $Y_{t+1} = (r_{t+1}^{(1)}, \dots, r_{t+1}^{(m)})'$. As in the previous subsection, we posit that the m-vector of excess bond returns Y_{t+1}, adapted to the increasing filtration $\mathcal{F}_t$, admits the representation (22) with common CH features.[13]

[12] In unreported results for $n = 50{,}000$ and $m = 2$, the empirical rejections for the standard J test based on $\chi^2(k_1 - p)$ are 18.2% and 11.3% for $\tau = 0$, and 16.9 and 10.1% for $\tau \neq 0$ at 10% and 5% nominal levels, respectively. A similar increase in overrejections for the J test are observed for $m = 3$ in sample sizes that exceed those reported in Table 2.

[13] The conventional term structure models impose no-arbitrage restrictions on the factor loading matrix Λ. Recent research (Duffee (2011); Joslin, Singleton, and Zhu (2011); among others) casts doubt on the role

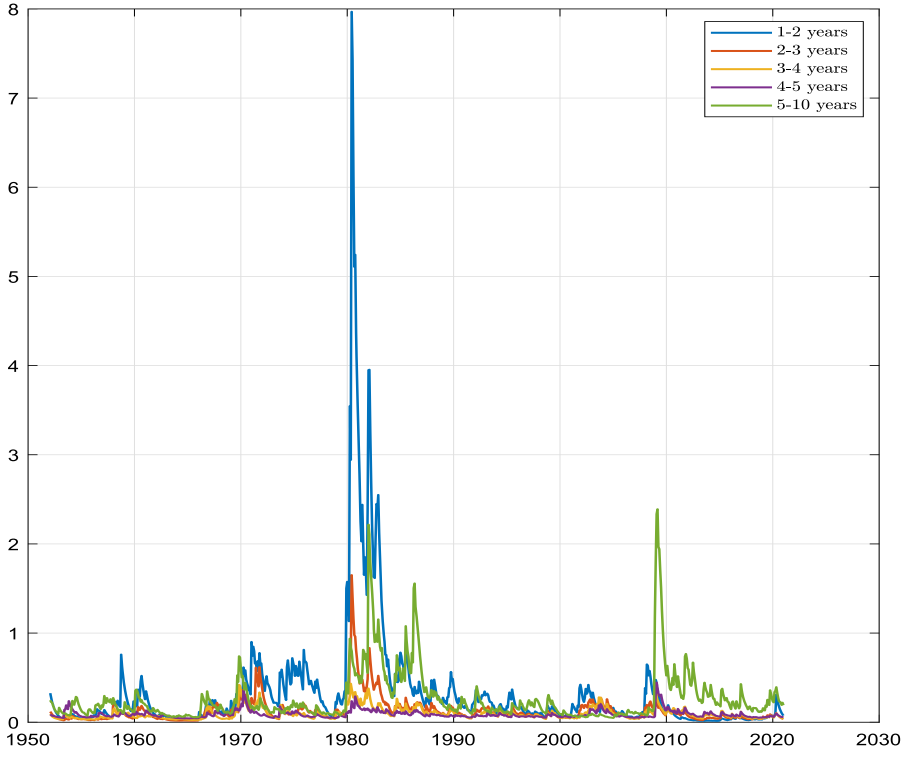

FIGURE 1. Estimated GARCH(1,1) volatilities for portfolio bond excess returns of different maturities.

This implies that there exists a vector $\beta \neq 0_m$ in $\mathbb{R}^m$ such that $\mathbb{E}((\beta' Y_{t+1})^2|\mathcal{F}_t)$ is constant, that is, $\mathbb{E}(u_{t+1}(\theta_0)|\mathcal{F}_t) = 0$ with $\theta = (\beta', c)'$ and $u_{t+1}(\theta) := (\beta' Y_{t+1})^2 - c$.

In the empirical analysis, we use the Fama bond portfolio returns from the Center for Research in Security Prices (CRSP) (2023) 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.[14] The data is at monthly frequency covering the period January 1952–December 2020. We construct excess bond returns by subtracting the 1-month risk-free rate (retrieved from Kenneth R. French—Data Library (2023)).

We start by fitting a GARCH(1,1) to each of these excess bond returns. The filtered GARCH volatilities are plotted in Figure 1. As the graph reveals, there appears to be a strong comovement 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.

