Econometrica Supplementary Material

SUPPLEMENT TO "TESTING FOR COMMON CONDITIONALLY HETEROSKEDASTIC FACTORS"

(*Econometrica*, Vol. 81, No. 6, November 2013, 2561–2586)

BY PROSPER DOVONON AND ERIC RENAULT

This supplementary material is organized as follows. The first part of Appendix A discusses the alternative regression-based approach for testing for common GARCH factors and highlights some singularities that make the standard Wald, likelihood ratio, and Lagrange multiplier tests invalid. The second part of Appendix A provides some additional interpretations of the mixture of rates of convergence that we get for the GMM estimators. It also highlights some key differences between the first-order local identification failure studied in this paper and the weak identification framework (see, e.g., Stock and Wright (2000), Kleibergen (2005), and Andrews and Mikusheva (2012)). Appendix B contains the proofs of all of the results stated in the main paper.

APPENDIX A: TESTING FOR CH EFFECTS ON COMMON FEATURES AND FURTHER REMARKS

A.1. *Testing for CH Effects on Common Features*

A REGRESSION-BASED APPROACH IS AKIN to considering an instrumental heteroskedasticity model that can be written

$$\xi_{t+1} = a + bz_t + \varepsilon_{t+1}, \tag{A.1}$$

where ξ_{t+1} is an m-dimensional vector that gathers some coefficients of the matrix $Y_{t+1}Y'_{t+1}$, z_t is again a vector of H $\mathfrak{F}_t$-measurable instruments, $E(\varepsilon_{t+1}) = 0$, $\text{Cov}(\varepsilon_{t+1}, z_t) = 0$, and $a \in \mathbb{R}^m$ and $b \in \mathbb{R}^{m\times H}$ are vectors of unknown parameters. It will be possible to check from this regression model that the portfolio $\theta' Y_{t+1}$ is a CH common feature insofar as

$$\left(\theta' Y_{t+1}\right)^2 = \theta' Y_{t+1} Y'_{t+1}\theta = \gamma'(\theta)\xi_{t+1}$$

for some known function γ from Θ to $\mathbb{R}^{mH}$. In other words, the regression model (1) must be rich enough to be such that the vector ξ_{t+1} gathers in particular the coefficients of $Y_{t+1}Y'_{t+1}$ that show up in $\theta' Y_{t+1}Y'_{t+1}\theta$. In these circumstances, the null hypothesis of interest for the test of CH common features will be

$$H_0 : \exists\theta: \quad \gamma'(\theta)b = 0.$$

Irrespective of the preferred test chosen from the trinity of asymptotic tests used to test such a composite hypothesis (see Gourieroux and Monfort (1989)),

 DOI: 10.3982/ECTA10082

the standard asymptotic chi-squared distribution under the null will be warranted only if

$$\left.\frac{\partial[\gamma'(\theta)b^0]}{\partial\theta}\right|_{\theta=\theta^0}$$

is full rank. However, from the regression model,

$$\left(\theta' Y_{t+1}\right)^2 = \gamma'(\theta)a^0 + \gamma'(\theta)b^0 z_t + \gamma'(\theta)\varepsilon_{t+1},$$

and thus

$$\frac{\partial(\theta' Y_{t+1})^2}{\partial\theta} = \frac{\partial\gamma'(\theta)}{\partial\theta}a^0 + \frac{\partial\gamma'(\theta)}{\partial\theta}b^0 z_t + \frac{\partial\gamma'(\theta)}{\partial\theta}\varepsilon_{t+1}$$

and

$$\mathrm{Cov}\left(\frac{\partial(\theta' Y_{t+1})^2}{\partial\theta}, z_t\right) = \frac{\partial\gamma'(\theta)}{\partial\theta}\mathrm{Cov}\left(b^0 z_t, z_t\right).$$

Therefore,

$$\left.\frac{\partial\gamma'(\theta)}{\partial\theta}b^0\right|_{\theta=\theta^0} = \left.\mathrm{Cov}\left(\frac{\partial(\theta' Y_{t+1})^2}{\partial\theta}, z_t\right)\right|_{\theta=\theta^0}\left(\mathrm{Var}(z_t)\right)^{-1} = 0$$

by Proposition 2.1 when θ^0 is a common feature.

A.2. *Heterogeneity of Rates of Convergence*

REMARK A.1: It is worth interpreting the heterogeneity of rates of convergence across the sample space in terms of the randomness of a (population) matrix that may be seen as a Fisher information matrix. While randomness of the information matrix is known to occur in some nonergodic settings, it has been recently considered as a possibility in the context of weak identification by Andrews and Mikusheva (2012), even though they eventually precluded this possibility by their maintained Assumption 1(b). By a slight abuse of language, we will use here their information theoretic terminology, even though we are in a GMM context that is more general than maximum likelihood for which Fisher information matrices are usually defined. The GMM analogs of the score vector and of the Hessian of the log-likelihood will be defined from the criterion function

$$Q_T^W(\theta) = \frac{T}{2}\bar{\phi}'_T(\theta)W_T\bar{\phi}_T(\theta).$$

The GMM analog of the outer product of the score is then

$$I_T^W(\theta) = \frac{1}{T}\frac{\partial Q_T^W(\theta)}{\partial \theta} \cdot \frac{\partial Q_T^W(\theta)}{\partial \theta'}$$
$$= \frac{\partial \bar{\phi}'_T(\theta)}{\partial \theta} W_T \big(\sqrt{T}\bar{\phi}_T(\theta)\big)\big(\sqrt{T}\bar{\phi}'_T(\theta)\big) W_T \frac{\partial \bar{\phi}_T(\theta)}{\partial \theta'}$$

such that, under regularity conditions, we have, at true value $\theta = \theta^0$,

$$I^W(\theta^0) = \lim_{T\to\infty} E\big(I_T^W(\theta^0)\big) = \Gamma'(\theta^0) W \Sigma(\theta^0) W \Gamma(\theta^0).$$

The GMM analog of the Hessian matrix of the log-likelihood is

$$H_T^W(\theta) = \frac{1}{T}\frac{\partial^2 Q_T^W(\theta)}{\partial \theta\, \partial \theta'} \tag{A.2}$$
$$= \frac{\partial \bar{\phi}'_T(\theta)}{\partial \theta} W_T \frac{\partial \bar{\phi}_T(\theta)}{\partial \theta'} + \big(h'_{ijT}(\theta) W_T \bar{\phi}_T(\theta)\big)_{1\le i,j\le p},$$

where

$$h_{ijT}(\theta) = \frac{\partial^2 \bar{\phi}_T(\theta)}{\partial \theta_i\, \partial \theta_j}.$$

In particular, under regularity conditions, we have, at true value $\theta = \theta^0$,

$$H^W(\theta^0) = \operatorname*{plim}_{T\to\infty} H_T^W(\theta^0) = \Gamma'(\theta^0) W \Sigma(\theta^0) W \Gamma(\theta^0).$$

In the (standard) strong identification case, $\Gamma(\theta^0)$ is full column rank. Then for large T, both the expected outer product matrix $E(I_T^W(\theta^0))$ and the Hessian matrix $H_T^W(\theta^0)$ are positive definite with probability 1. Moreover, for the efficient choice $W = \Sigma^{-1}(\theta^0)$ of the weighting matrix, the difference $(E(I_T^W(\theta^0)) - H_T^W(\theta^0))$ converges to zero in probability. This generalization to non-maximum-likelihood contexts of the so-called second informational equality was put forward by Gourieroux and Monfort (1989) as the necessary and sufficient condition to keep the asymptotic equivalence between the standard asymptotic tests.

The situation is much different in the context of weak identification (drifting data generating process (DGP) such that $\Gamma(\theta^0) = O(1/\sqrt{T})$; see, e.g., Kleibergen (2005)) or in our context ($\Gamma(\theta^0) = 0$). Then, as stressed by Andrews and Mikusheva (2012), "the difference between the two information matrices is asymptotically non-negligible compared with the information measure" itself. While they point this out in a maximum-likelihood context, this statement remains true in a GMM context, regardless of the choice of the weighting matrix W. There is, however, an important difference between our setting and the common weak identification framework.

In the common "GMM with weak identification" asymptotics as developed by Stock and Wright (2000), the drifting DGP introduces a perverse factor $(1/\sqrt{T})$ at the level of the moment condition itself and this factor will go through all derivatives of moment conditions. Then both $E(I_T^W(\theta^0))$ and $H_T^W(\theta^0)$, and their difference as well are all of order $1/T$.

In our framework, while $E(I_T^W(\theta^0))$ is still of order $1/T$, the Hessian matrix $H_T^W(\theta^0)$ is now dominating since we can deduce from (A.2) that

$$\sqrt{T}H_T^W(\theta^0) \xrightarrow{d} Z(X),$$

where

$$Z(X) = \big(h'_{ij}(\theta^0)WX\big)_{1\le i,j\le p}, \qquad h'_{ij}(\theta^0) = \operatorname*{plim}_{T\to\infty} h'_{ijT}(\theta^0),$$

and X is defined by the Gaussian limit in distribution of $\sqrt{T}\ \bar{\phi}_T(\theta^0)$:

$$\sqrt{T}\bar{\phi}_T(\theta^0) \xrightarrow{d} X \sim N\big(0, \Sigma(\theta^0)\big).$$

Note that since the variance matrix $\Sigma(\theta^0)$ is nonsingular, $Z(X)$ is a nondegenerate random matrix, the coefficients of which are all Gaussian with zero mean.

APPENDIX B: PROOFS

Throughout this appendix, we denote by Δ and $\bar{\Delta}$ the $\mathbb{R}^H$-valued functions defined by

$$\Delta(v) = \left(v'\frac{\partial^2\rho_h}{\partial\theta\,\partial\theta'}(\theta^0)v\right)_{1\le h\le H} \quad \text{and}$$

$$\bar{\Delta}(v) = \left(v'\frac{\partial^2\bar{\phi}_{h,T}}{\partial\theta\,\partial\theta'}(\theta^0)v\right)_{1\le h\le H}, \quad \forall v\in\mathbb{R}^p,$$

$p = n-1$ and $n = \dim(Y_t)$. We let G and $\bar{G}$ be two (H, p^2) matrices defined such that $\Delta(v) = G\,\mathrm{Vec}(vv')$ and $\bar{\Delta}(v) = \bar{G}\,\mathrm{Vec}(vv')$ for all $v \in \mathbb{R}^p$. By definition,

$$G = \left(\mathrm{Vec}\left(\frac{\partial^2\rho_1}{\partial\theta\,\partial\theta'}(\theta^0)\right), \mathrm{Vec}\left(\frac{\partial^2\rho_2}{\partial\theta\,\partial\theta'}(\theta^0)\right), \dots, \mathrm{Vec}\left(\frac{\partial^2\rho_H}{\partial\theta\,\partial\theta'}(\theta^0)\right)\right)'$$

and $\bar{G}$ has the same expression but with $\bar{\phi}_{h,T}$ instead of ρ_h, $h = 1, \dots, H$.

Lemmas B.1–B.6 below connect as follows in proving the main results in the paper. Lemma B.1 is relevant thanks to Lemma 2.3 and is useful to derive the rate of convergence as stated in Propositions 3.1 and 3.2. This lemma is also useful to establish part of Lemma B.6. Lemma B.2 is useful to establish Proposition 3.2. Lemma B.3 is used in the proof of Lemma B.5. Lemma B.5 is useful to establish Lemma B.6, part (iii) of which essentially proves Theorem 3.1, while Lemma B.4 is used in the proof of Theorem 3.2.

LEMMA B.1: *If* $(\Delta(v) = 0) \Rightarrow (v = 0))$, *then there exists* $\gamma > 0$ *such that for any* $v \in \mathbb{R}^p$,

$$\|\Delta(v)\| \geq \gamma \|v\|^2.$$

PROOF: The function $\Delta(v)$ is homogeneous of degree 2 with respect to v. Therefore, for all $v \in \mathbb{R}^p$,

$$\|\Delta(v)\| = \|v\|^2 \left\| \Delta\left(\frac{v}{\|v\|}\right) \right\|.$$

Define $\gamma = \inf_{\|v\|=1} \|\Delta(v)\|$. From the compactness of $\{v \in \mathbb{R}^p : \|v\| = 1\}$ and the continuity of $\Delta(v)$, there exists v^* such that $\|v^*\| = 1$ and $\gamma = \|\Delta(v^*)\|$. Then $\Delta(v^*) \neq 0$ since $v^* \neq 0$ and this shows the expected result. *Q.E.D.*

LEMMA B.2: *Let* $\{X_T : T \in \mathbb{N}\}$ *and* $\{\varepsilon_T : T \in \mathbb{N}\}$ *be two sequences of real-valued random variables such that* ε_T *converges in probability toward* 0 *and for all* T, $X_T \leq \varepsilon_T$ *almost surely* (*a.s.*). *Then*

$$\limsup_{T\to\infty} \text{Prob}(X_T \leq \epsilon) = 1, \quad \forall \epsilon > 0.$$

PROOF: Let $\epsilon > 0$. We have

$$\limsup_{T\to\infty} \text{Prob}(X_T \leq \epsilon) = 1 - \liminf_{T\to\infty} \text{Prob}(X_T > \epsilon).$$

But

$$\inf_{n\geq T} \text{Prob}(X_n > \epsilon) \leq \text{Prob}(X_T > \epsilon) \leq \text{Prob}(\varepsilon_T > \epsilon) \to 0$$

as $T \to \infty$. This establishes the result. *Q.E.D.*

LEMMA B.3: *Let* $\{X_T : T \in \mathbb{N}\}$ *be a sequence of real-valued random variables, let* X *a real-valued random variable, and let* $x \in \mathbb{R}$. *If* X_T *converges in distribution toward* X, *then*

$$\forall \epsilon > 0, \exists T_\epsilon \in \mathbb{N}: \quad \forall T \geq T_\epsilon,$$
$$\left|\text{Prob}(X_T \leq x) - \text{Prob}(X \leq x)\right| \leq \text{Prob}(X = x) + \epsilon.$$

PROOF: Since $\inf_{n\geq T}\text{Prob}(X_n\leq x)\leq \text{Prob}(X_T\leq x)\leq \sup_{n\geq T}\text{Prob}(X_n\leq x)$, we can write

$$\begin{aligned}&\left|\text{Prob}(X_T\leq x)-\text{Prob}(X\leq x)\right|\\&\quad=\max\big(\text{Prob}(X_T\leq x)-\text{Prob}(X\leq x);\\&\qquad-\text{Prob}(X_T\leq x)+\text{Prob}(X\leq x)\big)\\&\quad\leq\max\Big(\sup_{n\geq T}\text{Prob}(X_n\leq x)-\text{Prob}(X\leq x);\\&\qquad-\inf_{n\geq T}\text{Prob}(X_n\leq x)+\text{Prob}(X\leq x)\Big)\\&\quad\leq\max\Big(\Big|\sup_{n\geq T}\text{Prob}(X_n\leq x)-\text{Prob}(X\leq x)\Big|;\\&\qquad\Big|-\inf_{n\geq T}\text{Prob}(X_n\leq x)+\text{Prob}(X\leq x)\Big|\Big).\end{aligned}$$

Thus, to complete the proof, it suffices to show that for any $\epsilon>0$, there exists $T_\epsilon\in\mathbb{N}$: $\forall T\geq T_\epsilon$, both

$$\Big|\sup_{n\geq T}\text{Prob}(X_n\leq x)-\text{Prob}(X\leq x)\Big|\leq\text{Prob}(X=x)+\epsilon \tag{B.1}$$

and

$$\Big|-\inf_{n\geq T}\text{Prob}(X_n\leq x)+\text{Prob}(X\leq x)\Big|\leq\text{Prob}(X=x)+\epsilon \tag{B.2}$$

hold.

To establish (B.1), we observe that $\sup_{n\geq T}\text{Prob}(X_n\leq x)\to L=\limsup_{T\to\infty}\text{Prob}(X_T\leq x)$. Hence, there exists $T_{1,\epsilon}\in\mathbb{N}$ such that $\forall T\geq T_{1,\epsilon}$, $|\sup_{n\geq T}\text{Prob}(X_n\leq x)-L|<\epsilon$.

Hence,

$$\begin{aligned}&\forall T\geq T_{1,\epsilon},\\&\quad\Big|\sup_{n\geq T}\text{Prob}(X_n\leq x)-\text{Prob}(X\leq x)\Big|\leq\epsilon+\big|\text{Prob}(X\leq x)-L\big|.\end{aligned}$$

Now, since $X_T\xrightarrow{d}X$, by the portmanteau lemma (Lemma 2.2(vi) of van der Vaart (1998)), we have

$$L=\limsup_{T\to\infty}\text{Prob}(X_T\leq x)\leq\text{Prob}(X\leq x).$$

Hence, $|\operatorname{Prob}(X \leq x) - L| = \operatorname{Prob}(X \leq x) - L$. But

$$L \geq \limsup_{T\to\infty} \operatorname{Prob}(X_T < x) \geq \liminf_{T\to\infty} \operatorname{Prob}(X_T < x) \geq \operatorname{Prob}(X < x).$$

(The first two inequalities hold by definition, whereas the last one holds by the portmanteau lemma (Lemma 2.2(v) of van der Vaart (1998)).) Thus

$$\begin{aligned}\left|\operatorname{Prob}(X \leq x) - L\right| &= \operatorname{Prob}(X \leq x) - L \\ &\leq \operatorname{Prob}(X \leq x) - \operatorname{Prob}(X < x) = \operatorname{Prob}(X = x).\end{aligned}$$

This establishes (B.1).

We establish (B.2) along similar lines: $\inf_{n\geq T} \operatorname{Prob}(X_n \leq x) \to l = \liminf_{T\to\infty} \operatorname{Prob}(X_T \leq x)$. Hence, there exists $T_{2,\epsilon} \in \mathbb{N}$ such that $\forall T \geq T_{2,\epsilon}$, $|-\inf_{n\geq T} \operatorname{Prob}(X_n \leq x) + l| < \epsilon$. Hence,

$$\begin{aligned}&\forall T \geq T_{2,\epsilon}, \\ &\quad\left|-\inf_{n\geq T} \operatorname{Prob}(X_n \leq x) + \operatorname{Prob}(X \leq x)\right| \leq \epsilon + \left|\operatorname{Prob}(X \leq x) - l\right|.\end{aligned}$$

Since $\operatorname{Prob}(X \leq x) \geq L = \limsup_{T\to\infty} \operatorname{Prob}(X_T \leq x)$, we can also claim that $\operatorname{Prob}(X \leq x) \geq l$. Therefore, $|\operatorname{Prob}(X \leq x) - l| = \operatorname{Prob}(X \leq x) - l$. But

$$\begin{aligned}\operatorname{Prob}(X \leq x) &= \operatorname{Prob}(X < x) + \operatorname{Prob}(X = x) \\ &\leq \liminf_{T\to\infty} \operatorname{Prob}(X_T < x) + \operatorname{Prob}(X = x) \\ &\leq l + \operatorname{Prob}(X = x).\end{aligned}$$

Thus $\operatorname{Prob}(X \leq x) - l \leq \operatorname{Prob}(X = x)$. This establishes (B.2) and, therefore, the lemma with $T_\epsilon = \max(T_{1,\epsilon}, T_{2,\epsilon})$. *Q.E.D.*

LEMMA B.4: *Under the same conditions as Theorem* 3.2, *there exists an* (H, p) *matrix* G_1 $(p = n - 1)$ *and a* (p, p^2) *matrix* G_2 *such that*

$$G = G_1 G_2 \quad \textit{and} \quad \operatorname{Rank}(G) = \operatorname{Rank}(G_1) = \operatorname{Rank}(G_2) = p.$$

PROOF: Let $\theta_* = (\theta', 1 - \sum_{i=1}^{n-1} \theta_i)'$, $\theta \in \mathbb{R}^{n-1}$. We recall that $\rho(\theta) = E[z_t((\theta_*' Y_{t+1})^2 - c(\theta_*))]$. We have

$$\begin{aligned}\rho(\theta) &= E\left[\left(z_t - E(z_t)\right)\left(\theta_*' Y_{t+1}\right)^2\right] = E\left[\left(z_t - E(z_t)\right)\left(\theta_*' Y_{t+1} Y_{t+1}' \theta_*\right)\right] \\ &= E\left[\left(z_t - E(z_t)\right) E\left(\theta_*' Y_{t+1} Y_{t+1}' \theta_* | \mathfrak{F}_t\right)\right] \\ &= E\left[\left(z_t - E(z_t)\right) \theta_*' \Lambda D_t \Lambda' \theta_*\right] \\ &= E\left[\left(z_t - E(z_t)\right) \operatorname{tr}\left(D_t \Lambda' \theta_* \theta_*' \Lambda\right)\right]\end{aligned} \tag{B.3}$$

$$
\begin{aligned}
&= E\big[\big(z_t - E(z_t)\big)\,\mathrm{Diag}'(D_t)\big]\,\mathrm{Diag}\big(\Lambda'\theta_*\theta_*'\Lambda\big)\\
&= \mathrm{Cov}\big(z_t, \mathrm{Diag}(D_t)\big)\,\mathrm{Diag}\big(\Lambda'\theta_*\theta_*'\Lambda\big)\\
&\equiv G_1\,\mathrm{Diag}\big(\Lambda'\theta_*\theta_*'\Lambda\big),
\end{aligned}
$$

where $G_1 = \mathrm{Cov}(z_t, \mathrm{Diag}(D_t))$ is an (H, p) matrix of rank p by Assumption 3. Then, by computing the second-order derivatives at θ^0, we deduce that

$$G = G_1 G_2$$

for some (p, p^2) matrix G_2. We now show that G_2 has full row rank p. We proceed by contradiction. If G_2 does not have full row rank, G itself would be of rank smaller than p and the null space of G would be of dimension larger than $p^2 - p$. This cannot be true since, by Lemma 2.3,

$$G\,\mathrm{Vec}\big(vv'\big) = 0 \Rightarrow v = 0$$

and, clearly, none of the p linearly independent vectors $\mathrm{Vec}(e_i e_i')$, $i = 1, \dots, p$, where $\{e_i : i = 1, \dots, p\}$ is the canonical basis of $\mathbb{R}^p$ (all the components of e_i are zero except the ith one equal to 1), belongs to the null space of G. *Q.E.D.*