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.

[14]The data for the bond portfolio returns is obtained from the Wharton Research Data Services (WRDS), using database CRSP Treasuries—Fama bond portfolios ©2023 (CRSP).

Crump and Gospodinov (2022) argue that the cross-sectionally differenced returns, $dr_{t+1}^{(j)} = r_{t+1}^{(j)} - r_{t+1}^{(j-1)}$, reveal better the underlying factor structure since the term-structure identities induce elevated local correlations in $r_{t+1}^{(j)}$ across maturities that obscure the true signal.[15] For this reason, we use the vector of differenced returns $Y_{t+1} = (r_{t+1}^{(1)}, dr_{t+1}^{(2)}, \dots, dr_{t+1}^{(m)})'$, 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.

We estimate the parameters by the two-step GMM, based on the conditional restriction $\mathbb{E}(u_{t+1}(\theta_0)|\mathcal{F}_t) = 0$, where $u_{t+1}(\beta, c) := (\beta_1 Y_{1,t+1} + \cdots + \beta_4 Y_{4,t+1} + (1 - \beta_1 - \cdots - \beta_4)Y_{5,t+1})^2 - c$. 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 $(-0.209, 0.191, 0.642, 0.438, -0.062)'$ which, interestingly, produce a tent-shaped pattern as in Cochrane and Piazzesi (2005). Figure 2 plots $(\hat{\beta}_1 Y_{1,t+1} + \cdots + \hat{\beta}_4 Y_{4,t+1} + (1 - \hat{\beta}_1 - \cdots - \hat{\beta}_4)Y_{5,t+1})^2$. While this graph reveals that the strong CH structure, observed in individual series, is largely eliminated, we subject this hypothesis to a formal test using our $\hat{Z}$ test statistic. The p-value of the one-sided $\hat{Z}$ test is 0.533, suggesting that the null $H_0 : \mathbb{E}((\beta' Y_{t+1})^2 - c|\mathcal{F}_t) = 0$ cannot be rejected.

7. 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 a particular economic theory. It is possible, however, that while the model remains globally identified, it suffers from first-order local identification failure. 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 gains that drive the faster rate of convergence. We also characterize the asymptotic distribution of the estimator, which is nonstandard and has a mass point at zero. Finally, we establish the asymptotic normality of the conditional moment restriction test in our setup. This result is obtained for time series data by resorting to central limit theorems for degenerate U-statistics of weakly dependent processes. Importantly, the limiting behavior is robust to first-order local identification failure regardless of the number of directions in which the Jacobian of the conditional moment restrictions is degenerate. The test has the additional advantage that it retains its validity even when first-order local identification holds.

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

[15] 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)).