LEMMA B.5: *Let $\hat{M}_T(v)$ and $M(v)$ be two real-valued stochastic processes with continuous sample paths indexed by $\mathbb{R}^p$ and let $\{\mathbb{V}_T : T \in \mathbb{N}\}$ be a nondecreasing sequence of subsets of $\mathbb{R}^p$ such that $\bigcup_{T\geq 0} \mathbb{V}_T = \mathbb{R}^p$. If*

(i) *$\hat{M}_T(\cdot)$ converges in distribution toward $M(\cdot)$ in $\ell^\infty(\mathbb{K})$ for every compact $\mathbb{K} \subset \mathbb{R}^p$, where $\ell^\infty(\mathbb{K})$ is the set of all bounded real-valued functions on $\mathbb{K}$, endowed with the sup-norm,*

(ii) *there exists $\hat{v}_T \in \arg\min_{v\in\mathbb{V}_T} \hat{M}_T(v)$ that is uniformly tight, and*

(iii) *there exists $\hat{v} \in \arg\min_{v\in\mathbb{R}^p} M(v)$ which is tight,*

then

$$\hat{M}_T(\hat{v}_T) \xrightarrow{d} M(\hat{v}).$$

PROOF: We show that $\mathrm{Prob}(\hat{M}_T(\hat{v}_T) \leq x) \to \mathrm{Prob}(M(\hat{v}) \leq x)$ as $T \to \infty$ for any continuity point x of the cumulative distribution of $M(\hat{v})$. Let $x \in \mathbb{R}$ be such a point and let $\epsilon > 0$. Since $\hat{v}_T$ is uniformly tight and $\hat{v}$ is tight, there exists $m_\epsilon > 0$ such that

$$\sup_T \mathrm{Prob}\big(\|\hat{v}_T\| > m_\epsilon\big) < \frac{\epsilon}{4} \quad \text{and} \quad \mathrm{Prob}\big(\|\hat{v}\| > m_\epsilon\big) < \frac{\epsilon}{4},$$

and from condition (i) of the lemma, $\hat{M}_T(\cdot)$ converges toward $M(\cdot)$ in distribution in $\ell^\infty(\{v : \|v\| \leq m_\epsilon\})$. Since the function inf is continuous on $\ell^\infty(\mathbb{K})$, for any nonempty compact $\mathbb{K}$, we can apply the continuous mapping theorem and deduce that

$$\inf_{\|v\|\le m_\epsilon} \hat{M}_T(v) \xrightarrow{d} \inf_{\|v\|\le m_\epsilon} M(v).$$

Hence, from sample path continuity of $\hat{M}_T(\cdot)$ and $M(\cdot)$, we have

$$\min_{\|v\|\le m_\epsilon} \hat{M}_T(v) \xrightarrow{d} \min_{\|v\|\le m_\epsilon} M(v).$$

Hence, from Lemma B.3, there exists T_ϵ such that for all $T > T_\epsilon$, $\{v : \|v\| < m_\epsilon\} \subset \mathbb{V}_T$ and

$$\begin{aligned}&\left|\operatorname{Prob}\Bigl(\min_{\|v\|\le m_\epsilon} \hat{M}_T(v) \le x\Bigr) - \operatorname{Prob}\Bigl(\min_{\|v\|\le m_\epsilon} M(v) \le x\Bigr)\right| \\ &\quad < \frac{\epsilon}{4} + \operatorname{Prob}\Bigl(\min_{\|v\|\le m_\epsilon} M(v) = x\Bigr).\end{aligned}$$

But, clearly,

$$\begin{aligned}&\operatorname{Prob}\Bigl(\min_{\|v\|\le m_\epsilon} M(v) = x\Bigr) \\ &\quad = \operatorname{Prob}\Bigl(\min_{\|v\|\le m_\epsilon} M(v) = x, \|\hat{v}\| \le m_\epsilon\Bigr) \\ &\qquad + \operatorname{Prob}\Bigl(\min_{\|v\|\le m_\epsilon} M(v) = x, \|\hat{v}\| > m_\epsilon\Bigr) \\ &\quad \le \operatorname{Prob}\Bigl(\min_{\|v\|\le m_\epsilon} M(v) = x, M(\hat{v}) = \min_{\|v\|\le m_\epsilon} M(v)\Bigr) \\ &\qquad + \operatorname{Prob}\Bigl(\min_{\|v\|\le m_\epsilon} M(v) = x, \|\hat{v}\| > m_\epsilon\Bigr) \\ &\quad \le \operatorname{Prob}\bigl(M(\hat{v}) = x\bigr) + \operatorname{Prob}\bigl(\|\hat{v}\| > m_\epsilon\bigr) \\ &\quad \le 0 + \frac{\epsilon}{4}.\end{aligned}$$

It follows that $\forall T > T_\epsilon$,

$$\left|\operatorname{Prob}\Bigl(\min_{\|v\|\le m_\epsilon} \hat{M}_T(v) \le x\Bigr) - \operatorname{Prob}\Bigl(\min_{\|v\|\le m_\epsilon} M(v) \le x\Bigr)\right| < \frac{\epsilon}{2}.$$

Also,

$$\begin{aligned}&\bigl(\hat{M}_T(\hat{v}_T) \le x\bigr) \\ &\quad = \bigl(\hat{M}_T(\hat{v}_T) \le x; \|\hat{v}_T\| \le m_\epsilon\bigr) \cup \bigl(\hat{M}_T(\hat{v}_T) \le x; \|\hat{v}_T\| > m_\epsilon\bigr)\end{aligned}$$

$$= \Big(\min_{\|v\|\le m_\epsilon} \hat{M}_T(v) \le x;\ \|\hat{v}_T\| \le m_\epsilon\Big) \cup \big(\hat{M}_T(\hat{v}_T) \le x;\ \|\hat{v}_T\| > m_\epsilon\big)$$

$$= \Big[\Big(\min_{\|v\|\le m_\epsilon} \hat{M}_T(v) \le x\Big) \setminus \Big(\min_{\|v\|\le m_\epsilon} \hat{M}_T(v) \le x;\ \|\hat{v}_T\| > m_\epsilon\Big)\Big]$$
$$\cup \big(\hat{M}_T(\hat{v}_T) \le x;\ \|\hat{v}_T\| > m_\epsilon\big).$$

Thus,

$$\operatorname{Prob}\big(\hat{M}_T(\hat{v}_T) \le x\big) - \operatorname{Prob}\Big(\min_{\|v\|\le m_\epsilon} \hat{M}_T(v) \le x\Big) \le \operatorname{Prob}\big(\|\hat{v}_T\| > m_\epsilon\big).$$

We can actually replace $\hat{M}_T(\hat{v}_T)$ by $\min_{\|v\|\le m_\epsilon} \hat{M}_T(v)$ in the previous set operations and deduce that

$$\operatorname{Prob}\Big(\min_{\|v\|\le m_\epsilon} \hat{M}_T(v) \le x\Big) - \operatorname{Prob}\big(\hat{M}_T(\hat{v}_T) \le x\big) \le \operatorname{Prob}\big(\|\hat{v}_T\| > m_\epsilon\big).$$

Therefore,

$$\Big|\operatorname{Prob}\big(\hat{M}_T(\hat{v}_T) \le x\big) - \operatorname{Prob}\Big(\min_{\|v\|\le m_\epsilon} \hat{M}_T(v) \le x\Big)\Big|$$
$$\le \operatorname{Prob}\big(\|\hat{v}_T\| > m_\epsilon\big) < \frac{\epsilon}{4}.$$

In the same way, we also have

$$\Big|\operatorname{Prob}\big(M(\hat{v}) \le x\big) - \operatorname{Prob}\Big(\min_{\|v\|\le m_\epsilon} M(v) \le x\Big)\Big| \le \operatorname{Prob}\big(\|\hat{v}\| > m_\epsilon\big) < \frac{\epsilon}{4}.$$

Now we observe that

$$\big|\operatorname{Prob}\big(\hat{M}_T(\hat{v}_T) \le x\big) - \operatorname{Prob}\big(M(\hat{v}) \le x\big)\big|$$
$$\le \Big|\operatorname{Prob}\big(\hat{M}_T(\hat{v}_T) \le x\big) - \operatorname{Prob}\Big(\min_{\|v\|\le m_\epsilon} \hat{M}_T(v) \le x\Big)\Big|$$
$$+ \Big|\operatorname{Prob}\Big(\min_{\|v\|\le m_\epsilon} \hat{M}_T(v) \le x\Big) - \operatorname{Prob}\Big(\min_{\|v\|\le m_\epsilon} M(v) \le x\Big)\Big|$$
$$+ \Big|\operatorname{Prob}\Big(\min_{\|v\|\le m_\epsilon} M(v) \le x\Big) - \operatorname{Prob}\big(M(\hat{v}) \le x\big)\Big|.$$

Hence, for any $T > T_\epsilon$, $|\operatorname{Prob}(\hat{M}_T(\hat{v}_T) \le x) - \operatorname{Prob}(M(\hat{v}) \le x)| < 4\epsilon/4$. This completes the proof. *Q.E.D.*

LEMMA B.6: *Under the same conditions as Theorem* 3.1, *we have the following situations*.

(i) *The stochastic process* $\hat{J}^W(\cdot)$ *converges in distribution toward* $J^W(\cdot)$ *in* $\ell^\infty(\mathbb{K})$ *for every compact* $\mathbb{K} \subset \mathbb{R}^p$.

(ii) *Any* $\hat{v}_T \in \arg\min_{v\in\mathbb{H}_T} \hat{J}^W(v)$ *and* $\hat{v} \in \arg\min_{v\in\mathbb{R}^p} J^W(v)$ *is uniformly tight and tight, respectively.*

(iii) *In particular,* $\hat{J}^W(\hat{v}_T) \xrightarrow{d} J^W(\hat{v})$.

PROOF: We have

$$\bar{\phi}_T(\theta^0 + T^{-1/4}v) = \bar{\phi}_T(\theta^0) + T^{-1/4}\frac{\partial\bar{\phi}_T}{\partial\theta'}(\theta^0)v + \frac{1}{2}T^{-1/2}\bar{\Delta}(v)$$

and

$$\begin{aligned}\hat{J}^W(v) &= T\bar{\phi}'_T(\theta^0 + T^{-1/4}v)W_T\bar{\phi}_T(\theta^0 + T^{-1/4}v)\\ &= T\bar{\phi}'_T(\theta^0)W_T\bar{\phi}_T(\theta^0) + 2T^{1/2}\bar{\phi}'_T(\theta^0)W_T T^{1/4}\frac{\partial\bar{\phi}_T}{\partial\theta'}(\theta^0)v\\ &\quad + T^{1/2}\bar{\phi}'_T(\theta^0)W_T\bar{G}\operatorname{Vec}(vv') + T^{1/2}v'\frac{\partial\bar{\phi}'_T}{\partial\theta}(\theta^0)W_T\frac{\partial\bar{\phi}_T}{\partial\theta'}(\theta^0)v\\ &\quad + T^{1/4}v'\frac{\partial\bar{\phi}'_T}{\partial\theta}(\theta^0)W_T\bar{G}\operatorname{Vec}(vv')\\ &\quad + \frac{1}{4}\operatorname{Vec}'(vv')\bar{G}'W_T\bar{G}\operatorname{Vec}(vv').\end{aligned}$$

Hence

$$\begin{aligned}\hat{J}^W(v) &= T\bar{\phi}'_T(\theta^0)W\bar{\phi}_T(\theta^0) + T^{1/2}\bar{\phi}'_T(\theta^0)WG\operatorname{Vec}(vv')\\ &\quad + \frac{1}{4}\operatorname{Vec}'(vv')G'WG\operatorname{Vec}(vv') + o_P(1),\end{aligned} \tag{B.4}$$

where the $o_P(1)$ term is, in fact, uniformly negligible over any compact subset of $\mathbb{R}^p$.

(i) We apply Theorem 1.5.4 of van der Vaart and Wellner (1996). To deduce that the stochastic process $\hat{J}^W(\cdot)$ converges in distribution toward $J^W(\cdot)$ in $\ell^\infty(\mathbb{K})$, this theorem requires the following conditions.

(a) The marginals $(\hat{J}^W(v_1), \dots, \hat{J}^W(v_k))$ converge in distribution toward $(J^W(v_1), \dots, J^W(v_k))$ for every finite subset $\{v_1, \dots, v_k\}$ of $\mathbb{K}$.

(b) The random process $\hat{J}^W(\cdot)$ is asymptotically tight.

To show (a), we observe that since the $o_P(1)$ term in (B.4) is uniformly negligible over any compact, $(\hat{J}^W(v_1), \dots, \hat{J}^W(v_k))$ is asymptotically equivalent to a continuous function of $\sqrt{T}\bar{\phi}_T(\theta^0)$ whose components are

$$\begin{aligned}&T\bar{\phi}'_T(\theta^0)W\bar{\phi}_T(\theta^0) + T^{1/2}\bar{\phi}'_T(\theta^0)WG\operatorname{Vec}(v_iv_i')\\ &\quad + \frac{1}{4}\operatorname{Vec}'(v_iv_i')G'WG\operatorname{Vec}(v_iv_i'), \quad i = 1, \dots, k.\end{aligned}$$

By the continuous mapping theorem, this latter converges in distribution toward $(J^W(v_1), \dots, J^W(v_k))$. This establishes (a).

To establish (b), we rely on Theorem 1.5.7 of van der Vaart and Wellner (1996). This theorem gives some sufficient conditions for the random process $\hat{J}^W(\cdot)$ to be asymptotically tight. From (a), $\hat{J}^W(v)$ converges in distribution toward $J^W(v)$ for any $v \in \mathbb{K}$. In addition, as a compact subset, $\mathbb{K}$ equipped with the usual metric on $\mathbb{R}^p$ is totally bounded. It remains to show that $\hat{J}^W(\cdot)$ is asymptotically uniformly equicontinuous in probability. That is, for any $\epsilon, \eta > 0$, there exists $\delta > 0$ such that

$$\limsup_T \text{Prob}\Big(\sup_{v_1, v_2 \in \mathbb{K}: \|v_1 - v_2\| < \delta} \big|\hat{J}^W(v_1) - \hat{J}^W(v_2)\big| > \epsilon\Big) < \eta.$$

From (B.4), $\hat{J}^W(v)$ is essentially a polynomial function of v and since $\mathbb{K}$ is bounded, we can write

$$\big|\hat{J}^W(v_1) - \hat{J}^W(v_2)\big| = X_T\|v_1 - v_2\| + o_P(1), \tag{B.5}$$

where $X_T = O_P(1)$. Let $\epsilon, \eta > 0$. Since $X_T = O_P(1)$, there exists $m_\eta > 0$ such that $\sup_T \text{Prob}(|X_T| > m_\eta) < \eta$. Let $\delta = \epsilon/(2m_\eta)$ and $A_T = (\sup_{v_1, v_2 \in \mathbb{K}: \|v_1 - v_2\| < \delta} |\hat{J}^W(v_1) - \hat{J}^W(v_2)| > \epsilon)$. We have

$$A_T = \big(A_T, |X_T| > m_\eta\big) \cup \big(A_T, |X_T| \le m_\eta\big).$$

Because they are uniformly negligible over $\mathbb{K}$, we can safely ignore the $o_P(1)$ term in (B.5) and write

$$\begin{aligned}\big(A_T, |X_T| \le m_\eta\big) &\subset \Big(\sup_{\|v_1 - v_2\| < \delta} |X_T|\|v_1 - v_2\| > \epsilon, |X_T| \le m_\eta\Big) \\ &\subset \big(|X_T| > 2m_\eta, |X_T| \le m_\eta\big) = \emptyset.\end{aligned}$$

Thus

$$\text{Prob}(A_T) \le \text{Prob}\big(|X_T| > m_\eta\big) < \eta.$$

As a result, $\limsup_T \text{Prob}(A_T) < \eta$ and this completes the proof of (b) and thus (i).

(ii) By definition, $\hat{v}_T = T^{1/4}(\hat{\theta}_T - \theta^0)$ and the uniform tightness of $\hat{v}_T$ follows from Proposition 3.1. Next consider $\hat{v} \in \arg\min_{v \in \mathbb{R}^p} J^W(v)$. Let $\epsilon > 0$. We have $0 \le \min_{v \in \mathbb{R}^p} J^W(v) \le J^W(0) = O_P(1)$; hence, there exists $m_1 > 0$ such that

$$\text{Prob}\Big(\min_{v \in \mathbb{R}^p} J^W(v) > m_1\Big) < \frac{\epsilon}{2}.$$

Note that the leading term in $J^W(v)$ is $\mathrm{Vec}'(vv')G'WG\,\mathrm{Vec}(vv')$ and we know from Lemma B.1 that $\gamma\|v\|^4 \leq \mathrm{Vec}'(vv')G'WG\,\mathrm{Vec}(vv')$, $\gamma > 0$. Therefore, for $\|v\|$ large enough, we can make $J^W(v)$ as large as desired with arbitrarily large probability. That is,

$$\forall \alpha, \beta > 0, \exists m_2 > 0: \quad \mathrm{Prob}\Big(\inf_{\|v\|>m_2} J^W(v) > \alpha\Big) > 1 - \beta.$$

We apply this with $\alpha = m_1$ and $\beta = \frac{\epsilon}{2}$, and observe that

$$\big(\|\hat{v}\| > m_2\big) = \big(\|\hat{v}\| > m_2, J^W(\hat{v}) > m_1\big) \cup \big(\|\hat{v}\| > m_2, J^W(\hat{v}) \leq m_1\big).$$

Thus

$$\begin{aligned}\mathrm{Prob}\big(\|\hat{v}\| > m_2\big) &\leq \mathrm{Prob}\big(J^W(\hat{v}) > m_1\big) + \mathrm{Prob}\Big(\inf_{\|v\|>m_2} J^W(v) \leq m_1\Big)\\ &\leq \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.\end{aligned}$$

This shows that $\hat{v}$ is tight.

(iii) This last point follows from Lemma B.5: since θ^0 is an interior point for Θ, the sequence $\mathbb{H}_T$ verifies the condition of this lemma. *Q.E.D.*

PROOF OF LEMMA 2.1: Let $\theta \in \mathbb{R}^n$, $\theta \neq 0$. We know that

$$\mathrm{Var}\big(\theta' Y_{t+1} | \mathfrak{F}_t\big) = \theta' \Lambda D_t \Lambda' \theta + \theta' \Omega \theta.$$

If $\Lambda'\theta = 0$, then $\mathrm{Var}(\theta' Y_{t+1}|\mathfrak{F}_t) = \mathrm{cst}$ and θ is a common feature. Conversely, if $\theta' \Lambda D_t \Lambda' \theta + \theta' \Omega \theta = \mathrm{cst}$, writing $c = \Lambda'\theta$, we have

$$\sum_{k=1}^{K} c_k^2 D_{kk,t} = \mathrm{cst}.$$

Hence, we have a linear combination of the terms in $\mathrm{Diag}(D_t)$ that is constant. From Assumption 1(ii), we necessarily have $c_k^2 = 0$, $k = 1, \ldots, K$. Thus $\Lambda'\theta = 0$. *Q.E.D.*

PROOF OF COROLLARY 3.1: Note that

$$\begin{aligned}\psi_t(\theta) &= z_t\big(\big(\theta' Y_{t+1}\big)^2 - c(\theta)\big)\\ &= (z_t - \bar{z}_T)\big(\big(\theta' Y_{t+1}\big)^2 - \bar{c}_T(\theta)\big)\\ &\quad + (z_t - \bar{z}_T)\big(\bar{c}_T(\theta) - c(\theta)\big) + \bar{z}_T\big(\big(\theta' Y_{t+1}\big)^2 - c(\theta)\big)\end{aligned}$$

and thus,

$$\begin{aligned}
\sqrt{T}\bar{\psi}_T(\theta) &= \sqrt{T}\bar{\phi}_T(\theta) + \bar{z}_T \cdot \left(\frac{\sqrt{T}}{T}\sum_{t=1}^{T}\left[\left(\theta' Y_{t+1}\right)^2 - c(\theta)\right]\right) \\
&= \sqrt{T}\bar{\phi}_T(\theta) + \mu_z \cdot \left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[\left(\theta' Y_{t+1}\right)^2 - c(\theta)\right]\right) \\
&\quad + (\bar{z}_T - \mu_z)\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[\left(\theta' Y_{t+1}\right)^2 - c(\theta)\right]\right) \\
&= \sqrt{T}\bar{\phi}_T(\theta) - \sqrt{T}\bar{v}_T(\theta) + o_P(1),
\end{aligned}$$

since $\sqrt{T}(\bar{z}_T - \mu_z) = O_P(1)$ by Assumption 4. Hence,

$$\begin{aligned}
&\sqrt{T}\bar{\phi}_T(\theta) = \sqrt{T}\bar{\psi}_T(\theta) + \sqrt{T}\bar{v}_T(\theta) + o_P(1), \\
&\sqrt{T}\frac{\partial\bar{\psi}_T}{\partial\theta'}(\theta) = \sqrt{T}\frac{\partial\bar{\phi}_T}{\partial\theta'}(\theta) + \sqrt{T}\frac{\partial\bar{v}_T}{\partial\theta'}(\theta) + o_P(1).
\end{aligned} \tag{B.6}$$