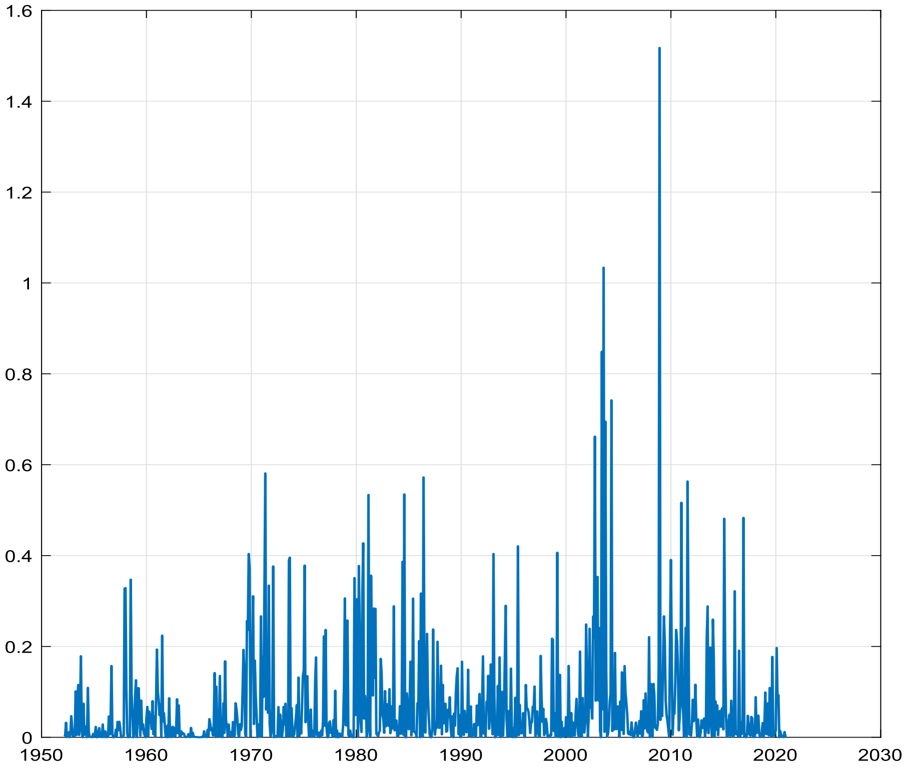

FIGURE 2. Plot of $(\hat{\beta}' Y_{t+1})^2$, where the estimates are based on $\mathbb{E}(u_{t+1}(\beta, c)|\mathcal{F}_t)) = 0$.

highly challenging nature of our setup: a conditional moment restrictions model with local first-order 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: IDENTIFICATION AND CONSISTENCY OF THE GMM ESTIMATOR

This Appendix covers some technicalities related to identification in conditional moment restrictions and consistency of GMM estimators with an increasing number of moment conditions. Proposition A.1 expresses in terms of unconditional moment restrictions, the notions of point identification and first-order local identification failure in conditional moment models. A general result on consistency is introduced in Theorem A.2, which is specialized to the two-step estimator by Corollary A.3. Assumption A.1 provides sufficient conditions for Assumption 3(ii) in the main text to hold as established by Lemma B.2.

A.1 *Identification*

PROPOSITION A.1. *Let* $(g_l)_{l\in\mathbb{N}}$ *be a countable basis* (*not necessarily orthonormal*) *of* $L^2(P)$. *Let* $\alpha_l(\theta) = \langle g_l(\cdot), \rho(\cdot, \theta)\rangle := \mathbb{E}(g_l(x)\rho(x, \theta))$ *and* $g^{(k)} := (g_1, \dots, g_k)'$.

(i) – *If the conditional moment restriction* (1) *satisfies the point identification property in* (4) *and if there exists* $k_0 \in \mathbb{N}$ *such that* $\alpha_l \equiv 0$ *for all* $l \geq k_0$, *then for all* $k \geq k_0$,

$$\forall \theta \in \Theta, \quad \mathbb{E}\big(g^{(k)}(x) \cdot u(y, \theta)\big) = 0 \quad \Leftrightarrow \quad \theta = \theta_0. \tag{A.1}$$

– *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*) $\lim_k [\sup_{\theta \in \Theta} \sum_{i \geq k} \alpha_i(\theta)^2] = 0$.

Then, if model (1) *satisfies the point identification property in* (4) *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*

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

(ii) *If* $\mathbb{E}(\nabla_\theta u(y, \theta_0)|x) \in (L^2(P))^p := L^2(P) \times \cdots \times L^2(P)$, *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* $k \geq k_0$,

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

Proof. See Appendix B. □

Part (i) of Proposition A.1 aims to establish an equivalence between point identification of conditional moment restriction (1)—as expressed by (4)—and point identification of the unconditional moment restriction $\mathbb{E}(g^{(k)}(x)u(y, \theta)) = 0$ 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.

A.2 *Consistency of GMM estimator*

Theorem A.2. *Under Assumptions* 1, 2(*i*), *and* 3, *if* $k/n \to 0$ *as* $n \to \infty$, *then the GMM estimator* $\hat{\theta}$*—defined by* (7)*—converges in probability to* θ_0.

The following assumption pertains to the consistency of the two-step 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) $\mathbb{E}[u(y,\theta_0)^{2\delta}|g_h(x)|^{\delta}|g_l(x)|^{\delta}] \le \Delta < \infty$, *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* $\mathbb{E}(g_l(x)^2 \sup_{\theta\in\mathcal{N}} u(y,\theta)^2) \le \Delta < \infty$ *and* $\mathbb{E}(g_l(x)^2 \sup_{\theta\in\mathcal{N}} \|\nabla_\theta u(y,\theta)\|^2) \le \Delta < \infty$.

(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.3. *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.2. Let $u_t := u(y_t,\theta_0)$. Note that

$$\mathbb{E}\big(\bar{f}_k(\theta_0)'\bar{f}_k(\theta_0)\big) = \frac{1}{n}\sum_{t,s=1}^{n}\mathbb{E}\big[g^{(k)}(x_t)'g^{(k)}(x_s)u_t u_s\big] = \frac{1}{n}\sum_{t=1}^{n}\mathbb{E}\big[g^{(k)}(x_t)'g^{(k)}(x_t)u_t^2\big]$$

$$= \mathbb{E}\big[g^{(k)}(x_t)'g^{(k)}(x_t)u_t^2\big] \le k\cdot \max_{1\le l\le k}\mathbb{E}\big[g_l(x_t)^2 u_t^2\big].$$

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 $\Theta\setminus\mathcal{N}$,

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

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

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

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

Proof of Corollary A.3. 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). □

Appendix B: Preliminary lemmas and proofs of main results

B.1 *Useful lemmas*

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

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

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

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*

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

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 $\bar{P}^{(k)} := \hat{W}_k^{1/2}\bar{D}_1(\bar{D}_1'\hat{W}_k\bar{D}_1)^{-1}\bar{D}_1'\hat{W}_k^{1/2}$, $M^{(k)} := I_k - P^{(k)}$, with $P^{(k)} := W_k^{1/2}D_1(D_1'W_kD_1)^{-1}D_1'W_k^{1/2}$, $\bar{f}_k := \bar{f}_k(\theta_0)$, and $\hat{M}^{(k)}$ defined as $\bar{M}^{(k)}$ but using $\hat{D}_1$. Finally, let*

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

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

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

and

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

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* $\|\hat{M}^{(k)} - \bar{M}^{(k)}\|_2 = O_P(n^{-1/2} \vee \|\hat{\theta} - \theta_0\|_2)$.

Lemma B.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 $\theta \in \Theta$,*

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

The proofs of Lemmas B.1, B.2, B.3, and B.4 are provided in the Supplemental Material. Lemma B.4 is a strengthened version of Lemma 1 of de Jong and Bierens (1994). 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 in Lemma B.4.

B.2 *Proofs of main results*

Proof of Proposition A.1. (i) It suffices to show that (A.1) 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

$$\mathbb{E}\big(u(y,\theta)|x\big) := \rho(x,\theta) = \sum_{l=1}^{k_0-1} \alpha_l(\theta) g_l(x) \quad \text{and}$$

$$\big[\rho(x,\theta) \equiv 0\big] \quad \Leftrightarrow \quad [\alpha_l = 0, \forall l = 1, \ldots, k_0 - 1].$$

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

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

Hence, by the definition of θ_k,

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

where the second equality holds by the stated assumption that $(g_l)_l$ are orthonormal and 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 (4), 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

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

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 $k \in \mathbb{N}$,

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

As a result,

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

Since $k \mapsto \text{Rank}(\mathbb{E}(g^{(k)}(x) \cdot \nabla_\theta u(y, \theta_0)))$ 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

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

Since $\mathbb{E}(\nabla_\theta u(y, \theta_0)|x) \in (L^2(P))^p$, its ith 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

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

and this completes the proof. □

Proof of Theorem 4.1. Let $R = (R_1|R_2)$ and consider the transformation $\theta = R\eta := R_1\eta_1 + R_2\eta_2$, with $\theta, \eta \in \mathbb{R}^p$, $\eta_1 \in \mathbb{R}^r$ and $\eta_2 \in \mathbb{R}^{p-r}$, 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

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

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

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

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

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

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

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

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

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

We show in the Supplemental Material that

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

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

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

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

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

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

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

Hence,

$$\begin{aligned}\|z_{1n}\|_2^4 \le{}& \|\gamma_k^{-1/2} H' W_k^{1/2} M^{(k)} W_k^{1/2}\bar{f}_k(\theta_0)\|_2 \cdot \|z_{1n}\|_2^2 \\ &+ o_P(1)\cdot\|z_{1n}\|_2 + o_P(1)\cdot\|z_{1n}\|_2^2 \\ &+ o_P(1)\cdot\|z_{1n}\|_2^3 + o_P(1)\cdot\|z_{1n}\|_2^4. \end{aligned} \tag{B.8}$$