The second equation in (B.6) is obtained similarly to the first one. For $\theta = \theta^0$, $\psi_t(\theta^0) = z_t((\theta^{0'} Y_{t+1})^2 - c(\theta^0))$ and $v_t(\theta^0) = -\mu_z((\theta^{0'} Y_{t+1})^2 - c(\theta^0))$ are two martingale difference sequences with respect to $\mathfrak{F}_t$. Hence, the asymptotic variance, Σ, of $\sqrt{T}\bar{\phi}_T(\theta^0)$ is equal to

$$\begin{aligned}
\Sigma &= \operatorname{Var}\left(\psi_t\left(\theta^0\right) + v_t\left(\theta^0\right)\right) \\
&= \operatorname{Var}\left(\psi_t\left(\theta^0\right)\right) + \operatorname{Var}\left(v_t\left(\theta^0\right)\right) \\
&\quad + \operatorname{Cov}\left(v_t\left(\theta^0\right), \psi_t\left(\theta^0\right)\right) + \operatorname{Cov}\left(\psi_t\left(\theta^0\right), v_t\left(\theta^0\right)\right) \\
&= E\left(\psi_t\left(\theta^0\right)\psi_t'\left(\theta^0\right)\right) + E\left(v_t\left(\theta^0\right)v_t'\left(\theta^0\right)\right) \\
&\quad + E\left(v_t\left(\theta^0\right)\psi_t'\left(\theta^0\right)\right) + E\left(\psi_t\left(\theta^0\right)v_t'\left(\theta^0\right)\right) \\
&= E\left(z_t z_t'\left(\left(\theta^{0'} Y_{t+1}\right)^2 - c\left(\theta^0\right)\right)^2\right) + \mu_z\mu_z' E\left(\left(\theta^{0'} Y_{t+1}\right)^2 - c\left(\theta^0\right)\right)^2 \\
&\quad - E\left(\left(\left(\theta^{0'} Y_{t+1}\right)^2 - c\left(\theta^0\right)\right)^2 z_t\right)\mu_z' - \mu_z E\left(\left(\left(\theta^{0'} Y_{t+1}\right)^2 - c\left(\theta^0\right)\right)^2 z_t'\right) \\
&= E\left((z_t - \mu_z)(z_t - \mu_z)'\left(\left(\theta^{0'} Y_{t+1}\right)^2 - c\left(\theta^0\right)\right)^2\right).
\end{aligned}$$

Regarding the Jacobian, we mention that, under the conditions of the corollary,

$$\frac{\partial\psi_t}{\partial\theta'}\left(\theta^0\right) = z_t\left(\left(\theta^{0'} Y_{t+1}\right)Y_{t+1}' - E\left(\left(\theta^{0'} Y_{t+1}\right)Y_{t+1}'\right)\right)$$

is a martingale difference sequence, and, thanks to Assumption 4, the central limit theorem of Billingsley (1961) for stationary and ergodic martingale difference sequences applies to $\sqrt{T}\frac{\partial\bar{\psi}_T}{\partial\theta'}(\theta^0)$, which is asymptotically normal; therefore, $\sqrt{T}\frac{\partial\bar{\psi}_T}{\partial\theta'}(\theta^0) = O_P(1)$. *Q.E.D.*

PROOF OF PROPOSITION 3.1: We want to show that $\hat{v}_T = T^{1/4}(\hat{\theta}_T - \theta^0)$ is bounded in probability. We observe that as a second-order polynomial,

$$\sqrt{T}\bar{\phi}_T(\hat{\theta}_T) = \sqrt{T}\bar{\phi}_T(\theta^0) + \sqrt{T}\frac{\partial\bar{\phi}_T}{\partial\theta'}(\theta^0)(\hat{\theta}_T - \theta^0) + \frac{1}{2}\sqrt{T}\bar{\Delta}(\hat{\theta}_T - \theta^0).$$

From Corollary 3.1, $\sqrt{T}\bar{\phi}_T(\theta^0)$ and $\sqrt{T}\partial\bar{\phi}_T(\theta^0)/\partial\theta'$ are bounded in probability. Hence,

$$\sqrt{T}\bar{\phi}_T(\hat{\theta}_T) = \sqrt{T}\bar{\phi}_T(\theta^0) + \frac{1}{2}\sqrt{T}\bar{\Delta}(\hat{\theta}_T - \theta^0) + o_P(1)$$

and

$$\begin{aligned} &T\bar{\phi}'_T(\hat{\theta}_T)W_T\bar{\phi}_T(\hat{\theta}_T) \\ &\quad = T\bar{\phi}'_T(\theta^0)W_T\bar{\phi}_T(\theta^0) + \frac{T}{4}\bar{\Delta}'(\hat{\theta}_T - \theta^0)W_T\bar{\Delta}(\hat{\theta}_T - \theta^0) \\ &\qquad + T\bar{\Delta}'(\hat{\theta}_T - \theta^0)W_T\bar{\phi}_T(\theta^0) + o_P(\|\sqrt{T}\bar{\Delta}(\hat{\theta}_T - \theta^0)\|) + o_P(1). \end{aligned}$$

By definition,

$$T\bar{\phi}'_T(\theta^0)W_T\bar{\phi}_T(\theta^0) - T\bar{\phi}'_T(\hat{\theta}_T)W_T\bar{\phi}_T(\hat{\theta}_T) \geq 0$$

and we can write

$$\begin{aligned} &\frac{T}{4}\bar{\Delta}'(\hat{\theta}_T - \theta^0)W_T\bar{\Delta}(\hat{\theta}_T - \theta^0) \\ &\quad \leq -T\bar{\Delta}'(\hat{\theta}_T - \theta^0)W_T\bar{\phi}_T(\theta^0) + o_P(\|\sqrt{T}\bar{\Delta}(\hat{\theta}_T - \theta^0)\|) + o_P(1). \end{aligned} \tag{B.7}$$

Let $\hat{\delta} \equiv \text{Vec}((\hat{\theta}_T - \theta^0)(\hat{\theta}_T - \theta^0)')$. By definition, $\bar{\Delta}(\hat{\theta}_T - \theta^0) = \bar{G}\hat{\delta}$ and we have

$$\begin{aligned} &\bar{\Delta}'(\hat{\theta}_T - \theta^0)W_T\bar{\Delta}(\hat{\theta}_T - \theta^0) \\ &\quad = \hat{\delta}'\bar{G}'W_T\bar{G}\hat{\delta} \\ &\quad = \hat{\delta}'G'WG\hat{\delta} + \hat{\delta}'(\bar{G} - G)'W_T\bar{G}\hat{\delta} \\ &\qquad + \hat{\delta}'G'(W_T - W)\bar{G}\hat{\delta} + \hat{\delta}'G'W(\bar{G} - G)\hat{\delta}, \end{aligned}$$

and from (B.7), we can write

$$\frac{T}{4}\hat{\delta}'G'WG\hat{\delta} \le -T\hat{\delta}'(\bar{G}-G)'W_T\bar{\phi}_T(\theta^0) - T\hat{\delta}'G'(W_T-W)\bar{\phi}_T(\theta^0)$$
$$- T\hat{\delta}'G'W\bar{\phi}_T(\theta^0)$$
$$- \frac{T}{4}\hat{\delta}'(\bar{G}-G)'W_T\bar{G}\hat{\delta} - \frac{T}{4}\hat{\delta}'G'(W_T-W)\bar{G}\hat{\delta}$$
$$- \frac{T}{4}\hat{\delta}'G'W(\bar{G}-G)\hat{\delta} + o_P(\|\sqrt{T}\bar{G}\hat{\delta}\|) + o_P(1).$$

By the Cauchy–Schwarz inequality,

$$\frac{T}{4}\hat{\delta}'G'WG\hat{\delta} \le \sqrt{T}\|\hat{\delta}\|\|\bar{G}-G\|\|W_T\|\|\sqrt{T}\bar{\phi}_T(\theta^0)\|$$
$$+ \sqrt{T}\|\hat{\delta}\|\|G\|\|W_T-W\|\|\sqrt{T}\bar{\phi}_T(\theta^0)\|$$
$$+ \sqrt{T}\|\hat{\delta}\|\|G\|\|W\|\|\sqrt{T}\bar{\phi}_T(\theta^0)\|$$
$$+ \frac{T}{4}\|\hat{\delta}\|^2\|\bar{G}'-G'\|\big[\|W_T\|\|\bar{G}\| + \|W\|\|G\|\big]$$
$$+ \frac{T}{4}\|\hat{\delta}\|^2\|G\|\|W_T-W\|\|\bar{G}\|$$
$$+ \sqrt{T}\|\hat{\delta}\|\|\bar{G}\|o_P(1) + o_P(1).$$

Noting that $\|\hat{\delta}\| = \|\hat{\theta}_T - \theta^0\|^2$ and that W is symmetric positive definite, and also using Lemma B.1, we can write

$$\hat{\delta}'G'WG\hat{\delta} \ge \gamma_0\|\hat{\delta}'G'G\hat{\delta}\| = \gamma_0\|\Delta(\hat{\theta}_T-\theta^0)\|^2 \ge \gamma\|\hat{\theta}_T-\theta^0\|^4$$

for some $\gamma_0, \gamma > 0$. Hence

$$\gamma\|\hat{v}_T\|^4 \le 4\|\hat{v}_T\|^2\|G\|\|W\|\|\sqrt{T}\bar{\phi}_T(\theta^0)\|$$
$$+ \|\hat{v}_T\|^2 o_P(1) + \|\hat{v}_T\|^4 o_P(1) + o_P(1).$$

Dividing each side by $\|\hat{v}_T\|^2$ and after some rearrangements, we have

$$\|\hat{v}_T\|^2\big(\gamma + o_P(1)\big) \le 4\|G\|\|W\|\|\sqrt{T}\bar{\phi}_T(\theta^0)\| + \frac{o_P(1)}{\|\hat{v}_T\|^2} + o_P(1)$$

and, for T large enough, we can write

$$\|\hat{v}_T\|^2 \le \frac{4}{\gamma}\|G\|\|W\|\|\sqrt{T}\bar{\phi}_T(\theta^0)\| + \frac{o_P(1)}{\|\hat{v}_T\|^2} + o_P(1).$$

Hence, for large values of $\|\hat{v}_T\|^2$, the term $o_P(1)/\|\hat{v}_T\|^2$ stays asymptotically negligible in probability. Therefore, $\|\hat{v}_T\|^2$ is at most of the same asymptotic order of magnitude as $\|\sqrt{T}\bar{\phi}_T(\theta^0)\|$. This establishes that $\|\hat{v}_T\|^2 = O_P(1)$ or, equivalently, $\|\hat{v}_T\| = O_P(1)$. *Q.E.D.*

PROOF OF PROPOSITION 3.2: Since $Z_T(\theta^0)$ is a continuous function of $\sqrt{T}\bar{\phi}_T(\theta^0)$, it suffices to show that the sequence $(T^{1/4}(\hat{\theta}_T - \theta^0)', \sqrt{T}\bar{\phi}_T(\theta^0)')'$ has a subsequence that converges in distribution. From Proposition 3.1, $T^{1/4}(\hat{\theta}_T - \theta^0)$ is uniformly tight and $\sqrt{T}\bar{\phi}_T(\theta^0)$ is also uniformly tight following Assumption 4. Thus, these two random vectors, which are measurable (we implicitly assume $\hat{\theta}_T$ is measurable; this is a common assumption in the literature on extremum estimators) on the same probability space, are jointly uniformly tight. Therefore, from the Prohorov's theorem (see Theorem 2.4 of van der Vaart (1998)), the joint sequence has a subsequence that converges in distribution. This establishes the first part of the proposition.