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

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

Hence, if $\|z_{1n}\|_2 > 1$, this inequality implies $(1 + o_P(1))\|z_{1n}\|_2 \le O_P(1) + o_P(1)$. Thus, we either have $(\|z_{1n}\|_2 < 1)$ or $(1 + o_P(1))\|z_{1n}\|_2 \le O_P(1)$, 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 the range of the transpose a matrix and the null space of that same matrix, respectively, 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 in the Supplemental Material. □

The validity of the results in Theorem 4.1 hinges on verifying Assumptions 3(ii) and 4(iii, iv). For this reason, some further remarks on the rates of convergence in Theorem 4.1, specialized to the first-step and two-step GMM estimators, are warranted.

Remark 1. For the first-step GMM estimator, the weighting matrix $\hat{W}_k := W_{k,0}$ is nonrandom with bounded eigenvalues from above and away from zero, and 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 Assumption A.1(ii).

Remark 2. For the two-step estimator with $\hat{W}_k := \hat{V}_k^{-1}$, we assume that the smallest eigenvalue of V_k is bounded away from 0, that is, $\lambda_{\min}(V_k) \geq \underline{\lambda} > 0$ for all k. This is a reasonable assumption since $\lambda_{\max}(V_k)$ is an increasing sequence and we 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) ensures that

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

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

Remark 3. Finally, from Lemma B.2(iv), 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 $n \to \infty$.

Proof of Theorem 4.2. 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

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

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

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

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

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

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

$$\begin{aligned}
&\frac{1}{\gamma_k}\big(\dot{h}'\hat{W}_k\ddot{h} - H'W_kH\big) = o_P(1), \qquad H'\hat{W}_k\bar{D}_2 = O_P(k), \\
&\|\dot{h}-H\|_2 = O_P\big(\sqrt{k}\big[n^{-1/2} \vee \|\hat{\theta}-\theta_0\|_2\big]\big) = O_P\big(k^{1/4}/n^{1/4}\big), \\
&(\dot{h}-H)'\hat{W}_k\bar{D}_2 = O_P\big(k^{3/4}/n^{1/4}\big), \quad \text{and} \quad \frac{1}{\sqrt{\gamma_k}}\big(\dot{h}'\hat{W}_k - HW_k\big)\bar{f}_k(\theta_0) = o_P(1).
\end{aligned}$$

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

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

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

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

Therefore, it is not difficult to see that $|V| = 1_{Z \le 0}\sqrt{-2Z}$ and, since as a Gaussian random variable Z has a symmetric distribution, $|V| = 1_{Z \ge 0}\sqrt{2Z}$. 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 $z_{1n}^2 \xrightarrow{d} V^2 = 1_{Z \ge 0}(2Z)$ and this completes the proof. □

Proof of Theorem 5.2. 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,

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

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,

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

Note that

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

From Theorem 4.1, 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,

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

As a result, we have

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

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

$$\begin{aligned} &\bar{f}_k(\hat{\theta})' \hat{V}_k^{-1} \bar{f}_k(\hat{\theta}) \\ &\quad = \big[\bar{f}_k(\hat{\theta})' \hat{V}_k^{-1/2} \bar{P}^{(k)} \hat{V}_k^{-1/2} \bar{f}_k(\hat{\theta}) - \bar{f}_k(\theta_0)' \hat{V}_k^{-1/2} \bar{P}^{(k)} \hat{V}_k^{-1/2} \bar{f}_k(\theta_0)\big] + \bar{f}_k(\theta_0)' \hat{V}_k^{-1} \bar{f}_k(\theta_0) \\ &\qquad + (\hat{\eta}_2 - \eta_{02})' \bar{D}_2' \hat{V}_k^{-1/2} \bar{M}^{(k)} \hat{V}_k^{-1/2} \bar{D}_2 (\hat{\eta}_2 - \eta_{02}) \end{aligned}$$

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

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

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

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

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

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 $r_h=\text{Rank}(M^{(k)}V_k^{-1/2}H)\le p^2$. 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

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

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 5.1. In addition, it is not hard to see that $\mathbb{E}(U_{2n}) = 0$. Using similar arguments as in the proof of Proposition S.2 in the Supplemental Material, we can show that $\mathbb{E}(U_{2n}^2) = o(1)$. 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 $M_1 = S_1 S_1'$. In that respect, $\bar{f}_k(\theta_0)' V_k^{-1/2} M_1 V_k^{-1/2} \bar{f}_k(\theta_0) = \bar{f}_k(\theta_0)' V_k^{-1/2} S_1 S_1' V_k^{-1/2} \bar{f}_k(\theta_0)$. Also, $\mathrm{Var}(S_1' V_k^{-1/2} \bar{f}_k(\theta_0)) = I_{k-r_h}$. Using Theorem 5.1 and similar to the lines above for $\hat{Z}_2$, we can claim that

$$\hat{Z}_3 := \frac{\bar{f}_k(\theta_0)' V_k^{-1/2} S_1 S_1' V_k^{-1/2} \bar{f}_k(\theta_0) - (k - r_h)}{\sqrt{2(k - r_h)}} \xrightarrow{d} N(0, 1).$$

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 Theorem 5.3. From Lemma B.4, there exist k_0 and $\delta_0 > 0$ such that $k > k_0$ and

$\mathbb{E}(f_{k_0}(x_t, y_t, \hat{\theta}))' \mathbb{E}(f_{k_0}(x_t, y_t, \hat{\theta})) \geq \delta_0$. We have

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

Also,

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

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

Proof of nondecreasing γ_k in Equation (15). We need to show that for $k \geq k_0$, $\gamma_k \leq \gamma_{k+1}$. For this, we note that R_2 is fixed for $k \geq k_0$. Therefore, it suffices to show that for any $a \in \mathbb{R}^k$ and $b \in \mathbb{R}$, and letting $v := (a', b)' \in \mathbb{R}^{k+1}$,

$$a' M^{(k)} a \leq v' M^{(k+1)} v.$$

Let $\mu = \mathbb{E}(g_{k+1}(x) \cdot (\nabla_\theta u(y, \theta_0)'))$, $G_1 = G^{(k)}$ and $G = G^{(k+1)}$. We have

$$G = \begin{pmatrix} G_1 \\ \mu' \end{pmatrix}, \qquad M^{(k)} = I_k - G_1\bigl(G_1'G_1\bigr)^{-1}G_1', \quad \text{and} \quad M^{(k+1)} = I_{k+1} - G\bigl(G'G\bigr)^{-1}G'.$$

Write $A := G_1'G_1$. Using the Woodbury formula, we can claim that

$$\bigl(G'G\bigr)^{-1} = \bigl(A + \mu\mu'\bigr)^{-1} = A^{-1} - \frac{A^{-1}\mu\mu'A^{-1}}{1 + \mu'A^{-1}\mu}.$$

Thus,

$$\begin{aligned} v'M^{(k+1)}v &= a'a + b^2 - a'G_1\bigl(A + \mu\mu'\bigr)^{-1}G_1'a - 2ba'G_1\bigl(A + \mu\mu'\bigr)^{-1}\mu \\ &\quad - b^2\mu'\bigl(A + \mu\mu'\bigr)^{-1}\mu \\ &= a'M^{(k)}a + \frac{\bigl(b - a'G_1A^{-1}\mu\bigr)^2}{1 + \mu'A^{-1}\mu}. \end{aligned}$$

It follows that, for any $a \in \mathbb{R}^k$ and $b \in \mathbb{R}$,

$$v'M^{(k+1)}v \geq a'M^{(k)}a$$

and this completes the proof. □

References

Andrews, Donald (1984), "Non-strong mixing autoregressive processes." *Journal of Applied Probability*, 21, 930–934. [852]

Andrews, Donald (1991), "Asymptotic normality of series estimators for nonparametric and semiparametric regression models." *Econometrica*, 59, 307–345. [854]

Antoine, Bertille, Kevin Proulx, and Eric Renault (2020), "Pseudo-true SDFs in conditional asset pricing models." *Journal of Financial Econometrics*, 18, 656–714. [853]

Berman, Simeon (1964), "Limit theorems for the maximum term in stationary sequences." *Annals of Mathematical Statistics*, 35, 502–516. [866]

Bierens, Herman (1982), "Consistent model specification tests." *Journal of Econometrics*, 20, 105–134. [850]