Next we show that

$$0 < \text{Prob}\big(Z(X) \geq 0\big) \leq \frac{1}{2}.$$

First note that for any vector d, we have

$$\text{Prob}\big(Z(X) \geq 0\big) \leq \text{Prob}\big(d'Z(X)d \geq 0\big).$$

Moreover, since $\text{Vec}(Z(X))$ is a Gaussian vector with zero mean, $d'Z(X)d$ is a Gaussian real random variable with zero mean. Therefore, we will deduce the inequality $\text{Prob}(Z(X) \geq 0) \leq 1/2$ if we can show that, for at least one vector d, $d'Z(X)d$ has a positive variance. However, an obvious implication of Lemma 2.3 is that for all $i = 1, \dots, p$,

$$\frac{\partial^2 \rho}{\partial \theta_i^2}\big(\theta^0\big) \neq 0. \tag{B.8}$$

Otherwise, if we had $\frac{\partial^2 \rho}{\partial \theta_i^2}(\theta^0) = 0$ for some specific component i, by considering a vector $u \in \mathbb{R}^p$ with all components equal to zero except the ith component, we would have

$$u' \frac{\partial^2 \rho_h}{\partial \theta \, \partial \theta'}\big(\theta^0\big)u = 0, \quad \forall h = 1, \dots, H.$$

This contradicts Lemma 2.3, since θ^0 being in the interior of the set Θ, we can always find $\theta \neq \theta^0$ such that $\theta \in \Theta$ and $(\theta - \theta^0)$ is proportional to the aforementioned vector u.

From (B.8), we deduce that if $d = e_i$, the vector of $\mathbb{R}^p$ with all components equal to 0 except the ith component that is equal to 1, we have

$$d'Z(X)d = e_i'Z(X)e_i = Z_{ii}(X) = \frac{\partial^2 \rho'}{\partial \theta_i^2}(\theta^0)WX$$

and

$$\mathrm{Var}\big(d'Z(X)d\big) = \frac{\partial^2 \rho'}{\partial \theta_i^2}(\theta^0)W\Sigma(\theta^0)W\frac{\partial^2 \rho}{\partial \theta_i^2}(\theta^0) \neq 0$$

since the matrix $W\Sigma(\theta^0)W$ is positive definite. Hence the announced upper bound for $\mathrm{Prob}(Z(X) \geq 0)$. The announced lower bound will be obtained by showing that

$$\text{(B.9)} \qquad \mathrm{Prob}\big(Z(X) \geq 0\big) \geq \mathrm{Prob}(X \in U)$$

for a nonempty open set U in $\mathbb{R}^H$. Since X is an H-dimensional Gaussian vector with a nonsingular variance matrix, we can be sure that it has a positive probability to take values in any nonempty open set of $\mathbb{R}^H$, and, thus, inequality (B.9) will give us the required lower bound for $\mathrm{Prob}(Z(X) \geq 0)$. The open set U will be defined as

$$U = \big\{\lambda \in \mathbb{R}^H : \mathrm{Cov}\big(\sigma_{kt}^2, \lambda' z_t\big) > 0, \forall k = 1, \ldots, K\big\}.$$

That U is open is by an obvious continuity argument.

By Assumption 3, we know that the range of the matrix $\mathrm{Cov}(\mathrm{Diag}(Dt), z_t)$ is $\mathbb{R}^K$. Therefore, if we consider any vector a in $\mathbb{R}^K$, we can find $\lambda \in \mathbb{R}^H$ such that

$$\mathrm{Cov}\big(\mathrm{Diag}(Dt), z_t\big)\lambda = a.$$

In particular, when choosing a with all components strictly positive, we get λ in the open set U. Thus U is not empty. Hence, we just have to prove the inequality (B.9). Of course, it is sufficient to prove that

$$U \subset \big\{\lambda \in \mathbb{R}^H : Z(\lambda) \geq 0\big\}.$$

We note that

$$Z(\lambda) = \left(\frac{\partial^2 \rho'}{\partial \theta_i \partial \theta_j}(\theta^0)\lambda\right)_{1 \leq i,j \leq p} = \sum_{h=1}^{H} \lambda_h \frac{\partial^2 \rho_h}{\partial \theta \partial \theta'}(\theta^0).$$

We recall that

$$\begin{aligned}\rho_h(\theta) &= E\big(\big(z_{ht} - E(z_{ht})\big)\big(\delta' Y_{t+1}\big)^2\big) \\ &= E\big(\big(z_{ht} - E(z_{ht})\big)\big((L\theta + r)' Y_{t+1}\big)^2\big)\end{aligned}$$

if we use the affine normalization condition

$$\delta = L\theta + r.$$

Hence

$$\frac{\partial^2 \rho_h}{\partial\theta\,\partial\theta'}(\theta^0) = 2 \times E\big((z_{ht} - E(z_{ht}))L'Y_{t+1}Y'_{t+1}L\big).$$

By the law of iterated expectations, the expectation in this second derivative's expression is equal to

$$\begin{aligned} &E\big((z_{ht} - E(z_{ht}))L'E(Y_{t+1}Y'_{t+1}|\mathfrak{F}_t)L\big) \\ &\quad = E\big((z_{ht} - E(z_{ht}))L'\operatorname{Var}(Y_{t+1}|\mathfrak{F}_t)L\big). \end{aligned}$$

Recalling that $\operatorname{Var}(Y_{t+1}|\mathfrak{F}_t) = \Lambda D_t \Lambda' + \Omega$ and that D_t is diagonal, we have $L'\Lambda D_t \Lambda' L = \sum_{k=1}^{K} \sigma_{kt}^2 l_k l_k'$, where l_k is the kth column of $L'\Lambda$. Thus,

$$\begin{aligned} E\big((z_{ht} - E(z_{ht}))L'\operatorname{Var}(Y_{t+1}|\mathfrak{F}_t)L\big) &= \sum_{k=1}^{K} E\big((z_{ht} - E(z_{ht}))\sigma_{kt}^2 l_k l_k'\big) \\ &= \sum_{k=1}^{K} \operatorname{Cov}(z_{ht}, \sigma_{kt}^2) l_k l_k'. \end{aligned}$$

Hence,

$$\sum_{h=1}^{H} \lambda_h \frac{\partial^2 \rho_h}{\partial\theta\,\partial\theta'}(\theta^0) = 2 \times \sum_{k=1}^{K} \operatorname{Cov}(\lambda' z_t, \sigma_{kt}^2) l_k l_k'.$$

This matrix is obviously positive semidefinite for all $\lambda \in U$. This completes the proof.

We now want to assess the probabilities $\operatorname{Prob}(V = 0|Z(X) \geq 0)$ and $\operatorname{Prob}(V = 0|\overline{(Z(X) \geq 0)})$. Let us start with the former.

Since $\hat{\theta}_T - \theta^0 = O_P(T^{-1/4})$, we have

$$\text{(B.10)}\quad \begin{aligned} &\sqrt{T}\bar{\phi}_T(\hat{\theta}_T) \\ &\quad = \sqrt{T}\bar{\phi}_T(\theta^0) \\ &\qquad + \frac{1}{2}\sqrt{T}\bigg((\hat{\theta}_T - \theta^0)' \frac{\partial^2 \rho_h}{\partial\theta\,\partial\theta'}(\theta^0)(\hat{\theta}_T - \theta^0)\bigg)_{1\leq h\leq H} + o_P(1). \end{aligned}$$

In particular, $\sqrt{T}\bar{\phi}_T(\hat{\theta}_T) = O_P(1)$ and, thus,

$$J_T = T\bar{\phi}'_T(\hat{\theta}_T)W\bar{\phi}_T(\hat{\theta}_T) + o_P(1).$$

For the sake of expositional simplicity, we will consider $W = \mathrm{Id}_H$. This is not restrictive as it amounts to rescaling $\phi_{t,T}(\theta)$ by $W^{1/2}$. We keep $\phi_{t,T}(\theta)$ for $W^{1/2}\phi_{t,T}(\theta)$ in the rest of this proof for economy of notation. Thus

$$\begin{aligned} J_T &= T\bar{\phi}'_T(\hat{\theta}_T)\bar{\phi}_T(\hat{\theta}_T) + o_P(1) \\ &= T\bar{\phi}'_T(\theta^0)\bar{\phi}_T(\theta^0) + \Delta'\big(T^{1/4}(\hat{\theta}_T - \theta^0)\big)\sqrt{T}\bar{\phi}_T(\theta^0) \\ &\quad + \frac{1}{4}\Delta'\big(T^{1/4}(\hat{\theta}_T - \theta^0)\big)\Delta\big(T^{1/4}(\hat{\theta}_T - \theta^0)\big) + o_P(1). \end{aligned}$$

By definition, $J_T \leq T\bar{\phi}'_T(\theta^0)\bar{\phi}_T(\theta^0)$. Hence

$$\begin{aligned} &\Delta'\big(T^{1/4}(\hat{\theta}_T - \theta^0)\big)\sqrt{T}\bar{\phi}_T(\theta^0) \\ &\quad + \frac{1}{4}\Delta'\big(T^{1/4}(\hat{\theta}_T - \theta^0)\big)\Delta\big(T^{1/4}(\hat{\theta}_T - \theta^0)\big) \leq o_P(1). \end{aligned} \tag{B.11}$$

It is worth noting that

$$\begin{aligned} &\Delta'\big(T^{1/4}(\hat{\theta}_T - \theta^0)\big)\sqrt{T}\bar{\phi}_T(\theta^0) \\ &\quad = \big(T^{1/4}(\hat{\theta}_T - \theta^0)\big)' Z_T(\theta^0)\big(T^{1/4}(\hat{\theta}_T - \theta^0)\big). \end{aligned} \tag{B.12}$$

Actually, each side of (B.12) is equal to

$$\sum_{i,j=1}^{p}\sum_{h=1}^{H}\left(\frac{\partial^2\rho_h(\theta^0)}{\partial\theta_i\,\partial\theta_j}\sqrt{T}\bar{\phi}_{T,h}(\theta^0)\right)\big(T^{1/4}(\hat{\theta}_{T,i} - \theta_i^0)\big)\big(T^{1/4}(\hat{\theta}_{T,j} - \theta_j^0)\big).$$