Bierens, Herman (1990), "A consistent conditional moment test of functional form." *Econometrica*, 58, 1443–1458. [869]

Bierens, Herman and Werner Ploberger (1987), "Asymptotic theory of integrated conditional moment tests." *Econometrica*, 65, 1129–1151. [850]

Bougerol, Philippe and Nico Picard (1992), "Stationarity of GARCH processes and some nonnegative time series." *Journal of Econometrics*, 52, 115–127. [868]

Carrasco, Marine and Xiaohong Chen (2002), "Mixing and moment properties of various GARCH and stochastic volatility models." *Econometric Theory*, 18, 17–39. [860]

Carrasco, Marine and Jean-Pierre Florens (2000), "Generalization of GMM to a continuum of moment conditions." *Econometric Theory*, 16, 797–834. [850, 867]

Cattaneo, Matias, Michael Jansson, and Whitney Newey (2018), "Inference in linear regression models with many covariates and heteroskedasticity." *Journal of the American Statistical Association*, 113, 1350–1361. [861]

Center for Research in Security Prices (CRSP) (2023), "CRSP treasuries—fama bond portfolios." *Wharton Research Data Services*, https://wrds-www.wharton.upenn.edu/pages/about/data-vendors/center-for-research-in-security-prices-crsp/. [872]

Chao, John and Norman Swanson (2005), "Consistent estimation with a large number of weak instruments." *Econometrica*, 73, 1673–1692. [864]

Cochrane, John and Monika Piazzesi (2005), "Bond risk premia." *American Economic Review*, 95, 138–160. [873]

Crump, Richard and Nikolay Gospodinov (2022), "On the factor structure of bond returns." *Econometrica*, 90, 295–314. [872, 873]

de Jong, Robert and Herman 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. [850, 851, 854, 855, 878, 879]

Dominguez, Manuel and Ignacio Lobato (2004), "Consistent estimation of models defined by conditional moment restrictions." *Econometrica*, 72, 1601–1615. [850]

Domowitz, Ian and Halbert White (1982), "Misspecified models with dependent observations." *Journal of Econometrics*, 20, 35–58. [861]

Donald, Stephen, Guido Imbens, and Whitney Newey (2003), "Empirical likelihood estimation and consistent tests with conditional moment restrictions." *Journal of Econometrics*, 117, 55–93. [851, 867]

Dovonon, Prosper, and Nikolay Gospodinov (2024), "Supplement to 'Specification testing for conditional moment restrictions under local identification failure'." *Quantitative Economics Supplemental Material*, 15, https://doi.org/10.3982/QE2242. [852]

Dovonon, Prosper and Alastair Hall (2018), "The asymptotic properties of GMM and indirect inference under second-order identification." *Journal of Econometrics*, 205, 76–111. [853, 856, 857, 864]

Dovonon, Prosper and Eric Renault (2013), "Testing for common conditionally heteroskedastic factors." *Econometrica*, 81, 2561–2586. [850, 853, 855, 857, 863, 870]

Dovonon, Prosper and Eric Renault (2020), "GMM overidentification test with first order underidentification." Working Paper, Concordia University. [851, 856, 857, 864]

Duffee, Gregory (2011), "Forecasting with the term structure: The role of no-arbitrage restrictions." Working Paper, Johns Hopkins University. [871]

Engle, Robert and Sharon Kozicki (1993), "Testing for common features." *Journal of Business & Economic Statistics*, 11, 369–380. [857]

Engle, Robert, Victor Ng, and Michael Rothschild (1990), "Asset pricing with a factor-ARCH covariance structure: Empirical estimates for treasury bills." *Journal of Econometrics*, 45, 213–237. [858]

Fan, Jianqing, Yuan Liao, and Martina Mincheva (2013), "Large covariance estimation by thresholding principal orthogonal complements (with discussion)." *Journal of the Royal Statistical Society, Series B*, 75, 603–680. [861]

Francq, Christian and Jean-Michel 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. [860]

Fryzlewicz, Piotr and Suhasini Subba Rao (2011), "Mixing properties of ARCH and time-varying ARCH processes." *Bernoulli*, 17, 320–346. [860]

Gallant, Ronald (1981), "On the bias in flexible functional forms and an essentially unbiased form: The Fourier flexible form." *Journal of Econometrics*, 15, 211–245. [854]

Gospodinov, Nikolay, Raymond Kan, and Cesare Robotti (2017), "Spurious inference in reduced-rank asset-pricing models." *Econometrica*, 85, 1613–1628. [850]

Han, Chirok and Peter Phillips (2006), "GMM with many moment conditions." *Econometrica*, 74, 147–192. [861, 864]

Hansen, Lars Peter (2014), "Uncertainty outside and inside economic models." *Journal of Political Economy*, 122, 945–987. [853]

Isaev, Mikhail, Igor Rodionov, Rui-Ray Zhang, and Maksim Zhukovskii (2020), "Extreme value theory for triangular arrays of dependent random variables." *Russian Mathematical Surveys*, 75, 968–970. [866]

Joslin, Scott and Anh Le (2021), "Interest rate volatility and no-arbitrage affine term structure models." *Management Science*, 67, 7391–7416. [872]

Joslin, Scott, Kenneth Singleton, and Haoxiang Zhu (2011), "A new perspective on Gaussian dynamic term structure models." *Review of Financial Studies*, 24, 926–970. [871]

Kenneth R. French—Data Library (2023), "Fama/French 3 factors." *U.S. Research Returns Data*, http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html. [872]

Kim, Tae, Zhi-Ming Luo, and Chiho Kim (2011), "The central limit theorem for degenerate variable U-statistics under dependence." *Journal of Nonparametric Statistics*, 23, 683–699. [866]

King, Mervyn, Enrique Sentana, and Sushil Wadhwani (1994), "Volatility and links between national stock markets." *Econometrica*, 62, 901–933. [858, 868]

Kitamura, Yuichi, Gautam Tripathi, and Hyungtaik Ahn (2004), "Empirical likelihood-based inference in conditional moment restriction models." *Econometrica*, 72, 1667–1714. [850]

Lee, Ji and Zhipeng Liao (2018), "On standard inference for GMM with local identification failure of known forms." *Econometric Theory*, 34, 790–814. [850]

Lee, Lung-Fei and Andrew Chesher (1986), "Specification testing when score test statistics are identically zero." *Journal of Econometrics*, 31, 121–149. [856]

Newey, Whitney and Daniel McFadden (1994), "Large sample estimation and hypothesis testing." In *Handbook of Econometrics*, Vol. 4, 2111–2245. Chapter 36. [876]

Pereira, Helena (1983), "Rate of convergence towards a Fréchet type limit distribution." *Annales scientifiques de l'Université de Clermont-Ferrand II*, 76, 67–80. [866]

Pötscher, Benedikt and Ingmar Prucha (1989), "A uniform law of large numbers for dependent and heterogeneous data processes." *Econometrica*, 57, 675–683. [861]

Rotnitzky, Andrea, David Cox, Matteo Bottai, and James Robins (2000), "Likelihood based inference with singular information matrix." *Bernoulli*, 6, 243–284. [856]

Roussas, George and Dimitrios Ioannides (1987), "Moment inequalities for mixing sequences of random variables." *Stochastic Analysis and Applications*, 5, 61–120. [865]

Sargan, John Denis (1983), "Identification and lack of identification." *Econometrica*, 51, 1605–1633. [853, 855]

Sentana, Enrique and Gabriele Fiorentini (2001), "Identification, estimation and testing of conditionally heteroskedastic factor models." *Journal of Econometrics*, 102, 143–164. [858]

Tripathi, Gautam and Yuichi Kitamura (2003), "Testing conditional moment restrictions." *Annals of Statistics*, 31, 2059–2095. [867]

Co-editor Stéphane Bonhomme handled this manuscript.

Manuscript received 30 August, 2022; final version accepted 6 May, 2024; available online 6 May, 2024.

The replication package for this paper is available at https://doi.org/10.5281/zenodo.11098557. The authors were granted an exemption to publish parts of their data because either access to these data is restricted or the authors do not have the right to republish them. Therefore, the replication package only includes the codes and the parts of the data that are not subject to the exemption. However, the authors provided the Journal with (or assisted the Journal to obtain) temporary access to the restricted data. The Journal checked the provided and restricted data and the codes for their ability to reproduce the results in the paper and approved online appendices.