Considering a subsequence of $(T^{1/4}(\hat{\theta}_T - \theta^0)', \mathrm{Vec}'(Z_T(\theta^0)))'$ that converges in distribution toward a certain random vector $(V', \mathrm{Vec}'(Z(X)))'$, we can write (for the sake of simplicity, we do not make explicit the notation for a subsequence)

$$\Delta'\big(T^{1/4}(\hat{\theta}_T - \theta^0)\big)\sqrt{T}\bar{\phi}_T(\theta^0) \xrightarrow{d} V'Z(X)V.$$

From (B.11) and by Lemma B.2, we deduce that

$$\begin{aligned} \limsup_{T\to\infty} \mathrm{Prob}\bigg(&\Delta'\big(T^{1/4}(\hat{\theta}_T - \theta^0)\big)\sqrt{T}\bar{\phi}_T(\theta^0) \\ &+ \frac{1}{4}\Delta'\big(T^{1/4}(\hat{\theta}_T - \theta^0)\big)\Delta\big(T^{1/4}(\hat{\theta}_T - \theta^0)\big) \leq \epsilon\bigg) = 1 \end{aligned}$$

for any $\epsilon > 0$, and by the portmanteau lemma (Lemma 2.2(vi) of van der Vaart (1998)), we have

$$\text{Prob}\left(V'Z(X)V + \frac{1}{4}\Delta'(V)\Delta(V) \leq \epsilon\right) = 1, \quad \forall \epsilon > 0.$$

We deduce, by right continuity of cumulative distribution functions, that

$$\text{Prob}\left(V'Z(X)V + \frac{1}{4}\Delta'(V)\Delta(V) \leq 0\right) = 1.$$

In particular, if $Z(X)$ is positive semidefinite, then

$$\Delta'(V)\Delta(V) = 0 \quad \text{almost surely}$$

and, thus,

$$\|\Delta(V)\| = 0 \quad \text{almost surely}.$$

But, by Lemma B.1,

$$\|\Delta(V)\| \geq \gamma \|V\|^2.$$

Thus, $V = 0$ almost surely. In other words, we have shown that

$$\text{Prob}\big(V = 0 | Z(X) \geq 0\big) = 1.$$

Now let us establish that $\text{Prob}(V = 0|\overline{(Z(X) \geq 0)}) = 0$.

The necessary second-order condition for an interior solution for a minimization problem implies that for any vector $e \in \mathbb{R}^p$,

$$e'\left(\frac{\partial^2}{\partial\theta\,\partial\theta'}\big(\bar{\phi}'_T(\theta)\bar{\phi}_T(\theta)\big)\bigg|_{\theta=\hat{\theta}_T}\right)e \geq 0.$$

This can be written

$$e'(\tilde{Z}_T + N_T)e \geq 0, \tag{B.13}$$

where

$$\tilde{Z}_T = \left(\frac{\partial^2 \bar{\phi}'_T}{\partial\theta_i\,\partial\theta_j}(\hat{\theta}_T)\sqrt{T}\bar{\phi}_T(\hat{\theta}_T)\right)_{1\leq i,j\leq p}$$

and

$$N_T = \sqrt{T}\frac{\partial\bar{\phi}'_T}{\partial\theta}(\hat{\theta}_T)\frac{\partial\bar{\phi}_T}{\partial\theta'}(\hat{\theta}_T).$$

By a mean value expansion, we have

$$\frac{\partial \bar{\phi}_T}{\partial \theta_i}(\hat{\theta}_T) = \frac{\partial^2 \bar{\phi}_T}{\partial \theta_i \partial \theta'}(\bar{\theta})\big(\hat{\theta}_T - \theta^0\big) + O_P\big(T^{-1/2}\big), \tag{B.14}$$

with $\bar{\theta} \in (\theta^0, \hat{\theta}_T)$, which may differ from row to row, and $i = 1, \ldots, p$. On the other hand, thanks to Equation (B.10), we have

$$\begin{aligned}&\frac{\partial^2 \bar{\phi}'_T}{\partial \theta_i \partial \theta_j}(\hat{\theta}_T)\bar{\phi}_T(\hat{\theta}_T)\\ &\quad= \frac{\partial^2 \rho'}{\partial \theta_i \partial \theta_j}(\theta^0)\left(\bar{\phi}_T\big(\theta^0\big) + \frac{1}{2}\Delta\big(\hat{\theta}_T - \theta^0\big)\right) + o_P\big(T^{-1/2}\big).\end{aligned}$$

Hence, with $h_{ij} = \frac{\partial^2 \rho}{\partial \theta_i \partial \theta_j}(\theta^0)$,

$$\begin{aligned}&\frac{\partial^2 \bar{\phi}'_T}{\partial \theta_i \partial \theta_j}(\hat{\theta}_T)\sqrt{T}\bar{\phi}_T(\hat{\theta}_T)\\ &\quad= h'_{ij}\sqrt{T}\bar{\phi}_T\big(\theta^0\big) + \frac{1}{2}h'_{ij}\Delta\big(T^{1/4}\big(\hat{\theta}_T - \theta^0\big)\big) + o_P(1).\end{aligned}$$

Thus

$$\tilde{Z}_T = Z_T\big(\theta^0\big) + \frac{1}{2}\big(h'_{ij}\Delta\big(T^{1/4}\big(\hat{\theta}_T - \theta^0\big)\big)\big)_{1 \le i,j \le p} + o_P(1)$$

and

$$\begin{aligned}N_T = {}&\left(T^{1/4}\big(\hat{\theta}_T - \theta^0\big)'\frac{\partial^2 \rho'}{\partial \theta_i \partial \theta}\big(\theta^0\big)\frac{\partial^2 \rho}{\partial \theta_j \partial \theta'}\big(\theta^0\big)T^{1/4}\big(\hat{\theta}_T - \theta^0\big)\right)_{1 \le i,j \le p}\\ &+ o_P(1).\end{aligned}$$

From the inequality (B.13) and some successive applications of the Cauchy–Schwarz inequality, we can find a constant $A > 0$ such that for any vector $e \in \mathbb{R}^p$ with unit norm,

$$-e'Z_T\big(\theta^0\big)e \le A\sqrt{T}\big\|\hat{\theta}_T - \theta^0\big\|^2 + o_P(1).$$

By Lemma B.2,

$$\limsup_{T \to \infty} \operatorname{Prob}\big(-e'Z_T\big(\theta^0\big)e - A\sqrt{T}\big\|\hat{\theta}_T - \theta^0\big\|^2 \le \epsilon\big) = 1, \quad \forall \epsilon > 0.$$

Considering again a subsequence along which $(T^{1/4}(\hat{\theta}_T - \theta^0)', \sqrt{T}\bar{\phi}_T(\theta^0)')'$ converges in distribution, we can write, using the portmanteau lemma (Lemma

2.2(vi) of van der Vaart (1998)), that

$$\text{Prob}\big(-e'Z(X)e - A\|V\|^2 \leq \epsilon\big) = 1, \quad \forall \epsilon > 0.$$

Thus, by right continuity of cumulative distribution functions,

$$\text{Prob}\big(-e'Z(X)e - A\|V\|^2 \leq 0\big) = 1$$

and, consequently,

$$\text{Prob}\left(\|V\|^2 \geq -\frac{e'Z(X)e}{A}\middle| Z(X) = z\right) = 1, \quad P^Z \text{ a.s.} \tag{B.15}$$

In particular, when $Z(X) = z$ is nonpositive semidefinite, we can find a vector $e \in \mathbb{R}^p$ with unit norm and such that $e'Z(X)e < 0$, and thus

$$\text{Prob}\big(\|V\| > 0 | Z(X) = z\big) = 1.$$

Therefore, $\text{Prob}(\|V\| > 0 | \overline{(Z(X) \geq 0)}) = 1$. *Q.E.D.*

The proof of Theorem 3.1 follows from Lemma B.6(iii).

PROOF OF LEMMA 3.1: The first-order condition associated to (10) is

$$G'WX + \frac{1}{2}G'WG\hat{u} = 0.$$

Moreover, we know from Lemma B.4 that we can decompose $G = G_1G_2$ with G_1 (resp. G_2) full column rank (resp. full row rank) p. Since G_2' is full column rank, the above first-order conditions are equivalent to

$$G_1'WX + \frac{1}{2}G_1'WG_1G_2\hat{u} = 0,$$

and since G_1 is full column rank, we deduce

$$G_2\hat{u} = -2\big(G_1'WG_1\big)^{-1}G_1'WX.$$

Defining

$$\tilde{X} = W^{1/2}X, \quad \tilde{G}_1 = W^{1/2}G_1,$$

we see that

$$\tilde{G}_1G_2\hat{u} = -2\tilde{G}_1\big(\tilde{G}_1'\tilde{G}_1\big)^{-1}\tilde{G}_1'\tilde{X} = -2P_1\tilde{X},$$

where P_1 stands for the matrix of orthogonal projection on the p-dimensional subspace of $\mathbb{R}^H$ spanned by the columns of $\tilde{G}_1$. Plugging in (10), we deduce

$$\begin{aligned} L &= \tilde{X}'\tilde{X} + \tilde{X}'\tilde{G}_1 G_2 \hat{u} + \frac{1}{4}\hat{u}' G_2' \tilde{G}_1' \tilde{G}_1 G_2 \hat{u} \\ &= \tilde{X}'\tilde{X} - 2\tilde{X}' P_1 \tilde{X} + \tilde{X}' P_1 \tilde{X} = \tilde{X}'(\mathrm{Id} - P_1)\tilde{X}. \end{aligned}$$

Thus

$$L = \left\| (\mathrm{Id} - P_1)\tilde{X} \right\|^2 \sim \chi^2(H - p)$$

since, for $W = \Sigma^{-1}(\theta^0)$, $\tilde{X} = W^{1/2}X$ is a standardized Gaussian vector. Since $J(0) = X'WX = \|\tilde{X}\|^2$,

$$S = J(0) - L = \|P_1\tilde{X}\|^2 \sim \chi^2(p).$$

Moreover, since $P_1\tilde{X}$ and $(\mathrm{Id} - P_1)\tilde{X}$ are stochastically independent (orthogonal projections of standard Gaussian vectors on two orthogonal subspaces), L is independent of S and, of course, $J(0) = S + L \sim \chi^2(H)$.

In addition, elementary computations give

$$\begin{aligned} \mathrm{Vec}\big(Z(X)\big) &= G_2' G_1' W X = G_2' \big(\tilde{G}_1' \tilde{G}_1\big)\big(\tilde{G}_1' \tilde{G}_1\big)^{-1} G_1' W X \\ &= G_2' \tilde{G}_1' P_1 \tilde{X}. \end{aligned}$$

Therefore, L is actually jointly independent of $(S, Z(X))$. *Q.E.D.*

PROOF OF THEOREM 3.2: By definition of the minimization problems, we obviously have

$$L \leq J \leq J(0).$$

Moreover, from the alternative expression of $J(v)$ given by (8), one can easily see that when $Z(X) \geq 0$, the minimum of $J(v)$ is reached at $v = 0$, leading to $J = J(0)$.

Part (ii) of Theorem 3.2 will be proved in two steps:

Step 1. We show that there exists some $\varepsilon > 0$ such that

$$\mathrm{Prob}\big(J > L + \varepsilon, Z(X) \geq 0\big) > 0.$$

To see this, first note that since $S = J(0) - L \sim \chi^2(p)$, $\mathrm{Prob}(J(0) > L) = 1$. Thus

$$\mathrm{Prob}\big(J(0) > L, Z(X) \geq 0\big) = \mathrm{Prob}\big(Z(X) \geq 0\big).$$

But we know that

$$Z(X) \geq 0 \quad \Rightarrow \quad J = J(0).$$

Therefore,

$$\text{Prob}\big(J > L, Z(X) \geq 0\big) = \text{Prob}\big(Z(X) \geq 0\big).$$

However,

$$\begin{aligned}\text{Prob}\big(J > L, Z(X) \geq 0\big) &= \text{Prob}\left(\bigcup_{n\geq 1}\left(J > L + \frac{1}{n}\right), Z(X) \geq 0\right)\\ &= \lim_{n\to\infty} \text{Prob}\left(J > L + \frac{1}{n}, Z(X) \geq 0\right).\end{aligned}$$

Then

$$\lim_{n\to\infty} \text{Prob}\left(J > L + \frac{1}{n}, Z(X) \geq 0\right) = \text{Prob}\big(Z(X) \geq 0\big) > 0$$

and we deduce that there exists $n \in \mathbb{N}$ such that

$$\text{Prob}\left(J > L + \frac{1}{n}, Z(X) \geq 0\right) > 0.$$

Step 2. Following Remark 3.2, we actually show that

$$\text{Prob}\big(L > c, Z(X) \geq 0\big) < \text{Prob}\big(J > c, Z(X) \geq 0\big) \quad \forall c > 0.$$

Since we always have $L \leq J$, we have for any measurable part B of the sample space,

$$\begin{aligned}&\text{Prob}\big(L > c, \big(Z(X) \geq 0\big) \cap B\big)\\ &\quad \leq \text{Prob}\big(J > c, \big(Z(X) \geq 0\big) \cap B\big) \quad \forall c > 0.\end{aligned}$$

We will then obviously be able to deduce the announced strict inequality if we show that

$$\begin{aligned}&\text{Prob}\big(L > c, Z(X) \geq 0, J > L + \varepsilon\big)\\ &\quad < \text{Prob}\big(J > c, Z(X) \geq 0, J > L + \varepsilon\big) \quad \forall c > 0.\end{aligned}$$

But since again $L \leq J$, then

$$\begin{aligned}&\text{Prob}\big(J > c, Z(X) \geq 0, J > L + \varepsilon\big)\\ &\quad = \text{Prob}\big(L > c, Z(X) \geq 0, J > L + \varepsilon\big)\\ &\qquad + \text{Prob}\big(L \leq c, J > c, Z(X) \geq 0, J > L + \varepsilon\big).\end{aligned}$$

Hence we only need to show that

$$\text{Prob}\big(L \le c, J > c, Z(X) \ge 0, J > L + \varepsilon\big) > 0 \quad \forall c > 0.$$

Since when $Z(X) \ge 0$, $J = J(0)$ and, thus, $L = J - S$, we want to show that

$$\text{Prob}\big(c - S < L \le c, S > \varepsilon, Z(X) \ge 0\big) > 0.$$

Let $F_{L,S,Z(X)}(l, s, z)$ (resp. $F_L(l)$ and $F_{S,Z(X)}(s, z)$) be the joint distribution of $(L, S, Z(X))$ (resp. L, and $(S, Z(X))$) and let $I(\cdot)$ be the usual indicator function. We have

$$\begin{aligned}
&\text{Prob}\big(c - S < L \le c, S > \varepsilon, Z(X) \ge 0\big)\\
&\quad = \int I(c - s < l \le c, s > \varepsilon, z \ge 0)\, dF_{L,S,Z(X)}(l, s, z)\\
&\quad = \int I(c - s < l \le c, s > \varepsilon, z \ge 0)\, dF_L(l)\, dF_{S,Z(X)}(s, z)\\
&\quad = \int_{s>\varepsilon, z\ge 0} \left(\int_{c-s<l\le c} dF_L(l) \right) dF_{S,Z(X)}(s, z)\\
&\quad = \int_{s>\varepsilon, z\ge 0} \big(\text{Prob}\big(c - s < \chi^2(H - p) \le c\big)\big)\, dF_{S,Z(X)}(s, z),
\end{aligned}$$

where second equality follows from the independence of L and $(S, Z(X))$, and the last equality follows from the fact that $L \sim \chi^2(H - p)$.

But

$$\begin{aligned}
&\forall s > \varepsilon \ge 0 \text{ and } \forall c,\\
&\quad \text{Prob}\big(c - s < \chi^2(H - p) \le c\big) \ge \text{Prob}\big(c - \varepsilon < \chi^2(H - p) \le c\big).
\end{aligned}$$

Hence,

$$\begin{aligned}
&\int_{s>\varepsilon, z\ge 0} \big(\text{Prob}\big(c - s < \chi^2(H - p) \le c\big)\big)\, dF_{S,Z(X)}(s, z)\\
&\quad \ge \int_{s>\varepsilon, z\ge 0} \big(\text{Prob}\big(c - \varepsilon < \chi^2(H - p) \le c\big)\big)\, dF_{S,Z(X)}(s, z)\\
&\quad = \text{Prob}\big(c - \varepsilon < \chi^2(H - p) \le c\big) \int_{s>\varepsilon, z\ge 0} dF_{S,Z(X)}(s, z)\\
&\quad = \text{Prob}\big(c - \varepsilon < \chi^2(H - p) \le c\big)\, \text{Prob}\big(S > \varepsilon, Z(X) \ge 0\big).
\end{aligned}$$

Since, by continuity and positivity on the positive half line of the $\chi^2(H - p)$ distribution, $\text{Prob}(c - \varepsilon < \chi^2(H - p) \le c) > 0$ for all $c, \varepsilon > 0$ and $\text{Prob}(S >$

$\varepsilon, Z(X) \geq 0) > 0$ from the Step 1, we conclude that $\text{Prob}(c - S < L \leq c, S > \varepsilon, Z(X) \geq 0) > 0$, which concludes the proof. *Q.E.D.*

PROOF OF COROLLARY 3.2: Since $p = 1$, G is the column vector $\frac{\partial^2 \rho}{\partial \theta^2}(\theta^0)$ and Lemma 2.3 guarantees that $G \neq 0$. Also, J now has the expression

$$J = \min_{v \in \mathbb{R}} \left(X'WX + X'WGv^2 + \frac{1}{4} G'WGv^4 \right).$$

The first-order necessary condition for optimality gives $v(2X'WG + G'WGv^2) = 0$, while the second-order sufficient condition for $v = 0$ to be a solution is $X'WG > 0$. If $X'WG < 0$, we can say, from the first-order condition, that any solution v satisfies $2X'WG + G'WGv^2 = 0$. In the event that $X'WG = 0$, it appears that $v = 0$ is a solution. In summary, we can write that if $X'WG \geq 0$, then

$$J = J(0) = X'WX,$$

and if $X'WG < 0$, $J(v)$ is minimized at $v^2 = -2X'WG/G'WG$ so that

$$\begin{aligned} J &= X'W^{1/2}\big(\text{Id}_H - W^{1/2}G\big(G'WG\big)^{-1}G'W^{1/2}\big)W^{1/2}X \\ &\equiv X'W^{1/2}\mathcal{P}W^{1/2}X \equiv L. \end{aligned}$$

The variable $\mathcal{P}$ is the orthogonal projection matrix on the orthogonal of the subspace generated by the column vectors of $W^{1/2}G$.

Letting $z = \frac{X'WG}{\sqrt{G'WG}}$, $z \sim N(0, 1)$ and

$$J = I(z \geq 0)J(0) + I(z < 0)L.$$

Now we show that $I(z \geq 0)$ is independent of both $J(0)$ and L.

We have $\text{Cov}(z, \mathcal{P}W^{1/2}X) = \text{Cov}(\frac{X'WG}{\sqrt{G'WG}}, \mathcal{P}W^{1/2}X) = 0$. Thus, since X is a Gaussian vector, z is independent of $\mathcal{P}W^{1/2}X$, and so are $I(z \geq 0)$ and L.

To see that $I(z \geq 0)$ is independent of $J(0)$, we write $W^{1/2}X$ in the orthonormal basis

$$\big(W^{1/2}a_1, W^{1/2}a_2, \ldots, W^{1/2}a_H\big)$$

of $\mathbb{R}^H$ such that $a_1 = \frac{G}{\sqrt{G'WG}}$. (We choose a_1 such that the first component of $W^{1/2}X$ in this new basis is z.) The coordinates of $W^{1/2}X$ in this basis are $(a_1'WX, a_2'WX, \ldots, a_H'WX)$ and by the invariance of the norm,

$$J(0) = X'WX = \sum_{h=1}^{H} \big(a_h'WX\big)^2.$$

Note that $\text{Cov}(a_j'WX, a_j'WX) = 0$ for $i \neq j$ so that $z = a_1'WX$ is independent of $a_j'WX$, $j = 2, \ldots, H$. Hence, to claim that $I(z \geq 0)$ is independent of $J(0)$, it is sufficient to show that $(a_1'WX)^2 = z^2$ is independent of $I(z \geq 0)$. This becomes obvious once we see that $z \sim N(0, 1)$ has a symmetric distribution about the origin. This completes the proof. *Q.E.D.*

REFERENCES

ANDREWS, I., AND A. MIKUSHEVA (2012): "Maximum Likelihood Inference in Weakly Identified DSGE Models," Working Paper, MIT, available at http://economics.mit.edu/files/6593. [1-3]

BILLINGSLEY, P. (1961): "The Lindeberg–Lévy Theorem for Martingales," *Proceedings of the American Mathematical Society*, 12, 788–792. [15]

GOURIEROUX, C., AND A. MONFORT (1989): "A General Framework for Testing a Null Hypothesis in a 'Mixed' Form," *Econometric Theory*, 5 (1), 63–82. [1,3]

KLEIBERGEN, F. (2005): "Testing Parameters in GMM Without Assuming That They Are Identified," *Econometrica*, 73, 1103–1123. [1,3]

STOCK, J. H., AND J. H. WRIGHT (2000): "GMM With Weak Identification," *Econometrica*, 68, 1055–1096. [1,4]

VAN DER VAART, A. W. (1998): *Asymptotic Statistics*. Cambridge: Cambridge University Press. [6,7,17,21,23]

VAN DER VAART, A. W., AND J. A. WELLNER (1996): *Stochastic Convergence and Empirical Processes*. New York: Springer-Verlag. [11,12]

Dept. of Economics, Concordia University, 1455 de Maisonneuve Blvd. West, Montreal, Quebec, H3G 1M8 Canada, and CIRANO, and CIREQ; prosper.dovonon@concordia.ca

and

Dept. of Economics, Brown University, 64 Waterman Street, Providence, RI 02912, U.S.A., and CIRANO, and CIREQ; Eric_Renault@brown.edu.

Manuscript received June, 2011; final revision received March, 2013.